Proceedings of the

1995 IEEE International
Frequency Control Symposium

(The 49th Annual Symposium)

Sponsored by
The Institute of Electrical and
Electronics Engineers, Inc.

Ultrasonics, Ferroelectrics, and
Frequency Control Society

with the participation of personnel of the
Army Research Laboratory, Ft. Monmouth, NJ, U.S.A.

IEEE Catalog No. 95CH35752

Library of Congress No. 87-654207

31 May - 2 June 1995

The Fairmont Hotel
San Francisco, California, U.S.A.

Copyright and Reprint Permission: Abstracting is permitted with the credit to the source. Libraries are permitted to photocopy beyond the limits of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. All rights reserved. Copyright © 1995 by the Institute of Electrical and Electronics Engineers, Inc.

IEEE Catalog Number: 95CH35752
ISBN: 0-7803-2500-1 (Softbound Edition)
ISBN: 0-7803-2501-X (Casebound Edition)
ISBN: 0-7803-2502-8 (Microfiche Edition)
Library of Congress Number: 87-654207
ISSN: 1075-6787

1995 IEEE International Frequency Control Symposium

SYMPOSIUM CHAIRMEN

General Chairman.....John R. Vig

Technical Program Chairman.....Lute Maleki

Finance Chairman.....Raymond L. Filler

Awards Chairman.....Roger Ward

Publicity Chairman.....John D. Prestage

Local Arrangements Chairman.....Jack Kusters

Tutorials Coordinator.....David W. Allan

TECHNICAL PROGRAM COMMITTEE VICE CHAIRMEN

Materials & Resonators.....John Kosinski
Oscillators & Circuits.....Victor Reinhardt
Atomic Standards & Timing.....Bernardo Jaduszliwer

TECHNICAL PROGRAM COMMITTEE

David Allan, Allan's TIME
Ivan Avramov, Inst. Solid State Physics, Bulgaria
Arthur Ballato, U.S. Army Research Laboratory
Martin Bloch, Frequency Electronics, Inc.
Jean-Simon Boulanger, NRC, Canada
Jan Brown
Leonard Cutler, Hewlett-Packard Company
Andrea DeMarchi, Politecnico di Torino, Italy
Michael Driscoll, Westinghouse Systems Group
Raymond Filler, U.S. Army Research Laboratory
Marvin Frerking, Rockwell International
Edward Garber, TRW
Michael Garvey, Frequency & Time Systems, Inc.
Helmut Hellwig, Air Force Office of Scientific Research
Bernardo Jaduszliwer, The Aerospace Corporation
Charles Jensik, Piezo Crystal Company
Gary Johnson, Sawyer Research Products, Inc.
Hirofumi Kawashima, Seiko Electronics, Japan
William Klepczynski, U.S. Naval Observatory
John Kosinski, U.S. Army Research Laboratory
Jack Kusters, Hewlett-Packard Company
Theodore Lukaszek, U.S. Army Research Laboratory
Lute Maleki, Jet Propulsion Laboratory
Donald Malocha, University of Central Florida
Dennis Marvin, Motorola
Gary Montress, Raytheon Research Division
Thomas Parker, NIST
John Prestage, Jet Propulsion Laboratory
Victor Reinhardt, Hughes Aircraft Company
Gerald Roberts, General Electric Company
Tadashi Shiosaki, Kyoto University, Japan
Robert Smythe, Piezo Technology, Inc.
Samuel Stein, Timing Solutions Corporation
Daniel Stevens, AT&T Bell Laboratories
Donald Sullivan, NIST
Joseph Suter, Johns Hopkins University
Richard Sydnor, Jet Propulsion Laboratory
John Vig, U.S. Army Research Laboratory
Fred Walls, NIST
Roger Ward, Quartzdyne, Inc.
Werner Weidemann, Ball Efratom Division
Joseph White, Naval Research Laboratory
Gernot Winkler, U.S. Naval Observatory
Lidiya Zhourkina, Russian Committee IEC, Russia

TECHNICAL SESSION CHAIRMEN

PLENARY SESSION
L. Maleki, Jet Propulsion Laboratory

NOISE MODELS AND ANALYSIS
L. Cutler, Hewlett-Packard Laboratories

PIEZO MATERIALS
Y. Pisarevsky, Institute of Crystallography, RAS, Russia

SURFACE TRANSVERSE WAVE DEVICES
J. Kosinski, U.S. Army Research Laboratory

SYNTHESIZERS
V. Reinhardt, Hughes Aircraft Company

ACCELERATION SENSITIVITY OF SURFACE WAVE RESONATORS
G. Montress, Raytheon Research Division

MICRORESONATORS AND SENSORS
J. Brown, Consultant

TRAPPED ENERGY RESONATORS
D. Stevens, AT&T Bell Laboratories

DIODE LASERS AND MINIATURE RUBIDIUM STANDARDS
M. Garvey, Frequency & Time Systems, Inc.

ASPECTS OF ATOMIC FREQUENCY STANDARDS
B. Jaduszliwer, The Aerospace Corporation

PIEZO MATERIALS II
J. Vig, U.S. Army Research Laboratory

TERAHERTZ FREQUENCY SYNTHESIS
F. Walls, National Institute of Standards and Technology

RESONATOR PLATE THEORIES
D. Malocha, University of Central Florida

POSTER SESSION

Time, Frequency & Atomic Standards: *J. White, U.S. Naval Research Laboratory*

Quartz Resonators: *R. Sydnor, Jet Propulsion Laboratory*

SAW Devices: *J. Kusters, Hewlett-Packard Company*

Frequency Synthesis: *R. Smythe, Piezo Technology, Inc.*

CRYSTAL OSCILLATORS
I. Avramov, Institute of Solid State Physics, Bulgaria

BEAM TYPE DEVICES
R. Filler, U.S. Army Research Laboratory

SATELLITE BASED TIME TRANSFER
T. Parker, National Institute of Standards and Technology

CRYSTAL PROCESSING TECHNIQUES
G. Johnson, Sawyer Research Products, Inc.

MILLIMETER WAVE AND MICROWAVE OSCILLATORS
D. Sullivan, National Institute of Standards and Technology

NEURAL NETWORKS
S. Cantor, Mitre Corporation

TUTORIAL TOPICS & PRESENTERS*

"The Mathematics of Quartz - Part 1 - X-Rays"
John A. Kusters, Hewlett-Packard Company

"Introduction to Atomic Clocks"
Robert E. Drullinger, National Institute of Standards and Technology

"Design of SAW Filters for Cellular Phones"
Clemens W. Ruppel, Siemens

"Clocks in Telecommunications"
George P. Zampetti, Telecom Solutions

"Quartz Crystal Material"
Gary Johnson, Sawyer Research Products, Inc.

"Fundamental Concepts and Definitions in AM & PM Noise Metrology"
Eva Ferre-Pikal, National Institute of Standards and Technology

"Discussion of Error Models for PM and AM Noise Measurements"
Franklin G. Ascarrunz, Spectra Dynamics, Inc.

"Introduction to High Stability SAW Oscillators, Design and Performance"
Thomas E. Parker, National Institute of Standards and Technology and
Gary K. Montress, Raytheon Research Institute

"Time-Domain Instability Measures in Time and Frequency and for Telecommunications"
David W. Allan, Allan's TIME

"Introduction to Quartz Resonators as Transducers"
Errol EerNisse, Quartztronics, Inc.

"State-of-the-Art Measurement Techniques for PM and AM Noise"
Craig W. Nelson, Spectra Dynamics, Inc.

"New Time and Frequency Opportunities Using GPS"
David W. Allan, Allan's TIME

PANEL SESSION: "Resonator Fabrication"
Chair: *John R. Vig, U.S. Army Research Laboratory*

PANEL SESSION: "Time and Frequency Transfer (State-of-the-Art and Traditional Techniques)"
Chair: *Thomas E. Parker, National Institute of Standards and Technology*

LECTURE: "Synthesis Techniques (DDS)"
Venceslav F. Kroupa, Institute of Radio Engineering and Electronics,
Academy of Sciences of the Czech Republic

---

* The Tutorial Session was held after the Symposium on Friday afternoon and Saturday morning. It was organized by David Allan, Allan's TIME. Attendees were provided with several hundred pages of handouts. The handouts are not published.

TABLE OF CONTENTS

PAGE

AWARD PRESENTATIONS 1

IN MEMORY OF JOHN F. SILVER 2

NEURAL NETWORKS

Physiological Description of the Neuron and the Human Nervous System..... 3
S.R. Cantor, The MITRE Corporation & S.G. Cantor, Boston University

Neural Networks Trends for Frequency Control: A Review 10
B. Hivert & R. Brendel, Laboratoire de Physique et Metrologie des Oscillateurs du CNRS, France

Evaluation of Oven Controlled Crystal Oscillator Performance Under Non-Continous Stressed Operating Conditions 20
A.M. Gjelsvik, S.R. Cantor & T.E. Ioakimidis, The MITRE Corporation

ATOMIC CLOCKS, SYNTHESIZERS, TIME TRANSFER

An Ultra-Miniature Rubidium Frequency Standard with Two-Cell Scheme..... 33
Y. Koyama, Y. Nakajima, K. Chiba, Fujitsu Limited, H. Matsuura & K. Atsumi, Fujitsu Tohoku Digital Technology Ltd., Japan

Subminiature Rubidium Frequency Standard: Manufacturability and Performance Results from Production Units 39
T. McClelland, I. Pascaru, I. Shtaerman, C. Stone, C. Szekely & J. Zacharski, Frequency Electronics, Inc.

Miniaturized Rubidium Clocks for Space and Industrial Applications..... 53
C. Couplet, G. Mileti, H. Schweda, P. Thomann, G. Busca, Observatoire Cantonal & P. Rochat, Tekelec Neuchâtel Time, Switzerland

Recent Results of the LPTF Cesium Fountain Primary Frequency Standard 60
G. Santarelli, S. Ghezali, Ph. Laurent, S.N. Lea, M. Bahoura, K. Szymaniec, E. Simon & A. Clairon, LPTF/BNM, France

Further Investigation of a Prototype Microwave Frequency Standard Based on Trapped $^{171}Yb^{+}$ Ions 66
M.J. Sellars, P.T.H. Fisk, M.A. Lawn & C. Coles, National Measurement Laboratory, Australia

Development of a Trapped Yb^{+} Frequency Standard: Efficient Quenching of Population Trapping 74
D.J. Seidel & L. Maleki, Jet Propulsion Laboratory

A Mercury Linear Ion Trap Frequency Standard for the USNO..... 79
R.J. Tjoelker, J.D. Prestage & L. Maleki, Jet Propulsion Laboratory

Progress Report on the Improved Linear Ion Trap Physics Package 82
J.D. Prestage, R.L. Tjoelker, G.J. Dick & L. Maleki, Jet Propulsion Laboratory

Eight Years of Experience with Mercury Stored Ion Devices 86
D.N. Matsakis, A.J. Kubik, J.A. DeYoung, U.S. Naval Observatory, R.P. Giffard & L.S. Cutler, Hewlett-Packard Laboratories

A Cryogenic Linear Ion Trap for $^{199}Hg^{+}$ Frequency Standards 110
J.D. Miller, M.E. Poitzsch, F.C. Cruz, D.J. Berkeland, J.C. Bergquist, W.M. Itano & D.J. Wineland, National Institute of Standards & Technology

Systematic Errors in Cesium Beam Frequency Standards Introduced by Digital Control of the Microwave Excitation 113
W.D. Lee, J.H. Shirley, F.L. Walls & R.E. Drullinger, National Institute of Standards & Technology

A Transportable Hydrogen Maser for Australia Telescope 118
J.W. He, C.F. Lin, Z.C. Zhai, Shanghai Observatory, P.R. of China, W.E. Wilson & H. Fagg, Australia Telescope National Facility, Australia

DORIS Precise Orbit Determination and Localization System Description and USOs Performances..... 122
M. Brunet, A. Auriol, P. Agnieray & F. Nouel, Centre National D'Etudes Spatiales, France

Performance of Global Positioning System (GPS) On-Orbit Navstar Clocks 133
T.B. McCaskill, W.G. Reid, O.J. Oaks, R.L. Beard, U.S. Naval Research Laboratory, J.A. Buisson & H.E. Warren, SFA, Inc.

Activities of the Russian Institute of Radionavigation and Time in Development of GLONASS On-Board Q-Enhanced Oscillating Compact H-Maser 140
Yu.G. Gouzhva, A.G. Gevorkyan, A.L. Myasnikov, V.Z. Gankin, O.N. Kornishov & S.D. Teplova, The Russian Institute of Radio-navigation and Time, Russia

Influence of Saturation Beam Parameters on Laser Frequency Locked to the Cesium Cycling Transition 149
R. Gamidov, I. Taskin & V. Sautenkov, UME, Turkey

External Cavity Diode Laser High Resolution Spectroscopy of the Ca and Sr Intercombination Lines for the Development of a Transportable Frequency/ Length Standard 153
A. Celikov, A. Zibrov, V. Velichansky, P.N. Lebedev Physics Institute, Russia, P. Kersten, F. Riehle, G. Zinner, J. Helmcke, Physikalish Technische Bundesunstalt, Germany & L. D'Evelyn, National Institute of Standards & Technology

Characteristics and Performance of a Novel Photonic Oscillator 161
S. Yao & L. Maleki, Jet Propulsion Laboratory

New Tests for Variations of the Fine Structure Constant *(INVITED)* 169
J.D. Prestage, R.L. Tjoelker & L. Maleki, Jet Propulsion Laboratory

Experimental Test of Local Oscillator-Induced Performance Limitation for Pulse-Mode Passive Frequency Standards 175
G.J. Dick & D.A. Stowers, Jet Propulsion Laboratory

Reducing the Effect of Interrogation Phase Noise on the Frequency Stability in Passive Frequency Standards: Experimental Results and New Investigation Means 178
R. Barillet, D. Venot & C. Audoin, Laboratoire de L'Horloge Atomique, France

Diode Lasers for Frequency Standards and Precision Spectroscopy *(INVITED)* 185
L. Hollberg, S. Waltman, J.H. Marquardt, M. Stephens, R.W. Fox, C.S. Weimer, H.G. Robinson, N. Mackie, National Institute of Standards & Technology, J. Aman, Chalmers University of Technology, Sweden, D.A. Van Baak, Calvin College, A.S. Zibrov, T.P Zibrova, Lebedev Institute of Physics, Russia & L. Pendrill, Swedish Testing and Research Institute, Sweden

W-Band MMIC-Based Component Development and Applications *(INVITED)* 190
H. Wang, K-W. Chang, D.C-W. Lo & G.S. Dow, TRW

Harmonic Generation and Mixing in High-T_C Josephson Junctions with Terahertz Bandwidth 205
E.N Grossman, L.R. Vale, L.R. Zink & K.M. Evenson, National Institute of Standards and Technology

A High Purity, High Speed Direct Digital Synthesizer 207
G.W. Kent & N-H. Sheng, Rockwell International Corporation

Digital Frequency-Lock Loop Using High-Resolution TAC and Advanced Control Algorithm for Trapped-Ion Frequency Standards 212
D.A. Stowers, R.L. Sydnor, R.L. Hamell, M.J. Grimm & G.J. Dick, Jet Propulsion Laboratory

A Narrow Band High-Resolution Synthesizer Using a Direct Digital Synthesizer Followed by Repeated Dividing and Mixing 217
R. Karlquist, Hewlett-Packard Company

Narrow-Band Direct Digital Synthesizers 236
V.N. Kochemasov & A.N. Zharov, Moscow Technical University of Communications and Informatics, Russia

A High-Speed Phase-Frequency Detector ASIC with Programmable Dividers for LO Generator and Indirect Synthesis Applications 250
R. Howald & T. Marra, Lockheed-Martin Astro-Space

GPS/LORAN in an Urban Environment-Oscillator Stability Considerations 259
B. Peterson, R.J. Hartnett, G. Ottman, U.S. Coast Guard Academy & L. Miller, Science Applications International Corporation

Prediction of the Time Accuracy and Integrity of GPS Timing 266
G.J. Geier, T.M. King, H.L. Kennedy, R.D. Thomas & B.R. McNamara, Motorola Position and Navigation Systems Business

Common-Clock Two-Way Satellite Time Transfer Experiments 275
C. Hackman, S.R. Jefferts & T.E. Parker, National Institute of Standards and Technology

NOISE AND OSCILLATORS

1/f Fluctuations Sources in Direct Digital Frequency Synthesizers and Their Contribution to the Output Oscillations Power Spectral Density 282
V.N. Kuleshov, H.Y. Liu & B.E. Leshukov, Moscow Power Engineering Institute, Russia

Fundamental Noise in Direct Digital Frequency Synthesizers 288
V.N. Kuleshov & H.Y. Liu, Moscow Power Engineering Institute, Russia

The Origin of 1/f PM and AM Noise in Bipolar Junction Transistor Amplifiers *(INVITED)* 294
F.L. Walls, E.S. Ferre-Pikal & S.R. Jefferts, National Institute of Standards and Technology

Design Criteria for BJT Amplifiers with Low 1/f AM and PM Noise..... 305
E.S. Ferre-Pikal, F.L. Walls, National Institute of Standards and Technology & C.W. Nelson, Spectradynamics, Inc.

Advanced Phase Noise Suppression Technique for Next Generation of Ultra Low-Noise Microwave Oscillators..... 314
E.N. Ivanov, M.E. Tobar & R.A. Woode, University of Western Australia, Australia

An Extension of the Allan Variance with Increased Confidence at Long Term ... 321
D.A. Howe, National Institute of Standards and Technology

Cut-Off Frequencies and Noise Power Law Model of Spectral Density: Adaptation of the Multi-Variance Method for Irregularly Spaced Timing Data Using the Lowest Mode Estimator Approach 330
F. Vernotte, G. Zalamansky, M. McHugh, Observatoire de Besancon & E. Lantz, Laboratoire d'Optique P.M. Duffieux, France

Circular Representation of Infinitely Extended Sequences..... 337
D.A. Howe, National Institute of Standards & Technology

Estimating the Modified Allan Variance 346
C.A. Greenhall, Jet Propulsion Laboratory

Some New Developments of Precision Frequency Measurement Technique 354
W. Zhou, Z. Xuan & J. Yu, Xidian University, P.R. China

A Test Suite for the Calculation of Time Domain Frequency Stability 360
W.J. Riley, EG&G, Inc.

Measurement of Voltage Noise in Chemical Batteries 367
C.K. Boggs, A.D. Doak & F.L. Walls, National Institute of Standards & Technology

Transposed Gain Microwave Oscillators with Low Residual Flicker Noise..... 374
J.K.A. Everard & M.A. Page-Jones, University of York, England

Designing SAW Resonators and DRO Oscillators Using Nonlinear CAD Tools ... 379
U.L. Rohde, Compact Software, Inc.

Improved Performance of a Temperature Compensated Ln_2 Cooled Sapphire Oscillator..... 397
D.G. Santiago, R.T. Wang & G.J. Dick, Jet Propulsion Laboratory

Spectral Performance of Sapphire Dielectric Resonator-Controlled Oscillators Operating in the 80K to 275K Temperature Range..... 401
M.M. Driscoll, Westinghouse Electronics Systems Group &
R.W. Weinert, Westinghouse Science & Technology Center

A New Method of Frequency Stabilization 413
D.P. Tsarapkin, Moscow Power Engineering Institute, Russia

An Ultra-Low Noise Microwave Oscillator Based on a High-Q Liquid Nitrogen Cooled Sapphire Resonator 420
R.A. Woode, M.E. Tobar & E.N. Ivanov, University of Western Australia, Australia

A Non-Linear CAD Model for the Regenerative Dielectric Resonator Oscillator..... 427
A.A. Behagi, Penn State University & S.D. Turner, MAXTECH, Inc.

Latest Results of the U.W.A. Cryogenic Sapphire Oscillator 433
A.N. Luiten, A.G. Mann, N.J. McDonald & D.G. Blair, University of Western Australia, Australia

1 GHz STW Based Oscillator with Continuous Temperature Compensation..... 438
M.A. Taslakov, Institute of Electronics, Bulgaria

Precise Coupling-of-Modes Simulation of Surface Transverse Wave Devices for Frequency Control Applications 442
K-Y. Hashimoto & M. Yamaguchi, Chiba University, Japan

A Phenomenological Coupling-of-Modes Formalism for Surface Transverse Wave Devices..... 452
B.P. Abbott, RF Monolithics, Inc. & K-Y. Hashimoto, Chiba University, Japan

2 GHz Surface Transverse Wave Oscillators. Design, Performance and Limitations 459
I.D. Avramov, Institute of Solid State Physics, Bulgaria, V.S. Aliev, S. Denissenko & A.S. Kozlov, Institute of Semiconductor Physics, Russia

Design and Test of 3 GHz, Fundamental Mode STW Resonators on Quartz 469
S. Denissenko, Institute of Semiconductor Physics, Russia,
E. Gavignet, S. Ballandras, E. Bigler, Laboratoire de Physique et Métrologie des Oscillateurs du CNRS, France & E. Cambril, Laboratoire de Microstructures et Microélectronique du CNRS, France

An Oscillator Design Using Low g-Sensitivity, Low Phase Noise STW Devices 476
M.S. Cavin & R.C. Almar, Sawtek, Inc.

An Analysis of the Normal Acceleration Sensitivity of Rotated Y-Cut Quartz Surface Transverse Wave Resonators Simply Supported Along Rectangular Edges 486
J.A. Kosinski, J.T. Stewart, A. Ballato, U.S. Army Research Laboratory & R. Almar, Sawtek, Inc.

Experimental Studies of SAW and STW Acceleration Sensitivity 494
D. Huynh, R. McGowan, J.A. Kosinski, J.T. Stewart, R. Pierkarz & C.D. Mulford, U.S. Army Research Laboratory

Semi-Analytical Finite Element Analysis of Acceleration-Induced Frequency Change in SAW Resonators 499
J.T. Stewart, R. McGowan, J.A. Kosinski & A. Ballato, U.S. Army Research Laboratory

An Analysis of the Dynamic Behavior and Acceleration Sensitivity of a SAW Resonator Supported by Flexible Beams 507
J.T. Stewart, J.A. Kosinski & A. Ballato, U.S. Army Research Laboratory

A SAWR Oscillator Vibration Sensitivity and Phase Noise Reduction Technique Using Multiple Resonators and RF Outputs 514
M.M. Driscoll, Westinghouse Electronic Systems Group

Acceleration Charge Sensitivity in AT-Quartz Resonators 519
C.L. Anderson, Motorola, J.S. Bagby, R.L. Barrett & V. Ungvichian, Florida Atlantic University

Surface Acoustic Wave Propagation on Piezoelectric Substrates with Thin-Film PECVD Silicon Nitride 530
F.S. Hickernell, H.D. Knuth & T.S. Hickernell, Motorola Government and Space Technology Group & L. Mang, Motorola Phoenix Corporate Research Laboratories

Low-Loss SAW Filters Using Thickness Difference Type of IDT on the NSPUDT Orientation Substrate 537
K. Yamanouchi, M. Takeuchi, H. Odagawa & M. Tanaka, Tohoku University, Japan

Modulation Type Dual-Mode Oscillator Intended for Micro-Chip Realization 542
A.V. Kosykh & S.A. Zavyalov, Omsk State Engineering University, Russia

Design of a 100KHz to 70MHz, 2.7 to 7 Volt, Clocking Oscillator Chip 546
H. Rokos, C-MAC Quartz Crystals Ltd., England

CMOS Crystal Oscillators Based on Current Conveyors 548
J. Popovic, D. Vasiljevic, University of Belgrade, A. Pavasovic, Lola Institute & Z. Zivkovic-Dzunja, Faculty of Technical Sciences, Yugoslavia

A Single Bridge Resonator for Low Consumption OCXO 557
S. Galliou, B. Dulmet & M. Mourey, Ecole Nationale Supérieure de Mécanique et des Microtechniques, France

Fast Start-Up Crystal Oscillator Circuits 565
Y. Tsuzuki, T. Adachi & J.W. Zhang, Yokohama National University, Japan

Universal, Computer Facilitated, Steady State Oscillator, Closed Loop Analysis Theory 569
B. Parzen, Consulting Engineer

DTXCO Thermal Converter Realized by Means of a Metal Film Temperature Sensor Deposited on the Crystal Plate 574
V.P. Bagayev, B.P. Ionov & S.A. Zavyalov, Omsk State Engineering University, Russia

The Modulational Method of the Precision Quartz-Crystal Oscillators and Standards Frequency Stabilization 579
Yu.S. Shmaly, Kharkiv Military University & "Sichron" Centre, Ukraine

BVA Resonators and Oscillators: A Review. Relation with Space Requirements and Quartz Material Characterization 590
R.J. Besson, J-J. Boy & M.M. Mourey, Ecole Nationale Supérieure de Mécanique et des Microtechniques, France

Satellite Master Oscillators 600
D.A. Emmons, D.W. Martin & R.M. Garvey, Frequency & Time Systems, Inc.

Advanced Technology Oscillator for Small Spacecraft 614
J.R. Norton, Johns Hopkins University

An Experimental Study of Long Term Aging of Quartz Oscillators 620
M. Koyama, Nihon Dempa Kogyo Co. Ltd., Y. Watanabe, H. Sekimoto & Y. Oomura, Tokyo Metropolitan University, Japan

Design and Fabrication of a Capacitance Switch Array Usable on a DTCXO 623
W. Daniau, G. Marianneau, LPMO/CNRS, E. Girardet & J. Marguier, AR Electronique, France

PIEZOELECTRIC MATERIALS, RESONATORS, SENSORS

New Results on High Perfection Langasite Crystals: Studies of Crystalline Defects and Modes Shapes 629
A. Zarka, B. Capelle, CNRS, J. Detaint, CNET & D. Cochet-Muchy, CRISMATEC, France

The Nature of Langasite Crystal's Coloration 638
M.F. Dubovik, K.A. Katrunov & T.I. Korshikova, Institute for Single Crystals, Ukraine

Surface and Volume Defects in Langasite Crystals 642
S.A. Sakharov, A.V. Medvedev, V. Lider, FOCO Co., Yu. Pisarevsky & P.A. Senushencov, Institute of Crystallography RAS, Russia

New Data on Temperature Stability and Acoustical Losses of Langasite Crystals 647
S. Sakharov, A. Medvedev, FOCO Co., P. Senushencov & Yu. Pisarevsky, Institute of Crystallography RAS, Russia

New Strong Piezoelectric $La_3Ga_{5.5}Nb_{0.5}O_{14}$ with Temperature Compensation Cuts 653
Yu. Pisarevsky, P.A. Senushencov, P.A. Popov & B.V. Mill, Institute of Crystallography RAS, Russia

Langasite Crystal Quality Improvement Aimed at High-Q Resonators Fabrication 657
A.N. Gotalskaja, D.I. Dresin, S.N. Schegolkova, N.I. Saveleva, V.V. Bezdelkin & G.N. Cherpoukhina, "LANTAN" Co. Ltd., Russia

Bridgman Growth of $Li_2B_4O_7$ Crystal of 4-Inch Diameter 667
S-J. Fan, J. Wang & Y-B. Xu, Shanghai Institute of Ceramics, P.R. China

The Anomalous Elastic Anisotropy of $Li_2B_4O_7$ and its Influence on SAW Properties 671
N.F. Naumenko, Moscow Steel & Alloys Institute, Russia

Synchrotron Radiation Topography of Thick Swept Quartz 679
C.K. Suzuki, University of Campinas, Brazil, X.W. Zhang, M. Ando, National Laboratory for High Energy Physics, Y. Yoda, S. Kikuta, University of Tokyo, K. Hamaguchi, K. Nagai & S. Taki, Toyo Communications Equipment Co. Ltd., Japan

Specific Features of Quartz Crystals Lamellar Structure 685
N.M. Bauer, A.P. Pogrebnyak, S.N. Abdrafikov, N.A. Mamaev & B.V. Shulgin, The Plant "Kristall", Russia

Thermal Desorption of Quartz Crystals 696
R.A. Murray, D.E. Pierce, R. Lareau, J.R. Vig, U.S. Army Research Laboratory & S. Laffey, Vitronics, Inc.

Piezoelectrically Forced Thickness-Shear and Flexural Vibrations of Contoured Quartz Resonators 705
P.C.Y. Lee & J. Wang, Princeton University

A New Model for the Thickness Shear Resonators with Spherical Contours: Application to Bevelled Devices 716
J. Detaint, France Telecom, B. Capelle, A. Zarka & Y. Zheng, LMCP/CNRS, France

An Analysis of Contoured Quartz Resonators with Beveled Cylindrical Edges 727
J.S. Yang & H.F. Tiersten, Rensselaer Polytechnic Institute

The Influence of an Asymmetric Air Gap on the Vibration Characteristics of Quartz Trapped Energy Resonators 740
H.F. Tiersten, Rensselaer Polytechnic Institute

Lee's Plate Equations with New Correction Factors and Analysis of Spurious Vibrations of Rectangular AT-Cut Quartz Plates 746
H. Sekimoto, D. Tajima, Y. Watanabe & A. Ishizaki, Tokyo Metropolitan University, Japan

Eigenfunction Expansions in Studies of Multimode and Transient Vibrations of Piezoelectric Transducers 750
O.Yu. Zharii, Kiev University, Ukraine

On the Accuracy of Plate Theories for the Prediction of Unwanted Modes Near the Fundamental Thickness Shear Mode 755
Y-K. Yong, Z. Zhang, Rutgers University & J. Hou, Allied Signal, Inc.

Two-Dimensional Analysis Using 1D-FEM for the Straight-Crested Waves in Arbitrary Anisotropic Crystal Plates and the Axisymmetric Piezoelectric Vibrations in the Ceramic Disks 761
A. Ishizaki & H. Sekimoto, Tokyo Metropolitan University, Japan

Electromagnetic Waves in an Infinite Isotropic Dielectric Plate Enclosed by Two Parallel Perfect Conductors 769
P.C.Y. Lee & J.D. Yu, Princeton University

The Shear Coefficient for Quartz Crystal of Rectangular Cross Section in Timoshenko's Beam Theory 778
H. Kawashima, Seiko Instruments Inc., Japan

Mathematical Theory of the Fork-Type Wave Gyroscope 786
I.A. Ulitko, Kiev University, Ukraine

Dynamic Bimorph Matrix of End-Loaded Bimorphs 794
J. Goli, J.G. Smits, Boston University & A. Ballato, U.S. Army Research Laboratory

Torsional Vibrations of Quartz Crystal Beams 798
H. Kawashima & K. Sunaga, Seiko Instruments Inc., Japan

Dependence of Frequency-Temperature Characteristics of Quartz Resonators on Form, Dimensions and Local Variations of Thickness of Piezoelectric Crystal 804
I.I. Postnikov, Moscow Power Engineering Institute, Russia

Intensity of Resonance Vibration of AT-Cut Quartz Resonators with Circular Electrodes 810
I.I. Postnikov & T.V. Christophorova, Moscow Power Engineering Institute, Russia

High Frequency Fundamental Resonators and Filters Fabricated by Batch Process Using Chemical Etching 818
O. Ishii, T. Morita, T. Saito & Y. Nakazawa, Toyo Communications Equipment Co. Ltd., Japan

Self Limiting Etching of Piezoelectric Crystals 827
K.M. Lakin, G.R. Kline & K.T. McCarron, TFR Technologies, Inc.

Experiments to Minimize Frequency Shifts and Yield Losses for Surface Mount Filters 832
I. Alhayek & J. Voissem, Motorola

Internal Heated Quartz Resonator with Low Sensitivity to an Acceleration 838
I.V. Abramson, Omsk Research Instrumentation Institute, Russia

Frequency-Temperature Characteristics of Quartz Crystal Units of Different Cuts Operating Over a Wide Temperature Range Including Helium Temperatures 843
A.G. Smagin, Crystal Co., Russia

A High Sensitivity Mechanical Loss Spectrometer 848
R.A. Kant & C. Carosella, Naval Research Laboratory

Microresonator Sensor Arrays 852
J.R. Vig, R.L. Filler, U.S. Army Research Laboratory & Y. Kim, National Research Council

Direct Simulation of Next Generation Devices by Atomistic Simulation 870
C. Meli, J. Broughton, Naval Research Laboratory, P. Vashishta & R. Kalia, Louisiana State University

A Smart H_2 Sensor Array of QCMs Coated With PdNi Alloy of Various Thicknesses 876
G. Liu, M.D. Porter, G.J. Bastiaans & S. Burns, Iowa State University

A Micromachined Sensor Array Using Thin Film Resonators 879
J. Xia, S. Burns, M. Porter, T. Xue, G. Liu, R. Wyse & C. Thielen, Iowa State University

Spatial Frequency-Based Microbending Fiber Optic Sensor 885
J.A. Brosig, Motorola

Aspects of Activity of IEC Technical Committee No. 49 "Piezoelectric and Dielectric Devices for Frequency Control and Selection" 894
L.I. Zhourkina, Joint Stock Company "Phonon", Russia

POST DEADLINE SUBMISSION

Monofrequency Control of Low-Frequency Quartz Resonators by Holographic Interferometry 903
P. Borovtsov, Perm State Technical University, Russia

PROCEEDINGS ORDERING INFORMATION 908

SPECIFICATIONS AND STANDARDS RELATED TO FREQUENCY CONTROL 909

AUTHOR INDEX 913

1995 Award Winners

W. G. Cady Award - The Cady Award was presented to Jean-Jacques Gagnepain, LPMO-CNRS, France, "for major contributions to our understanding of nonlinear and environmental effects, and pioneering results on 1/f noise in piezoelectric resonators." The award was presented by Errol P. EerNisse, Quartzdyne.

I. I. Rabi Award - The Rabi Award was presented to Fred L. Walls, NIST, "...for major contributions to the characterization of noise and other instabilities of local oscillators and their effects on atomic frequency standards." The award was presented by Samuel R. Stein, Timing Solutions.

C. B. Sawyer Award - The Sawyer Award was presented to Lidya I. Zhourkina, Scientific Res. Inst. Phonon, Russia, "for outstanding contributions to international standardization, and 30 years of leadership on the IEC Technical Committee No. 49, "Piezoelectric and Dielectric Devices for Frequency Control and Selection." The award was presented by Gary R. Johnson, Sawyer Research Products.

1995 Lifetime Achievement Award - The 1995 Lifetime Achievement Award was presented to Arthur W. Warner, Jr. "for his nearly 60 years of pioneering and wide-ranging contributions to the design and fabrication of high-stability quartz resonators." The award was presented by John R. Vig, US Army Research Laboratory.



Jean-Jacques Gagnepain, Lidya I. Zhourkina, Fred L. Walls, and Arthur W. Warner, Jr.

JOHN FRANKLIN SILVER
1911-1990

Dr. Virgil E. Bottom
3442 S. Willis
Abilene, TX 79605

John Franklin Silver died on December 20, 1990, just 21 days short of his eightieth birthday. With his death the Engineering profession lost a valuable colleague, the business world a competent manager, and the QCU (Quartz Crystal Unit) industry a true friend.

John was born January 11, 1911, in Cincinnati, Ohio. He received the degree of Bachelor of Arts in Engineering from Ohio State University in 1935. The same year he was married to Miss Thelma Cooper. They were married more than 55 years.

His first employment was with the Crosley Radio Company where he was responsible for the triple-tuned I-F transformer which was a feature of Crosley receivers. In 1941 he transferred to the Collins company and later to the Motorola Corporation where he worked on problems associated with two-way radio communication systems. He remained with Motorola from 1945 to 1954 during which he sponsored a program at Colorado State University resulting in the development of the IT-cut piezoid to replace the AT-cut in transmitters operating at temperatures above 70 degrees C. This was the first doubly rotated thickness-shear unit and the predecessor of the SC-cut. For a short time he was VP of the Motorola Corporation.

He earned the degree of Master of Arts in Business from the University of Chicago in 1947 and it was the union of Electrical Engineering and Business Management which most characterizes the professional career of John Silver.

In 1966 John became President of the James Knights Company of Sandwich, Illinois. He remained in this position for 14 years during which J-K became the largest manufacturer of QCUs in the United States, if not in the entire world. John realized that the Quartz Crystal business lacked strong technical support and assisted educational programs at Northern Illinois University, New Mexico State University, and McMurry College in Abilene, Texas with equipment and money for student grants. He then offered summer employment to these students, many of whom became permanent employees of J-K and other QCU companies. Some of the leaders in the QCU industry began their careers in the educational programs begun by John Silver at the J-K Company in Sandwich, Illinois.

But John did not devote all of his energy to engineering and business. He was active in civic and church programs and in Boy Scout activities. He served on the School Board of Park Ridge Elementary School and was its president in 1952. He supported cooperative programs for students of Electrical Engineering and strove to make Engineering a "true profession"; as essential and recognizable as Law and Medicine. He was opposed to anything which tended to diminish the professional ideals of Engineering and did everything in his power to enhance the prestige of the profession. He is greatly missed.



1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PHYSIOLOGICAL DESCRIPTION OF THE NEURON AND THE HUMAN NERVOUS SYSTEM

Stephen R. Cantor*, Shira G. Cantor**

*The MITRE Corporation, 202 Burlington Rd., Bedford, MA 01730
**Sargent College of Allied Health Professions, Boston University, Boston, MA 02215

*/** The views and opinions expressed in this paper are those of the authors, and do not reflect those of MITRE nor of Boston University.

Abstract

In order to maintain equilibrium, the human body is constantly reacting to and adjusting to changes in the outside environment, as well as within the body (internal environment). Such environmental changes, or *stimuli*, initiate impulses in sensory receptors. These impulses (or inputs) pass via the nerves to the brain and spinal cord, where they are analyzed, combined, compared, and coordinated. After being sorted out, impulses are conveyed by nerves to the muscles and glands of the body. The nervous system expresses itself through muscles and glands, causing muscles to contract or relax, and glands to secrete or not secrete their products.

Stimuli received by the nervous system are processed rapidly through a combination of electrical impulses and chemical substances called *neurotransmitters* for communication between a nerve cell and a muscle cell, or between a nerve cell and glandular cells.

Although the nervous system is a single, unified communications network, it is divided on a gross anatomical basis into the central nervous system and the peripheral nervous system. The central nervous system consists of the brain and spinal cord; the peripheral nervous system includes all the nerve cells and fibers that lie outside the brain and spinal cord. The peripheral nervous system connects the brain and spinal cord with the rest of the body.

Neurons And Their Functions

The nervous system of the human body contains over 100 billion nerve cells, called neurons. Neurons are specialized to transmit impulses from short to relatively long distances, from one part of the body or nervous system to another. They have two important properties: (1) *excitability*, which is the ability to respond to stimuli, and (2) *conductivity*, which is the ability to conduct a signal. Neurons are very specialized types of cells. Although they vary in size and shape, all neurons contain three principal parts, each associated with a specific function: the cell body, dendrites, and an axon, as shown in Figure 1.

Cell Body

The cell body of a neuron is also called a *soma*. It may be star-shaped, roundish, oval, or even pyramid-shaped; but its distinguishing structural features are its complex, spreading processes (branches or fibers) that pass on to or receive impulses from other cells.

Dendrites and axons

Neurons have two types of processes, dendrites and axons. The typical neuron has many short, threadlike branches called *dendrites*, which are actually extensions of the cell body. Dendrites conduct nerve impulses *toward* the cell body. A neuron may have as many as 200 dendrites.

A neuron generally has just one *axon*, which is a slender process that extends from the cell body from less than a millimeter (as in the brain) to more than a meter (as in the axons of the sciatic nerve, which extend from the spinal cord to the foot). An axon carries nerve impulses *away* from the cell body to the next nerve cell, muscle cell, or gland cell. The axon may have some side

0-7803-2500-1/95 $4.00 © 1995 IEEE

branches, called *collateral branches*, that leave the main axon. The axon and its collateral branches end in a spray of tiny branches called *telodendria*. The branches of the telodendria usually have tiny swellings called end bulbs or synaptic boutons. A *synapse* (or connection) is formed where an end bulb associates with the plasma membrane of another neutron.

Nerves

A nerve is simply a bundle of fibers enclosed in a connective tissue sheath, like many telephone wires in a cable.

Types of Neurons: Based on Function

The neurons of the peripheral nervous system may be classified according to the direction in which they transmit nerve impulses. Neurons conveying information from sensory receptors in the body *to* the central nervous system are called *afferent*, or sensory, neurons. Afferent neurons, whose cell bodies are in clusters, located close to the central nervous system, also carry impulses from lower to higher centers in the central nervous system. The distal dendrites of these afferent fibers are the *sensory receptors* (or are connected to the sensory receptors), which are responsive to internal or external stimulation.

Efferent or motor, neurons or fibers convey nerve impulses from higher to lower centers in the central nervous system, or *away from* the central nervous system to the effectors (muscles and glands), where the response actually takes place. Most cranial and spinal nerves (nerves connected to the brain and spinal cord) are called *mixed nerves* because they are composed of hundreds, or even thousands, of afferent *and* efferent nerve fibers.

Interneurons lie mostly within the central nervous system, but a few are distributed between the central and peripheral nervous systems. Their functions are to carry impulses from sensory neurons to motor neurons and to process incoming neural information. Interneurons with long axons, called *relay neurons*, convey signals over long distances. Interneurons with short, usually branching, axons are called *local circuit neurons*. They convey signals locally, over short distances. Local circuit neurons play an important role in learning, emotions, and language. Complex functions such as learning and memory may depend on thousands of local circuit neurons. Less complicated activities, such as the simpler reflexes, may involve only a few or no interneurons, since sensory and motor neurons may be directly connected.

Types of Neurons: Based on Structure

Neurons may also be classified according to their structure, or more specifically, according to the number of their processes. *Multipolar neurons* generally have many dendrites radiating from the cell body, but only one axon. Most of the neurons of the brain and spinal cord are multipolar. *Bipolar neurons* have only two processes. Essentially, all neurons in the adult develop from bipolar cells, which are the neurons of the embryonic nervous system. In the adult, bipolar neurons are located in only a few structures, including the retina of the eye, cochlear and vestibular nerves of the ear, and the olfactory nerve in the upper nasal cavity. *Unipolar neurons* are the most common sensory neurons in the peripheral nervous system. The single process of the unipolar neuron divides into two branches. One branch extends into the brain or spinal cord, conducting nerve impulses away from the cell body, and the other branch extends to a peripheral sensory receptor in some distal part of the body.

For all their structural differences, most neurons are composed of four distinct functional segments:

1. The *receptive segment* of a multipolar neuron is composed of the cell body and its dendrites. It is the segment that receives a continuous bombardment of synaptic impulses from numerous other neurons on the receptor sites on its plasma membrane. These complex

inputs are processed in the receptive segment, and if sufficiently stimulated, the resolution of this processing is conveyed to the initial segment at the junction of the cell body and axon hillock.

2. The *initial segment* is the trigger zone of the neuron, where the processed neural information of the receptive segment is converted into a nerve impulse. As a result, this segment, with its low threshold, is critical for the generation of a nerve impulse.

3. The *conductive segment* (the axon) is specialized for the conduction of neural information over relatively long distances in the body. It acts to convey results of the neural processing of the receptive segment via nerve impulses to the terminal transmissive segment.

4. The *transmissive* (effector) *segment* contains the axon terminals that convert the stimulation of the nerve impulse to release chemical neurotransmitters at its synapses. These chemicals exert influences upon the receptor sites on an effector cell (another neuron, a muscle cell, or a glandular cell).

Physiology of Neurons

Neurons are specialized to respond to stimulation. They do this by generating electrical impulses. These impulses are expressed as changes in the electrical potentials conducted along the plasma membranes of the dendrites, cell body, and axon of each neuron. The difference in potential across the plasma membrane of the neuron results from differences in the concentration of certain ions on either side of the membrane.

The plasma membrane of a neuron contains proteins that are specialized as *channels* and *pumps*. The *channels* act as pores through which ions can pass (for example, sodium ion channels). Associated with the channels are *receptor sites*, which open or close the channels in response to stimuli.

A "resting" neuron is one that is not conducting a nerve impulse, but is electrically charged. The plasma membrane of such a "resting" neuron is said to be *polarized*, meaning that there is a difference in electrical charge between the inside and the outside of the membrane. The intercellular fluid on the inner side of the membrane is negatively charged in relation to the positively charged extracellular fluid outside the membrane. The difference in the electrical charge across the membrane in a "resting" neuron is called the *resting membrane potential*. Normally, the resting electrical potential is about -70 millivolts, and is due to the unequal distribution of various electrically charged ions (sodium, potassium, chloride, and protein molecules) and to the selective permeability of ion channels in the plasma membrane to the different ions (upper right portion of Figure 1).

Na^+ and K^+ ions constantly diffuse through the ion channels of the plasma membrane, moving from regions of high concentration to regions of lower concentration. Yet their relative concentration remain remarkably constant. The extracellular fluid usually contains about 10 Na^+ ions for every K^+ ion. Within the cell the ratio is reversed, with K^+ ions outnumbering Na^+ ions by at least 10 to 1. Such ionic consistency is maintained by a self-regulating mechanism located within the plasma membrane, known as the *sodium-potassium pump*. The more sodium that leaks into the neuron through the plasma membrane, the more active the pump becomes, pumping actively to restore the ionic concentrations that maintain the resting membrane potential. The sodium-potassium pump transports three sodium ions out of the cell for every two potassium ions it brings into the cell. So even though both ions are positive, more positive ions are moving out through the membrane than are moving in. Thus, the inner side of the cell is more negative than the outer side.

The Mechanism of Nerve Impulses

The *change* in the electrical potential difference across a plasma membrane is the key factor in the initiation and subsequent conduction of a nerve impulse, which occurs as follows:

1. A stimulus that is strong enough to initiate an impulse in a neuron is called a *threshold stimulus*. When such a stimulus is applied to a polarized resting membrane of an axon, sodium ion channels into the cell open, and sodium ions rush in, reversing the electrical charge at the point of stimulus (In the 1 millisecond that a channel is open, about 20,000 sodium ions flow through). Thus, at the point of stimulus, the *inside* of the membrane becomes positively charged relative to the *outside*, a condition known as *depolarization*. When a stimulus is strong enough to cause depolarization, the neuron is said to *fire* (upper right portion of Figure 1)

2. Once a small area on the neuron is depolarized, it stimulates the adjacent area, and an *action potential*, or *nerve impulse*, is initiated and conducted along the plasma membrane.

3. Shortly after depolarization, the original balance of sodium and potassium ions is restored by the action of the membrane pumps; the relative electrical charges inside and outside the membrane are also restored. The membrane is then said to be *repolarized*. The transmission of a nerve impulse along the plasma membrane may be visualized as a *wave* of depolarization and repolarization.

4. After each firing, there is an interval of from 0.5 to 1 millisecond before it is possible for an adequate stimulus to generate another action potential. This period is called the *refractory period*. Most nerve fibers are capable of generating about 300 impulses per second.

All-or-None Law

A stimulus of a minimum strength is necessary to initiate a nerve impulse, but an increase in stimulus strength above the minimum does not increase the strength of the impulse. The principle that states that a neuron will fire at full power or not at all is known as the *all-or-none law.*

How is the difference we feel between a light touch and a strong one, or can distinguish between a soft sound and a loud one, explained by the all-or-nothing law? Such differences can be perceived when the *frequency* of the impulses, not their strength, is changed. Some nerve fibers can conduct at different frequencies. The more frequent the impulses, the higher the level of excitation. Other fibers respond to *different thresholds*, but all impulses carried on any given fiber are of the same strength. Also, the *number of neurons* can make a difference in how strong the stimulus is perceived to be. A person feels the difference between a light push and a strong one, for example, because the strong push affects more of his neurons.

Synapses

A nerve impulse travels along an axon and eventually reaches the branching axon terminals in the transmissive segment of the neuron. If the impulse is to be effective, it must be conveyed to another neuron, a muscle cell, or gland cell.

Let us consider the junction between the axon terminal of one neuron and the dendrite, cell body, or axon of the next neuron. This junction between neurons is called a *synapse*. The neuron carrying the impulse *toward* a synapse is called the *presynaptic neuron.* It initiates a response in the receptive segment of a *postsynaptic neuron* leading *away from* the synapse. The presynaptic cell is almost always a neuron, but the

postsynaptic cell can be a neuron, muscle cell, or gland cell.

The transmission of a nerve impulse at a synapse can be either chemical or electrical, but chemical synapses are far more common than electrical ones.

In a *chemical synapse*, two cells communicate by way of a chemical agent called a *neurotransmitter*, which is released by the transmissive segment of the presynaptic neuron (lower left portion of Figure 1). A neurotransmitter is capable of changing the resting potential in the plasma membrane of the receptive segment of the postsynaptic cell.

When the nerve impulse of a presynaptic neuron reaches the end bulb of the axon, it depolarizes the presynaptic plasma membrane, causing voltage-gated channels to open. The structure of the synapse is such that impulses can travel in only one direction in the nervous system. The transmitters can pass only from axon terminals across the synaptic cleft to the receptor cell, but not the other way. This directional control prevents neural effects from traveling in all directions at once. If they did, neural messages would be garbled and scrambled.

The short trip across the synaptic cleft takes less than a millisecond. That interval results in a *synaptic delay*. Some of the released neurotransmitter binds with *receptor protein sites* in the postsynaptic membrane, causing changes in the membrane's permeability to certain ions. As a result of the change in permeability in the postsynaptic membrane, sodium and potassium ions rush through their open ion channels and cause the same kind of depolarization that the presynaptic cell experienced. A wave of depolarization and repolarization passes along the postsynaptic neuron, and the nerve impulse continues along its path to an effector.

The effect of a neurotransmitter on the postsynaptic membrane may be to produce either an excitatory or inhibitory response.

Learning occurs by changes that take place in the effectiveness of the synapses, so that the influence of one neuron on another changes.

Organization of the Nervous System

The brain and spinal cord make up the *central nervous system*, and all nerve cells outside the brain and spinal column make up *the peripheral nervous system*. Cranial nerves (which are connected to the brain) and spinal nerves (which are connected to the spinal cord), are part of the peripheral nervous system. The central nervous system may be thought of as the body's central control system, receiving and interpreting or integrating all stimuli, and relaying nerve impulses to muscles and glands, where the designated actions actually take place. The nerve cells and their fibers that make up the peripheral nervous system allow the brain and spinal column to communicate with the rest of the body. In terms of function, two types of nerve cells are present in the peripheral nervous system: (1) *afferent*, or sensory, nerve cells carry nerve impulses from sensory receptors in the body *to* the central nervous system, where the information is processed; (2) *efferent*, or motor, nerve cells convey information *away from* the central nervous system to the effectors (muscles and glands).

Neuronal Circuits

The nervous system is a structured network of neurons arranged in synaptically connected sequences called *neuronal circuits*.

Divergence and Convergence

Some neurons in the central nervous system may synapse with as many as 25,000 other neurons. When the

transmissive segment of a presynaptic neuron branches out to have synaptic connections with the receptive segment of many other neurons, it is an example of *divergence*. In diverging synapses, one neuron may excite or inhibit many others. In a *converging* synapse, the receptive segment of a postsynaptic neuron is excited or inhibited by the axon terminals of many presynaptic neurons.

Fibers from as many as several thousand presynaptic neurons may converge on one postsynaptic neuron. If the excitatory influences on the receptive segment are sufficient, an action potential can be generated.

Feedback Circuit

A *feedback circuit* is a mechanism for returning some of the output of a neuron (or neurons) for the purpose of modifying the output of a prior neuron (or neurons). Feedback in the nervous system is *negative feedback*, because the feedback modulates an effect by *inhibiting* the prior output. An example of negative feedback occurs in the regulation of body temperature.

Parallel Circuits

As a result of the neural processing of convergence and divergence, information is often conveyed by relay neurons to other neuron levels through *parallel circuits*. Through these circuits, different forms of neural information can be relayed, and ultimately recombined at the same time, at other levels.

Two-Neuron Circuit

The simplest neuronal circuit is the *two-neuron circuit*. It consists of a sequence of an afferent (sensory) neuron, one synapse, and an efferent (motor) neuron. A familiar two neuron sequence is the "knee-jerk" (patellar) extension reflex, where a tap on the patellar tendon of the flexed knee produces an extension of the knee.

Three-Neuron Circuit

A *three-neuron circuit* is a sequence of an afferent (sensory) neuron, an interneuron, and an efferent (motor) neuron. Such a chain is used in flexor reflexes, such as the flexion of the elbow when one touches a hot stove and pulls his hand away. The circuit involved in the elbow flexion consists of: (1) sensory neurons from the pain receptors in the hand to the spinal cord, (2) interneurons, located entirely within the spinal cord, which connect the sensory neurons with (3) the lower motor neurons to the biceps brachii muscle, which is stimulated to contract. There are synapses between sensory neurons and the lower motor neurons.

References

1. Robert Carola, John P. Harley, Charles R. Noback, "Human Anatomy," McGraw Hill, Inc., New York, 1992

2. Gerald D. Fischbach, "Mind and Brain," *Scientific American, Special Issue on Mind and Brain*, September 1992

3. Geoffrey E. Hinton, "How Neural Networks Learn from Experience," *Scientific American, Special Issue on Mind and Brain*, September 1992

4. Ed Reitman, "Experiments In Artificial Neural Networks," *Tab Books, Inc., Advanced Technology Series*, Blue Ridge Summit, PA, 1988

5. Voula C. Georgopoulos, "The Broad Use of Neural Networks in Real-Time Engineering Systems" *Proceedings of Electro/94 International*, June 1994

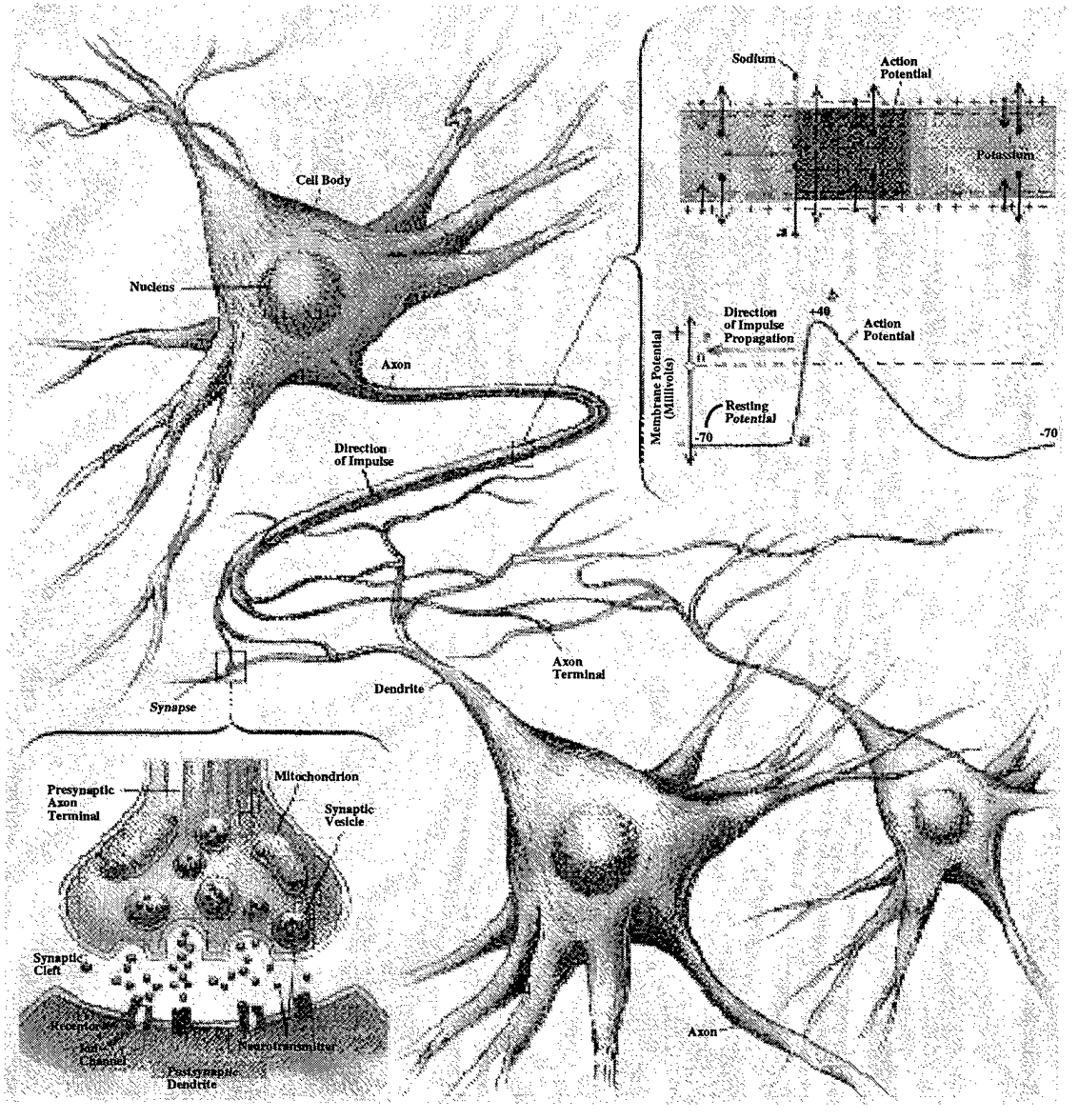


Figure 1. A Neuron, and Transmission of a Nerve Impulse across a Synapse (Courtesy of Prof. G.D. Fischbach, Harvard Medical School)

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

NEURAL NETWORKS TRENDS FOR FREQUENCY CONTROL : A REVIEW

B. Hivert, R. Brendel

* Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon - 32 avenue de l'Observatoire - 25044 Besançon Cedex - France

Abstract

The purpose of this paper is first, to give the reader a comprehensive "state of the art" in the neural networks field, then to discuss what kind of help such processing techniques may bring in the frequency control area.

So, we will give basic considerations about neural networks, explaining their principle. Then, we will show how they work and what kind of task they might perform, principally in the point of view of function approximation and pattern recognition. We will use this background to show some applications and to be critical on how neural networks may be used.

Then, a review of frequency control fields where neural networks may be applied will be presented. In order to let the reader make up its own opinion, particular attention will be paid to literature.

Introduction

Nowadays, Neural Networks (NN) are leaving the field of fashionable exotic techniques to enter the scientific toolboxes. But what are their capabilities ? This paper will try to depict the outlines of NN techniques. Furthermore, we will try to stay technical enough to allow the reader to make up its mind about the subject. Especially, some references will be given for each important point, and the common vocabulary of NN will be defined along, using **this typographical style**.

Before going any further, let us specify what type of Neural Networks we will talk about. Our tutorial will aim at the description of **Formal Neural Network**, or **Artificial Neural Networks**. That is to say, purely mathematical entities directly inspired by real living biological neural networks. But for the sake of simplification, we will use the acronym NN, with some precision when needed.

Background and Basic Tools

The formal Neuron - Mac Cullogh. & Pitts [1]

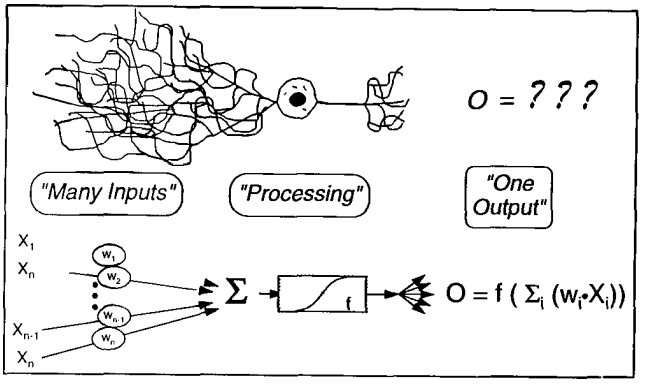


Fig. 1 : *The biological-formal neuron relationships*

The great majority of NN architectures are based upon a very simple basic cell, the **Artificial Neuron (AN)** also called **Mac Cullogh & Pitts Neuron**.

Let's have a brief look at a classical scheme of Biological Neuron (BN) [2]. We may divide it in three main parts, which are of special interest for us (see Fig. 1). First, to the left of the figure, we notice a huge interlaced tree of connections, the dendrite. With such a tree, the biological neuron is able to collect the informations from other neurons (roughly hundreds). To simplify, we will call that "the inputs of the neuron". The counterpart for the AN is clearly visible in Fig. 1.

Secondly, like any biological cell, the BN has a cell body with a nucleus. For simplification purpose we will assume that the cell body is doing the whole input processing task. The AN counterpart of the BN body has processing task too, which will be further explained.

Finally, for so much inputs (roughly hundreds), the BN has only one single output way, the axon.

0-7803-2500-1/95 $4.00 © 1995 IEEE

To summarize, the formal neuron, like its biological counterpart, has many inputs, which are processed together in order to compute one single output.

Unlike BN, AN have only ridiculous functionality. Thus, the artificial "Neuron" name is an heresy for most biologists. Otherwise, this name and the related terminology are widely used nowadays for artificial neural network. For example a connection between two AN may be called **Synapse**, like its biological counterpart.

However, let's have an insight in the AN. As stated before, it has multiple inputs and it processes them in order to give one single output. One of the most intuitive and simple operation to do is to sum the inputs, giving the same influence for all inputs on the output. Besides, to give more flexibility to this operation, each input is also multiplied by a dedicated coefficient. Such a coefficient is called the **Weight** of the connection.

Thus, the first operation performed by the AN is a weighted sum of its inputs

$$S = \Sigma\, w_i \cdot x_i$$

When Mac Culloch & Pitts designed there AN model, their aim was to compute logical function. For this purpose they add a threshold logical operation, which returns 0 if the sum S is beyond some threshold, and 1 otherwise. For the present neuron, this **decision** or **activation** function may have been replaced by a more continuous one, like the well-known **Sigmoid** or other classical ones (see Fig. 2).

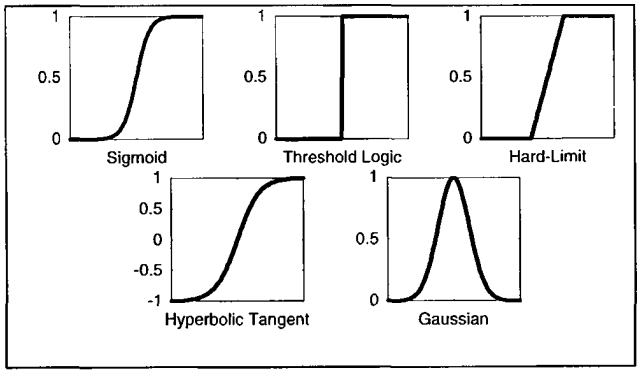


Fig. 2 : *Some classical activation functions*

At this step, let's have some graphical explanations upon the AN behavior. Assume for example that we have an AN with two inputs (see Fig. 3). Each input value will be represented along a classical 2D view with x_1 on an axis and x_2 on the other. So, one input vector will be represented by one point in the (x_1,x_2) plan. Moreover, w_1 and w_2 are the associated weights and θ the unit threshold. Therefore, the output 0 of the AN is given by

$$0 = f(w_1 x_1 + w_2 x_2 + \theta) = f(L)$$

Everyone may notice that the linear expression "L"

$$L = w_1 x_1 + w_2 x_2 + \theta$$

has a well-known behavior versus the $M(x_1, x_2)$ point placement over the plane.

- $(L) = 0$ when M is on the unique line defined by w_1, w_2 and θ
- $(L) > 0$ when M is on one side of the line
- $(L) < 0$ when M is on the other side.

Thus, using a threshold logical function, the AN allows us to classify different points in the plane. And, if x_1 and x_2 are some characteristics of an object to classify, one single AN is able to separate between two linearly separable classes. But such an important point will be discussed further.

The former result is easily extendable to n inputs, n greater than 2. For example, in three dimensions, the separating line becomes a plane, and with more dimensions, an hyperplane.

Since 1943, the Mac Culloch & Pitts neuron has been used as it is or with modifications. But the basic brick of NN remains with the same functionalities :

- several inputs
- one single output
- some simple nonlinear output function.

Although we know how one AN works, it may be of little help if we forget to use it in networks. In fact, the interest and the limitation of the AN is precisely the simplicity of its operation [3]. Furthermore, properly putting neurons together has saved the entire neural networks research field to be given up. See for example [4] or [5]. This leads us naturally to the study of neural networks.

Neural Networks

Biological NN processing power is almost entirely defined by their internal connectivity, rather than by the amount of neurons. For example, the weight of the mammal brain, hence the number of its neurons, is not directly related to its "processing power". Such a power seems to be much more related to the internal connectivity of the brain. This is also true for formal neural networks too.

Basically, an artificial NN is defined by its connection architecture, and the model of AN used. This involves at least three degrees of freedom :
- the connection placement
- the connection weights and other characteristics
- the neuron processing functions

Thus, it is necessary to have some algorithm for solving the problem of network definition to achieve a particular task.

Such an algorithm is generally named **Learning Algorithm**, in reference to biological systems. Learning is a common feature in evolved biological neural structures, human and mammal brain for example. Despite of this, the learning process in human brain and even simpler organizations is badly known for a biological sight. Learning with machines is a difficult issue too. As a bio-inspired approach, NN have tried to mimic such a capability.

Supervised Learning

One intuitive learning scheme is to proceed like a teacher. When a pupil learns how to read, the teacher shows the letter "A" and pronounces it. Then, the pupil is expected to establish the link between the "A" and its related sound. In the NN case, the process is roughly the same. The main difference is that NN are principally numerical algorithm, not symbolic ones. So, the "teacher" of a NN must find a way to digitize symbolic informations, if needed. In numerical oriented cases, that may be a problem of pre or post-processing (normalization, FFT, ...) But for some symbolic task, it may be a blocking factor.

But let assume that the whole problem of pre- and post-processing has been solved. The teaching algorithm, better called **Learning Algorithm** as stated before must therefore use sample input-output vectors or **Patterns**, in order to fix the NN weights.

From a mathematical sight, such a situation belongs to the function approximation field. The position of the problem is very simple. Some unknown function f is only known through some observations, i.e. n points of its curve (see Fig. 3). With these points, the goal is to design a function f' which is "as close as possible" to the original one.

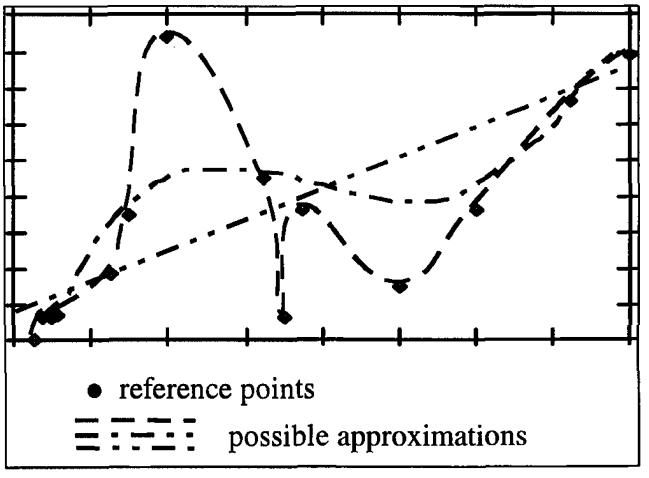


Fig. 3 : *The approximation scheme*

This is obviously an ill-posed problem [6], because there is an infinity of functions f' that may exactly go through the given points. For example, polynomials of degree greater than n. But with some more constraints over f' (smoothness, continuity, ...), one can expect a satisfying approximation of the desired function.

Notice that when f' is computed between the reference points, the task is called **Interpolation**. When f' is computed outside the reference point set, the task is called **Extrapolation**.

In the literature, NN capabilities to do **Pattern Recognition** or **Classification** are commonly quoted. The meaning of classification is rather self-explanatory. This consists in partitioning a set of objects (images, sounds, flowers, ...) in different clusters. For example, in the "flower" set, we may separate classes like red flowers, roses or graminae.

Doing classification by numerical means involves two steps. The first step consists in giving each object a set of numerical attributes. The second step, which may be the most difficult, is the use of these attributes to cluster the objects in determined classes. Character recognition is one of the most spectacular examples of pattern recognition.

The reader may have noticed that pattern recognition may be expressed via the function approximation theory. Pattern classification involves to design a function that returns 1 for a particular cluster and 0 otherwise. Assume that a function f takes its values from the pixel plane of an image. If f returns the value 1 when a "A" is recognized in the image, and 0 otherwise. Thus f is a "recognition function" of "A".

Unsupervised Learning

Unsupervised learning addresses the same problem as supervised learning, i.e. defining the network architecture in order to accomplish some task. But unsupervised learning algorithm is working with the input set only, and the output of the networks is driven by some characteristics of the input set.

Such an algorithm is less intuitive than the supervised one. We invite the reader to refer to the section about Feature Maps.

To conclude we will point out some common facts about current learning algorithms. We have stated that they are dealing with (at least) three degrees of freedom :
- the connection weight and other characteristics
- the connection placement
- the neuron processing functions.

The first one is the more studied problem. Generally, the network architecture and the AN activation function are fixed by the user before learning. The learning algorithm is thus generally in charge to compute the weights.

But relatively recent algorithm, namely **Pruning Algorithm** [9-11] and **Cascade Correlation** [12] are trying to solve the connection placement problem. To our knowledge, the problem of the modification of the neuron activation is not yet addressed.

In order to clarify the previous section, some common NN will now be described.

Static Neural Networks

This section will deal with NN that have an intrinsic static behavior. To be more specific, when they receive an input vector, the output computation involves no dynamic process.

Feedforward Neural Networks

Historically, this kind of NN was the first. The way it works is explained by its name. See Figure 4.

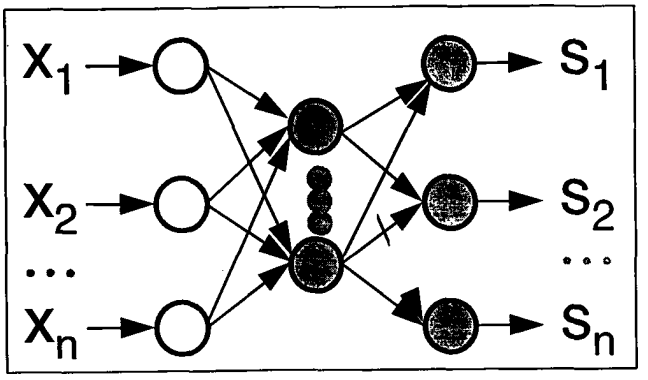


Fig. 4 : *A common Feedforward Neural Network*

A Feedforwad NN is split in layers of AN. When a layer has done his calculations, it feeds its results into the following one. The first layer contains only input points, with no processing capabilities. The last layer output define the output of the network.

This algorithm is claimed as massively parallel, because all AN of a layer may process the input together.

Such an architecture covers a crow of NN names such as : **Perceptron**, **Madaline** [13] (with only one processing layer), or **Multi-Layer Perceptron**, **Back-Propagation Network**, **Neocognitron** [14], (with multiple layers).

Let us point out the operation involved in Fig. 2. Assume for example that the network has three layers, and sigmoid activation function for each neuron. Recall that the first layer has only input points. For the second layer, the j^{th} neuron output is :

$$S_{1,j} = sigm(\Sigma\, w_{ji}\, x_i + \theta_{1j})$$

And for the third layer :

$$S_{2,k} = sigm(\Sigma\, w_{kj}\ S_{1,j})$$

so

$$S_{2,k} = sigm(\Sigma\, w_{kj}\, sigm(\Sigma\, w_{ji} x_i + \theta_{1j}) + \theta_{2k})$$

This type of sum may recall to the reader some well-known formulas. If we remove the rightmost sigmoid function :

$$\Sigma\, w_{kj}\, sigm(\Sigma\, w_{ji} x_i + \theta_{1j}) + \theta_{2k}$$

such a factorization may look like other classical ones. From a mathematical point of view it is like a polynomial or Fourier expansion, but in a nonlinear way.

Such an expression may be also interpreted from the pattern recognition point of view (see Fig 5).

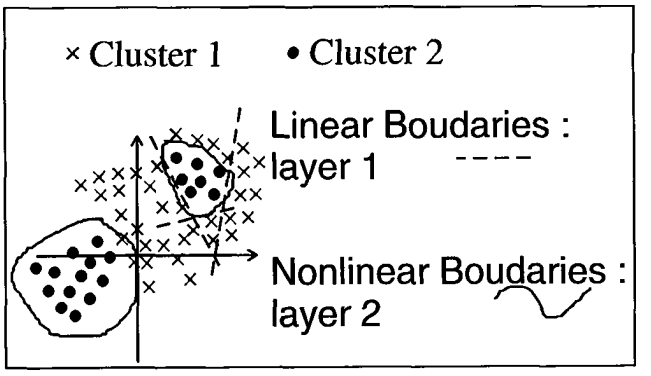


Fig. 5 : *Pattern recognition in a Feedforward Neural Network*

As we pointed out in the first section, one single AN is able to split the input space into two regions. If there are N inputs, then one AN separates them with a $N-1$ dimensions hyperplane. But in many cases, such a simple boundary is unusable, because the clusters have no such linear separation (see Fig. 5). For this purpose, one nonlinear boundary is needed. Such a nonlinear separation is performed by the third layer, which classify the second layer output.

This is a very interesting result. During the last decades, several theorems have been proved that show the universality of the model shown in Fig. 5 [15-20]. To summarize, even for function approximation or for pattern recognition, a three-layer network (one for input points, two for AN) is enough, provided that the desired function is square-integrable. But like any good mathematical theorem, they prove that three layers are enough, but they don't give any pointer to the dimension, connectivity, etc. of the network. With this three-layered simple architecture, this remains an open problem.

However, for much complicated recognition tasks, more than three layers are commonly used [21].

A common problem of this approach is the number of samples necessary to achieve a suitable approximation. Although this remains an open problem too, there is some features that must not be ignored. Despite their unusual background, NN are parametric methods that must follow the same rules as much more "classical" ones. For example, the samples in the learning base must at least be as numerous as the NN parameter. Otherwise, NN may have free parameters and hence undetermined behavior. Moreover, over-parametrisation may lead to learn the noise in the sample set rather than the desired function. This generally leads to poor interpolation capabilities. Such a capability is either called **Generalization** in the pattern recognition context and is essential.

Feedforward NN, and especially three-layer NN, are commonly called back-propagation network. But it is a misnomer, because back-propagation is much more a learning algorithm than a computing scheme.

Let us have a closer look to back-propagation algorithm. When this algorithm is used to train some NN, an input pattern is fed into, and the ouput of the NN is computed. Then, the square of the error between the desired and the computed output is back-propagated toward the NN in order to compute weight changes. In a more mathematical way, the square of the output error is used to initiate a gradient descent method for weight computation. Gradient descent method is a well-known optimization technique, that has many drawbacks (slowness and local minima) and is no longer used for optimization purposes. However, the back-propagation, which is a first order method, is still in use because its simplicity and because second order methods have generally higher calculation cost.

Such NN have a very large spectrum of applications : classification, process command and identification, time series predictions, nonlinear filtering, etc. Three layers NN with back-propagation have been widely used, and are still in use now although much more improved algorithms are available. For example **Pruning Algorithms** [7-12] and **Cascade Correlation** [13] try to address the problem of network connection automatic design. To overcome the limits of Backpropagation algorithm, simulated annealing has been used for the **Boltzmann** and **Cauchy Machine** [22-23]. The **Radial-Basis Function Network** [24-25] is also an "universal approximator" and seems to have some promising capabilities.

Feature maps

Feature maps and other projectors have a different type of behavior than feedforward NN. Our principal example in this section will be the well-known **Kohonen Feature Map** [26] (KFM) (see Fig. 6).

A KFM has slightly different neurons from the standard AN description. A KFM neuron has connections from the KFM inputs, but also from neighboring neurons. Such connections define the neighborhood of the KFM neuron. These connections are called **Inhibitory Synapses**, because when a neuron is activated, it tends to activate the nearest neighbors and to inhibit the rest of the network (see Fig. 6). In order to understand how a KFM works, let us examine its learning process (see Fig 7). Notice that this is a typical example of unsupervised learning. When an input vector is fed into the network, each neuron computes the distance between that input vector and its own weight vector. The neuron that has the greatest output is the winner and its weights are modified so as to be closer to the current input. So, the KFM neuron tends to react to only a small region of input vectors, in a way of preserving the input vector topology. When the learning comes to its end, the KFM has done an auto-organization process, which only depends on the inputs fed when learning.

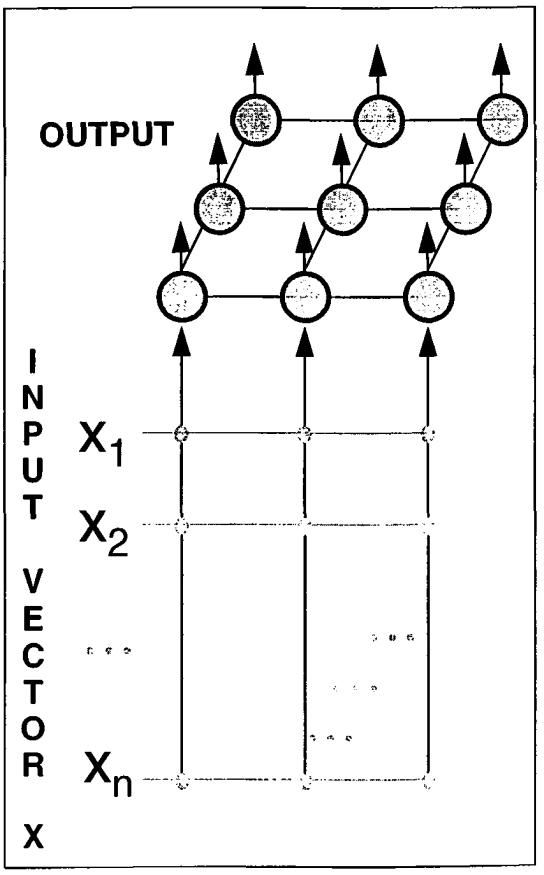


Fig. 6 : *Some example of KFM connectivity*

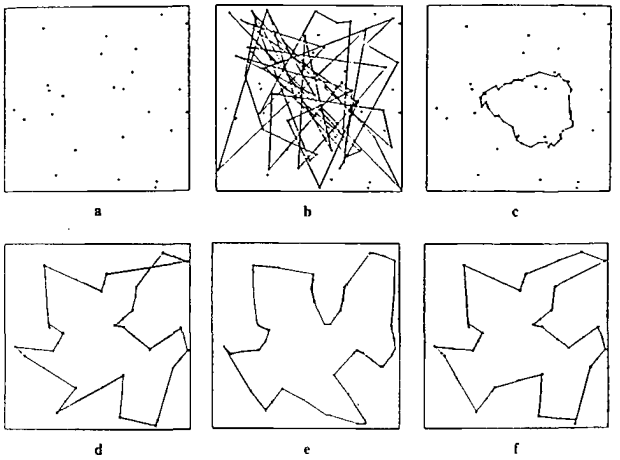
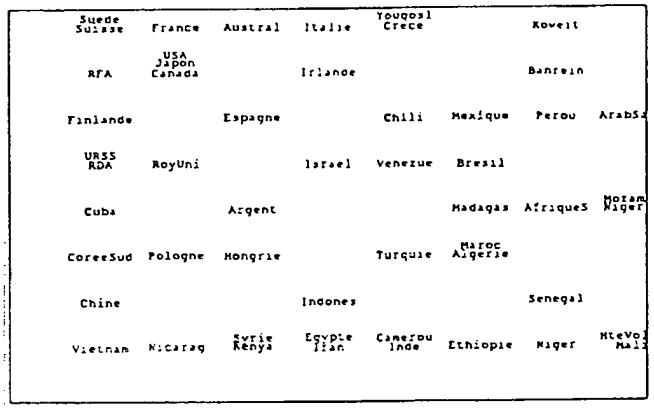


Fig. 7 : *Kohonen Feature Map during learning steps*

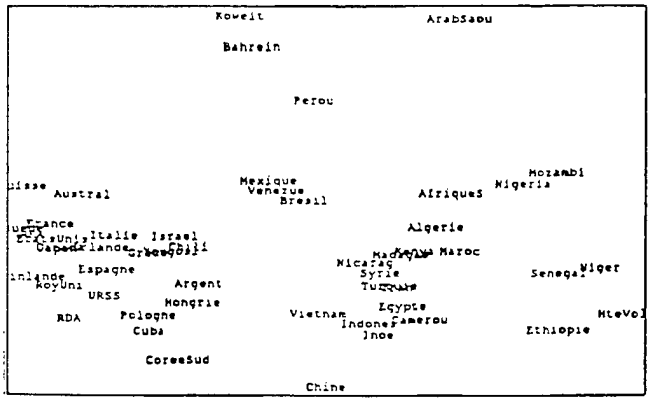
Such a process is also called Vector Quantization (VQ), because it reduces the great number of input vectors to those matching the KFM neurons. It looks like the K-means algorithm, which is also a VQ process, but KFM tends to keep the input topology.

Vector Quantization tends by definition to reduce the dimension number of the inputs. This may lead to pattern recognition [27] too, because several nearly identical inputs are detected by one KFM neuron. In

particular cases, dimension reduction may imply data compression [28]. When the topology of the KFM is a line, it has been used to solve the travelling salesman problem. The feature of the KFM used here is that each neuron must have neighbors that have close weight vector from them. For the travelling salesman problem, KFM have less complexity and better results than Hopfields networks [29] (see next section). The latest application presented here is data analysis from the statistical point of view. KFM have results that may be compared to principal component analysis [30] (see Fig. 8).



a



b

Fig 8 : *KFM doing statistical analysis*

The learning has been performed with some socio-economical data of 52 countries and the reader may notice that the KFM has successfully clustered the G7 members, the African countries, and the petroleum producers.

Summarizing KFM is able to do vector Quantization and nonlinear projection. In fact, they are relatively robust and easy to use. Their major drawback is that the topology (style of neighborhood) is user-defined and may not be suitable for a particular problem (see Fig. 9)

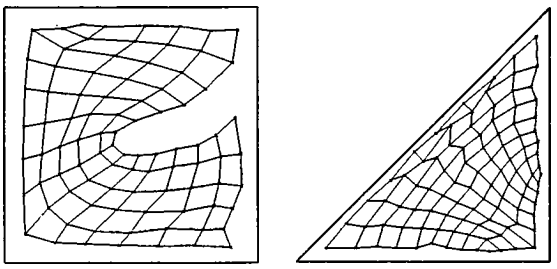


Fig 9 : *KFM having a bad topology with respect to its input set*

To conclude, some other works have tried to design more flexible vector quantization techniques, such as **Elastic Net, Vector Quantization and Projection** [31].

Dynamic Neural Networks

This type of architecture covers a type of NN structures that react with a dynamical behavior to an input vector. To introduce the reader in dynamical NN, we will use the Hopfield Network example. Such an architecture has been widely studied and inspired by statistical physics and particularly spin glasses.

The AN used in Hopfield Networks is the classical one, with a threshold logic activation function. Besides, each neuron is connected to all others (but not to itself). The weight matrix is said to be symmetric, because the weight that connects a neuron A to B is the same as the one between B and A.

The Hopfield NN has no input, the input of such a structure is the initial state of its neurons. Once the initial state has been fixed, the evolution rule is quite simple : a neuron is randomly selected, and it computes its new state with the standard method for AN. Then another neuron is selected, and this continues until there is no more state change at all.

The interest of such a process is that there is an energy function which may be linked with the update process. This energy function is positive and also proves to decrease along the neuron state change. This implies that the Hopfield NN has a stable state, defined by its connection matrix.

When a stable state has some significance, this may lead to the example shown in Fig. 10.

In Fig. 10, the Hopfield net has managed to associate the letter "A" to a noisy image of it. So, Hopfield nets may be named **Associative Memories**. More precisely, these are **Autoassociative Memories**, because they establish the link between same objects. Heteroassociative Memories [32] make links between different objects.

But some theoretical results have proved that Hopfield NN cannot store a number of patterns greater than 15 % of its neuron number. Trying to store more patterns causes the NN to have meaningless minima. This basic limitation explains the lack of application for Hopfield net. However, if the activation function is changed from threshold logic to a continuous one, the Continuous Hopfield NN exhibits some interesting dynamic capabilities. It has been used for transverse filter identification and noise cancellation.

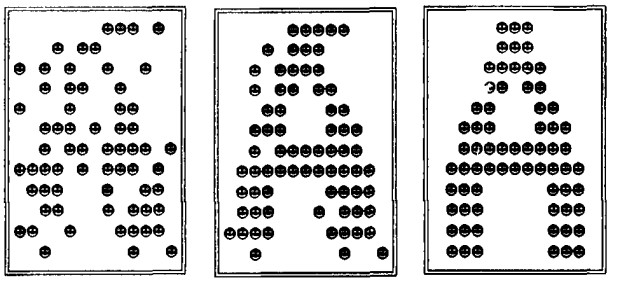


Fig. 10 : *An Hopfield Net recognizing a "A"*

One of the interesting aspects of Hopfield Network is that it evolves until its energy reaches a minimum. Thus, it has been used to solve optimization problem. The first historical consisted in solving the travelling salesman problem [33].

In a similar approach the Boltzmann Machine (BM) [22-23] tries to use the Simulated Annealing performances in order to achieve better optimization and to avoid trapping in local minima. The BM neuron has the classical weighted sum for its input, but the activation function is a probabilistic one. The state of one neuron depends on the value of its inputs, of its former state and of a parameter called the temperature. Like for real annealing, the temperature T is decreasing during the evolution of the Network. The neuron states are updated as for Hopfield NN. At the beginning of a relaxation process, when T is high, the network tends to get out of local minima. When T progressively decreases, the network slowly converges to the global minimum. This architecture exhibits interesting results for optimization and pattern recognition [34].

The two former architectures seem to show promising features, however they have few applications now. Especially, the Boltzmann machine is waiting for a fast hard implementation, so as to reach real-time application needs.

Associative Memories have aroused much interest, and Hopfield Networks have had much successors. Here are a few :
- Adaptative Resonance Theory 1 & 2 [35]. ART is able to do pattern recognition, with the ability to learn new unknown pattern
- Bidirectional Associative Memory [35], with the same capabilities
- Brain State in a Box Network [36].

Otherwise there are a few other dynamic Neural Networks. These are modified feedforward networks involving recurrent link. A recurrent link is a backward delayed connection giving the NN a short-term memory of its previous states. This allows to implement nonlinear recursive filters for example. Such structures have been successfully applied for time series predictions, and adaptative filtering.

To conclude this section, we may say that Formal Neural Network user and researcher must try not to forget the biological origin of their tool. Formal Neural Networks may always be expressed via classical mathematical relations, thus hiding the biological roots of their algorithms. But their inherent philosophy, bio-mimetism is generally the source of their processing power.

Neural Networks Trends for Frequency Control

To our knowledge, there are nowadays very few real applications of NN for typical Frequency Control purpose. This is rather surprising, because NN are already in industrial applications in several domains, and especially for artificial vision purpose [37-38]. We hope that the previous tutorial has briefly pointed out some evidence, and the reader must have guessed at least some part of this section content.

In order to conclude this paper, we shall try to list some needs of Frequency Control techniques, and try to identify where NN may fill some lack or bring some improvements.

The reader is certainly aware of Frequency Control goals and needs, let us briefly summarize some of them :
- frequency stabilization
- frequency tuning
- noise, drift and ageing detection and cancellation.

It may be obvious that NN may be of little help to the "hard" part of Frequency Control, i.e. device and electronic fabrications. But NN processing capabilities may be of greatest interest for general signal processing purpose, and especially :
- filtering (nonlinear or not, adaptative or not)
- identification and control.

Cristal Oscillators

Let us study the particular case of crystal oscillator (XO). Unlike primary references they are often on-board devices. Thus, their sensitivity to external influence must be reduced to ensure that they show suitable behavior. This may be in very tough condition of vibration, magnetic field, temperature, pressure, etc. The primary controlled parameter is the temperature, because crystal oscillators are particularly sensitive to it. They are many solutions for isolating the oscillator from external temperature fluctuations. The most drastic one is by using Oven Controlled Crystal Oscillators (OCXO). With such a device, the oscillator has a dedicated oven, which

allows to regulate oscillator temperature. Another case are Temperature Controlled Crystal Oscillators (TCXO). The basic difference between OCXO and TCXO is that in TCXO there is only a signal processing aimed at temperature compensation of the frequency drift. Generally, a few sensors are used to measure the environment parameters.

In our sense, the main problem of XO temperature control is that there is often only one temperature sensor and, in the best case the only device whose temperature is constant is the sensor itself [39-40].

At this point, we must point out a lesson of the bio-mimetic approach. Engineers and scientists tend to use one single sensor, and then to build a huge sophisticated signal processing afterward in order to obtain suitable results. If we observe biological sensor systems, the main feature is redundancy. On a square millimeter of a human digit, there are hundreds of sensors of the same type. In the human nose, the convergence ratio is about one neuron for 1 000 sensory cells.

We are just explaining that, in place of one poor single temperature sensor, why not use tens of them ? For example, even in a well-insulated OCXO, the crystal temperature is difficult to assess, because it involves the modelling of all the heat transfer effects in the oven. This involves conduction effects, radiation effects, convection effects. With so much parameters, it may be of great help to measure the temperature of the oven in several points, the remaining problem being the processing of the resulting data. At this point, we think that NN may bring some help. They may allow to identify the behavior of the output frequency versus the various sensor outputs and the XO driving parameters. Using the results of this first step, NN may also be used to control the output frequency, with the sole knowledge of the sensor value. Such a scheme is described in [41], with some applications. With such a scheme, one can even imagine a more general structure which could be called Environment Controlled Crystal Oscillator (ECXO) where the XO has no particular protection unless its packaging.

Noise cancellation

One of the greatest limitation in signal processing becomes generally from the noise. NN have been used in many applications for noise cancellation [5] [42]. It should be surprising that Frequency Control scientists would not be able to find some applications where this may be applied. Furthermore, a particular architecture, namely the Herault-Jutten Network, has shown the ability to separate mixed sources. Assume that two sensors (s_1, s_2) are measuring the signal from two sources (S_1, S_2), for example in a linear way. The resulting signals are :

$$s_1 = a_{11} S_1 + a_{12} S_2$$
$$s_2 = a_{21} S_1 + a_{22} S_2$$

Provided that S_1 and S_2 are uncorrelated, the network shown in Fig. 11 is able to restore S_1 and S_2 at its output. The learning is on-line, the Herault-Jutten network uses the s_1 and s_2 signals only (and a clever learning rule...) in order to work properly. An analog implementation of such a NN has been made, leading promising results. The reader may have further explanations in [44].

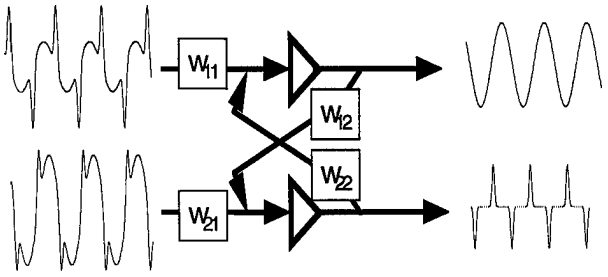


Fig. 11 : *Herault-Jutten Neural Network for two sources separation*

One may object that the only measure of an oscillator device is its single output. But an answer may be that the more sensors you have, the best the result is... However, noise source separation may allow to identify them and to decrease their influence on frequency.

Variance and other statistics computing

The most often used statistics is the Allan variance or the modified Allan variance which remain stationary measurements. It may be useful to use some NN in order to investigate what sort of improvement may bring nonlinear dynamic measurements in such a domain. NN and statistics have common issue, and there is some hope that a common research will bring some interesting results [45, 46].

International Atomic Time (TAI) and other Atomic Time computing

From the previous remarks, we may notice that TAI and other atomic time standard computation seems to near the NN capabilities field. The basic principle of TAI computing relies on the use of many primary cesium clocks. The TAI is the weighted sum of each contributing clock. The weight attributed to a clock depends on its particular behavior versus the mean of the others. The greater the difference is, the smaller the attributed weight is. The weight attribution algorithm behaves thus like a filter versus each clock behavior. Our opinion is that Neural Networks may be tried for such a purpose. The TAI computation algorithm is about to be revised. Further improvements may perhaps involve Neural Networks.

One of the most important issue lies in the real-time capability of neural networks. This is also a background question for this whole section. In the case of XO we may assume that external influence variation may be sufficiently slow to allow a direct processing with NN.

NN hardware is generally working with digital principles, and thus has a limited operational

bandwidth (see comp.ai.neural-nets newsgroup frequently asked questions). Here is although a few analog NN implementation. It is obvious that NN will not be able to work directly on oscillators output, but rather on frequency sample or other acquired data.

This review comes to its end, we hope that the reader found it interesting and useful. However, the whole content of this paper may reveal as garbage in a few year. Dear reader, please do not forget that we are human. We have tried to be as objective as possible in our obviously limited knowledge. And please notice that we are not clairvoyants, but scientists.

Conclusion

Fundamental aspects of Neural Networks have been reviewed. Frequency Control and Neural Network are currently separated fields, but they seem to have many possible connections. The future of Neural Networks seems to lie in a more bio-mimetic approach. This is not obvious for Frequency Control, because neuron and gigahertz seems to have no special link actually. However, we must not forget that the universe, and hence the living realms is made of vibrating phenomenons. The season cycle, the night and day one, the heartbeats are natural frequencies. Remember that the neuron itself codes its output through peak frequency. Like the Neural Networks, may be the next oscillators will be more bio-mimetic ones.

Aknowledgments:

Many thanks to Steve Cantor who has initiated and supported the neural networks session in AFCS. Many thanks also to the whole people who have granted the material of this article, especially, ACTH and NSI fellows.

References

[1] W.S. Mc Culloch, W.H. Pitts "A logical calculus of ideas immanent in nervous activity". Bull. Math. Bioph. 3, 115-33, 1943.

[2] G.M. Shepherd. *Neurobiology*, Oxford University Press, New York 1988.

[3] M.L. Minsky, S. Papert, (1969, 1987) *Perceptrons*, MIT Press, Cambridge, MA.

[4] J.A. Anderson, E. Rosenfeld, *Neurocomputing* Foundations of Research, MIT Press, Cambridge, MA. (1988).

[5] R. Hecht-Nielsen, *Neurocomputing*, Addison-Wesley San-Diego, California (1990).

[6] A.N. Tickhonov, V.Y. Arsenin, *Solutions of ill-posed problems*, Washington D.C. (1974).

[7] Y. Le Cun, J.S. Denker, S.A. Solla, "Optimal Brain Damage", Advances in Neural Information Processing Systems, vol. 2, R.P. Lippmann, J.E. Moody, D.F. Touretzky Ed., Morgan Kaufmann (Proceedings of NIPS 89).

[8] D.L. Reilly, L.N. Cooper, C. Erlbaum, "A neural model for category learning", Biological Cybernetics, vol. 45, pp. 35-41 (1982).

[24] M.J.D. Powell, "Radial basis functions for multivariable interpolation : a review", in Mason and Cox : Algorithms for Approximation, Clarendon Press, Oxford, pp. 143-167 (1987).

[25] T. Poggio, F. Girosi, "A theory of networks for approximation and learning", AI Memo n° 1140, CBIP paper n° 31, july 1989, MIT Press, pp. 1-85.

[26] E. T. Kohonen, *Self-Organization and Associative Memory*, 2nd ed., Springer-Verlag, Berlin (1987).

[2è] E. Farges, "Systèmes neuromimétiques de diagnostic automobile temps réel", Proceedings de Neuro-Nîmes '91 (1991), Nîmes (France), pp. 769-771.

[28] A. Gersho, R. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Press, London, 1992.

[29] J.C. Fort, "Solving a combinatorial problem via self-organizing process : an application of the Kohonen algorithm to the travelling salesman problem", Biological Cybernetics, vol. 59, pp. 33-40 (1988).

[30] M. Cotrell, J.C. Fort, "A stochastic model of retinotopy : a self-organization process", Biological Cybernetics, vol. 53, pp. 405-411 (1986).

[31] P. Demartines, J. Hérault, "Representation of nonlinear date structures through a fast VQP neural networks", Proceedings de Neuro-Nîmes '93 (1993), Nîmes (France), pp. 411-424.

[32] G.A. Carpenter, S. Grossberg, (The ART of adaptive pattern recognition", Computer, pp. 77-88 (1988).

[33] J.J. Hopfield, D.V. Tank, "Neural computation of decisions in optimisation problems", Biol. Cybern, vol. 52, pp. 141-152 (1985).

[34] J.O. Klein, P. Garda, "La machine de Boltzmann est-elle vraiment lente ?", Neuro-Nîmes 92, EC2 Eds., (1992).

[35] B. Kosko, "Constructing an associative memory", Byte, pp. 137-144 (1987).

[36] J.A. Anderson, "A simple neural network generating an interactive memory", Math. Biosciences, vol. 14, pp. 197-220 (1972).

[37] Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel (1990). Handwritten digit recognition with a back-propagation network. *Advances in Neural Information Processing Systems*, vol. 2, R.P. Lippmann, J.E. Moody, D.F. Touretzky Eds., Morgan Kaufmann (Proc. of NIPS 89).

[38] J.D. Muller, "La perception structurante", Thèse de Doctorat, Ecole Nationale Supérieure de l'Aéronautique et de l'Espace, Toulouse, 1993.

[39] F.L. Walls, "Analysis of high performance compensated thermal enclosures", Proc. 41st Ann. symp. Freq. Cont., pp. 439-443 (1987).

[40] R. Brendel, G. Marianneau, F. Djian, E. Robert, "Improved OCXO's oven using active thermal insulation", IEEE UFFC, vol. 41, n° 2, pp. 269-274 (1994).

[41] K. Warwick, G.W. Irwin and K.J. Hunt, *Neural networks for control and systems*, Peter Peregrinus, London, 1992.

[9] Y. Le Cun, "A learning scheme for asymmetric threshold network", COGNITIVA 85, Paris (France), pp. 599-604 (1985).

[10] Y. Le Cun, Modèle connexioniste de l'apprentissage, Thèse de Doctorat de l'Université de Paris VI (1987).

[11] M. Cottrell, B. Girard, Y. Girard, M. Mangeas, "Times series and neural network : a statistical method for weight elimination", 1st European Symposium on Artificial Neural Networks, Brussels (Belgium), D facto Ed., pp. 157-64 (1993).

[12] S.E. Fahlman, C. Lebiere, *The Cascade-Correlation Learning Architecture*, Carnegie Mellon University Report CMU-CS-90-100 (1990).

[13] B. Widrow, "30 years of adaptative neural networks", Proceedings of the IEEE, vol. 78, n° 9 (1990).

[14] K. Fukushima, "Neocognitron : a hierarchical neural network capable of visual pattern recognition", Neural Networks, vol. 1, pp. 119-130 (1987).

[15] P.L. Bartlett, "Vapnik-Chervonenkis dimension bounds for two- and three-layer networks", Neural Computation, vol. 5, pp. 371-373 (1993).

[16] G. Cybenko, "Approximation by superpositions of sigmoidal functions", Mathematics of Control, Signals and Systems, vol. 2, n° 4, pp. 303-314 (1989).

[17] K.L. Funahashi, "On the approximate realization of continuous mappings by neural networks", Neural Networks, vol. 2, pp. 183-192 (1989).

[18] R. Hecht-Nielsen, "Kolgmogorov's mapping neural network existence theorem, "Proc. of the IEEE 1st Conf. on Neural Networks, San Diego (USA, Ca), vol. 3, pp. 11-14 (1987).

[19] K. Hornik, M. Stinchcombe, H. White, "Multilayer feedforward networks are universal approximators", Neural Networks, vol. 2, pp. 359-366 (1989).

[20] K. Hornik, M. Stinchcombe, H. White, "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks", Neural Networks, vol. 3, pp.551-560 (1990).

[21] D.L. Chester, "Why two hidden layers are better than one", in Proceedings of the Int. Joint Conf. on Neural Networks, vol. 1, Erlbaum, pp. 265-268 (1990).

[22] G.E. Hinton, T.J. Sejnowski, "Learning and relearning in Boltzmann machines", in D.E. Rumelhart, J.L. McClelland, and the PDP Research Group (Eds.), Parallel Distributed Processing, vol. 1, MIT Press, Cambridge, MA (1986).

[23] H. Szu, R. Messner, "Adaptative invariant novelty filter", Proc. IEEE, vol. 24, p. 519 (1986).

[42] L. Fu, "Adaptation signal detection in noisy environments", The Journal of Neural Network Computing, Spring Issue, pp. 42-50 (1990).

[43] B. Widrow, J.R. Glover, Jr., J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong, Jr., R.C. Goodlin, "Adaptive noise cancelling : principles and applications", Proc. of the IEEE, vol. 63, n° 12 (1975).

[44] J. Hérault, C. Jutten, *Réseaux neuronaux et traitement du signal*, Hermes, Paris 1994.

[45] F. Blayo, P. Demartines, "Data analysis : how to compare Kohonen Neural Networks to other techniques ?", in A. Pietro (Ed.), International Workshop on Artificial Neural Networks (1991). Lecture notes in Computer Science, vol. 540, pp. 469-476, Springer Verlag (1991).

[46] W.S. Sarle, "Neural networks and statistical models", Proc. of the 19th Annual SAS Users Group International Conference, april 1994, pp. 1538-1550.

Useful books for beginners

R. Beale, T. Jackson , *Neural computing*, Adam Hilger Bristol, England 1990.

B. Muller, J. Reinhardt, *Neural Networks: an introduction*, Springer Verlag 1990.

D.E. Rumelhart, J.L. Mc Clelland and the PDP research group, *Parallel Distributed Processing. Exploration in the microstructure of cognition. Vols. I, II and III.*, A Bradford book, MIT Press Cambridge MA, 1986.

A few journals on NN

IEEE Transactions on Neural Networks, Publisher : IEEE.

Neural Networks, Publisher : Pergamon Press, ISSN 0893-6080.

Neural Computation, Publisher : MIT Press, ISSN 0899-7667.

Neural Processing Letters, Publisher : D facto Publications, ISSN : 1370-4621.

Finally, we encourage the interested reader to have a look at the newsgroup comp.ai.neural-nets, and especially to the Frequently Asked Questions (FAQ). Such a document may be found at :
URL :
http.//wwwipd.ira.uka.de/~prechelt/FAQ/neural-net-faq.html.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

EVALUATION OF OVEN CONTROLLED CRYSTAL OSCILLATOR PERFORMANCE UNDER NON-CONTINUOUS STRESSED OPERATING CONDITIONS

Asbjorn M. Gjelsvik, Stephen R. Cantor, Theodore E. Ioakimidis

The MITRE Corporation
202 Burlington Rd.
Bedford, MA 01730

Abstract

The frequency stability of a set of six oven controlled crystal oscillators has been measured under non-continuous and other stressed operating conditions. The oscillators were subjected to a range of environmental conditions which would be experienced in an airborne communications terminal. A database, corresponding to those measurements, was generated for support of a Kalman Filter/Neural Network modeling and training effort.

The tests were performed at a MITRE environmental control laboratory utilizing a vibration table which was enclosed within a heating chamber. An automated measurement system was used for control and data logging. Initial observations of the test results indicate that non-continuous operation in a benign environment has no noticeable impact on the aging characteristics, while aging under stress is harder to discern from frequency measurements made under composite environmental stress conditions. More elusive are the infrequent but sudden and permanent changes in frequency that appear to occur at random. The magnitude of those frequency jumps was often significant, and justifies further investigation.

Introduction

This paper describes the measurements made on a set of six oven controlled crystal oscillators (OCXOs). This was done in support of an investigation into the feasibility of improving the frequency accuracy of a crystal oscillator by real time compensation based on calibration data derived from Neural Network training. The frequency stability of the OCXOs was measured under a range of vibration, temperature, and non-continuous operating conditions which would be experienced by an airborne communications terminal. A database, corresponding to those measurements, was generated for use in the training of our Kalman Filter/Neural Network Model. Since our ultimate goal was to demonstrate the degree of compensation possible, the choice of the particular type of crystal oscillator was not a driver. The OCXO we used was a single-oven enhanced miniature crystal oscillator (EMXO), because it was state-of-the-art and we had six of them available in our laboratory.

Test Configuration

The test set-up used for the EMXO evaluation is shown in Figure 1. The system is designed to monitor the frequency stability of oscillators relative to a cesium beam time and frequency standard, as the oscillators are subjected to various types and levels of environmental stress. The monitoring is accomplished by means of phase measurements from which frequency offset relative to the cesium can be derived, rather than by direct frequency measurements. This method has the advantage of yielding data at the precision needed to determine both long-term and short-term stability on multiple oscillators under test, simultaneously, without requiring continuous and dedicated monitoring of each, as would be required with the direct frequency measurement method.

Unit Under Test Assembly

The unit under test (UUT) assembly, as shown in the bottom right section of Figure 1, was comprised of six

0-7803-2500-1/95 $4.00 © 1995 IEEE

EMXOs and a synchronized set of divide-by-10^7 circuits. The oscillators were all mounted in-line between two aluminum plates in order to maintain the same case temperature. This mounting arrangement is shown in Figure 2, before the aluminum plates with the oscillators had been clamped down to the mechanical interface on the shake table. The mechanical interface was needed to adapt the large mounting bolts on the shaker head to the lighter structure holding the oscillators. When mounted, the nameplate side of the oscillators seen in the picture was parallel to the shake table, and consequently perpendicular to the vibration vector. A thermocouple was attached to one of the side plates for monitoring of the oscillator case temperature. A second thermocouple was positioned next to the support plates for measurement of the ambient temperature. The dividers and synchronization circuitry were housed in a separate chassis mounted in the rack with the rest of the test equipment. Only the oscillators were subjected to vibration and thermal stress.

As shown, the EMXOs were operated at a frequency of 10 MHz, and thus the corresponding outputs (divide-by-10^7 outputs) of the UUT assembly were each at a one pulse-per-second (1PPS) rate. The synchronization circuit included with the set of dividers functioned to establish the initial alignment of these outputs to within one microsecond upon receipt of the "counter reset" signal, which was initiated by the System Controller via the Data Acquisition/Control Unit. The synchronization process was executed at the start of each test relative to the cesium reference, such that the UUT outputs all occurred initially at some preset delay following the cesium output. This was done to get a small positive reading on the time interval counter between the 1 PPS reference signal and each of the UUTs. Care was taken, however, during initial synchronization to select a UUT delay such that none of the oscillators whose frequency was higher than the cesium would drift more during the test than the preset delay and thereby cause a negative time interval reading. In addition to the inconvenience and computational complexity, a small negative measurement would also cause the time interval counter to use two seconds for measurement of UUT relative phase versus the one second measurement interval required for the rest of the UUTs that remain at a positive delay throughout the test. This would perturb the established data sampling sequence and cause erroneous measurements.

Frequency Measurement

The equipment used for measurements and to compute the actual frequency offset of each oscillator included the following:

Cesium Beam Time and Frequency Standard: Provided a 10 MHz reference and associated 1 PPS reference output, both of which were used by the HP5370B in making its measurement.

HP59307A Very High Frequency Switch: Multiplexed the six UUT outputs for measurement against the 1 PPS signal from the Cesium Beam Time and Frequency Standard.

HP5370B Time Interval Counter: Measured the time delay between the 1 PPS signal from the Cesium Beam Time and Frequency Standard, and the output signal from the HP59307A VHF switch, which corresponded to one of the six UUT outputs. Two such delay measurements (at one second intervals) were made consecutively for each of the six switch positions, and the difference between them was used by the System Controller to derive the frequency error. The controller stored all such measurement data (including time of measurement), enabling it to ascertain both short-term and long-term stability.

HP9816 System Controller: Controlled the UUT assembly via the Data Acquisition/Control Unit and the

VHF switch. It commanded the Time Interval Counter to take readings and used the resulting data to calculate the amount of frequency error in each of the oscillators under test as compared to the Cesium Standard.

Oscillator frequency error and change in time delay measurements are related by the following expression:

$$\Delta f/f = -\Delta t/t$$

where $\Delta f/f$ is the fractional frequency error, and Δt is the change in delay over the interval t.

Data Collection/Storage/Display

The following equipment was used to collect and store test data on the oscillators under test:

HP9816 System Controller: Selected desired channels from the Data Acquisition/Control Unit to read thermocouple voltage measurements and converted them to temperature readings. Read the power supply voltage and the total current drawn by the oscillators under test. It also actuated the appropriate channels on the Data Acquisition/Control Unit to reset (synchronize) the dividers in the UUT Assembly. For EMXO evaluation, 12 delay measurements (two for each oscillator) along with the time of measurement, case and ambient temperatures, and voltage and current readings, were recorded on 3.5 inch disks every 20 seconds. During the test, the first set of readings and every ninth set thereafter were sent to the HP2673A printer for display. The rest of the real-time display was provided on the System Controller screen for observation as the test progressed.

HP3497A Data Acquisition/Control Unit: Provided an interface and multiplexed thermocouple voltages, power supply voltage, and current monitor voltage to an internal voltmeter. It also provided actuators, which were used by the System Controller to reset/synchronize the UUT assembly counters, and to automatically start the test and execute the test sequences under program control.

HP9121D Dual Disk Drive: Provided storage medium (3.5 inch floppy diskette with 270 k-byte single-sided capacity) for test data from the System Controller.

Environmental Facilities

A thermal chamber with a companion shake table was used for the EMXO environmental stress evaluation. In the operating configuration, the thermal chamber was positioned to straddle the shake table with the shaker head protruding into the bottom of the chamber under a teflon membrane as seen in Figure 2. The oscillator assembly was mounted on a support fixture secured to the shaker head on the chamber side of the membrane. All the stress tests were performed with the oscillators mounted in this manner. Some general characteristics for the environmental equipment are included below:

Thermotron ESS-1200 Thermal Chamber: The Thermotron model ESS-1200 chamber is capable of sustaining temperatures from -50°C to +150°C and can change the temperature at a rate of up to 5°C per minute, depending on the test specimen thermal mass. The bottom of the chamber is fitted with a teflon membrane to facilitate a mechanical interface to a companion shake table.

LDC DPA8 Shake Table: The Ling Dynamic Systems model DPA8 shake table is a single axis vibration unit that can be used independently or in combination with the Thermotron ESS-1200 thermal chamber. Vibration levels can be programmed and controlled from essentially zero to 20 g_{rms}.

Test Results

This section contains the results of the measurements performed on the six EMXOs during the period of

August 1992 through June 1993. Prior to August 1992, those oscillators had been stored in a laboratory ambient environment, normally at 25°C, without power for approximately six months. The tests performed were: aging, retrace, cold and hot temperature storage, temperature cycling, vibration testing, and combined vibration-temperature testing. The objective of these tests was to measure the frequency stability of the EMXOs under varying stressed and unstressed conditions, and collect data for a study to determine the feasibility of compensating for the frequency fluctuations by means of Kalman Filter/Neural Network techniques. The testing started with six EMXOs in the testbed. All of the oscillators exhibited typical operation at the outset; and EMXO number six, in particular, appeared to be very stable. However, the performance of EMXO number six started to deteriorate about two months into the test phase, and eventually failed. This failure left five oscillators remaining for the duration of the tests. The sequence of tests performed is shown in Figure 3. The details of the individual tests are described below:

Baseline: In order to establish a datum from which to determine the frequency sensitivity of the oscillators to environmental stress conditions, a set of measurements was first performed in a benign (unstressed) situation. Even in a benign environment there will be changes in the output frequency of an oscillator due to the internal effects of aging. Random short-term fluctuations and even permanent frequency shifts, with no apparent correlation to operating or environmental conditions, may also occur in an unstressed environment. This can mask frequency sensitivity to some environmental stimuli, and such random frequency shifts may be indicative of an undesirable unit for real time frequency compensation.

As shown in Figure 3, commencing in August 1992, the tested EMXOs were operated under intermittent conditions (which were four hours on, followed by 20 hours off per day) for more than six weeks to provide retrace data. That was then followed by seven weeks of continuous operation (on for 24 hours per day). Following the continuous operation, the testing proceeded with six days of intermittent operation before the oscillators were powered down and moved to the environmental test facilities where the remaining tests were conducted. All the tests up to this point were performed under laboratory ambient conditions at a room temperature of about 25°C. The continuous operation was selected to measure frequency stability for comparison against the oscillator data sheets, and to establish a baseline for temperature cycling and vibration testing.

Cold Storage: The cold storage tests were conducted in the thermal chamber using the same test sequence as for intermittent operation (4 hours on / 20 hours off) except that the chamber was cooled down to -40°C for 16 of the 20 hours between tests. This left two hours for natural oscillator cooldown from the end of the test until ramping down the chamber temperature, and another two hours for the oscillator to stabilize at 25°C before power-on and start of test.

Hot Storage: These tests were conducted in a similar manner to the cold storage tests except that the chamber temperature was ramped up from +25°C to +95°C (instead of being ramped down to -40°C) for 16 hours for the storage period between tests.

Environmental Stress: This test sequence included: vibration, temperature cycling, and temperature cycling under vibration. In order to minimize the impact from retrace and any other sporadic frequency perturbation mechanisms, the oscillators were continuously powered for the duration of the environmental stress testing. The temperature profile and the vibration levels were selected to yield data suitable for inputs to construct neural network models for those oscillators. The resulting temperature profile as measured at the

oscillator case, is shown overlayed the frequency data for the temperature cycling tests. The vibration history is shown in Table 1. For each test in Table 1, data is collected continuously for a period of seven hours.

Table 1. Environmental Stress Tests

TEST #	LEVEL (g_{rms})	TEMP (°C)	DATE (1993)	Time of Vibration (UTC)
1	1.00	25	Mon. 6/14	17:00:00 - 18:00:00
	5.00	25	Mon. 6/14	19:00:00 - 20:00:00
2	3.00	25	Tue. 6/15	14:30:00 - 16:30:00
	8.00	25	Tue. 6/15	17:50:12 - 19:50:12
3	10.00	25	Wed. 6/16	12:15:00 - 13:15:00
	0.33	25	Wed. 6/16	16:00:00 - 17:00:00
	13.50	25	Wed. 6/16	18:00:00 - 19:00:00
4	5.00	Temp. Cycling	Wed. 6/16	20:10:00 - 02:50:00
5	1.00	Temp. Cycling	Thu. 6/17	13:16:30 - 19:50:00
6	3.00	Temp. Cycling	Fri. 6/18	12:05:00 - 19:50:00
7	10.00	Temp. Cycling	Mon. 6/21	16:01:00 - 16:39:10
				19:40:00 - 20:13:00
8	13.50	Temp. Cycling	Tue. 6/22	16:10:00 - 16:40:00
				19:43:00 - 20:13:00
9	No Vib	Temp. Cycling	Wed. 6/23	12:00:00 - 19:00:00
10	No Vib	Temp. Cycling	Thu. 6/24	12:00:00 - 19:00:00

Aging

Plots of fractional frequency versus time were generated for five EMXOs. Figure 4 depicts the results for three of those five units. As shown in that figure, there is a variability in drift from one oscillator to the next. The plot of each oscillator's performance shows a combination of the short-term fluctuations and long-term aging for the test sequence shown in Figure 3. Also, Figure 4 shows large day-to-day frequency changes while the EMXOs were at rest following the vibration and combined vibration/temperature tests. Frequency fluctuation measurements were made during and immediately after all imposed environmental stress conditions. However, the data shown in Figure 4 represents only the frequency fluctuations following the removal of the stresses in order to be consistent with the aging measurements depicted in the rest of the figure.

Cold and Hot Temperature Storage (Aging/Retrace)

Figures 5 and 6 show, on an expanded scale, sample retrace measurements from the periods of cold and hot storage, respectively. Each figure covers a two week period, resulting in four weeks of contiguous retrace measurements. That four week period includes retrace measurements done at cold storage, 25°C, and hot storage. The retrace measurements done at cold storage

are shown in the first portion of Figure 5, followed by the retrace at 25°C shown on the latter portion of Figure 5 and the first portion of Figure 6, concluded by the retrace at hot storage shown on the latter portion of Figure 6. The frequency stabilization performance at 25°C storage, which is shown on the second half of Figure 5 (starting on Feb. 7, 1993) and the first half of Figure 6 (ending Feb. 19, 1993), is in close agreement with the initial retrace measurements made in the laboratory environment. That indicates repeatable performance for the same operating condition. For cold and hot storage, the general trend appears to be more of a frequency change during the first few hours of operation following hot storage; whereas, operation following cold storage resulted in oscillator frequency stabilization similar to or better than the 25°C storage condition.

When one views all the data presented in Figure 4, the long-term aging trend when the oscillators are exposed to hot and cold storage does not appear to be significantly different from the initial baseline measurements. In the short-term, there was a reversal in the aging rate at 25°C immediately following hot storage.

Environmental Stress Testing

Figure 7 depicts the overnight frequency measurements where the oscillators stabilized between the vibration, the temperature, and combined vibration-temperature tests. EMXOs numbers two and three appear to be fairly insensitive to vibration, whereas EMXO number one appears to be highly sensitive to vibration. The figure demonstrates that different units of the same model oscillator are affected differently by vibration.

In particular, the data taken 20 through 23 June 1993 shows a significant frequency shift due to the high vibration levels of 10 to 13.5 g_{rms} during the combined high vibration/temperature cycling for all three EMXOs. Also, on 16 June, EMXO number one experienced a 2.5 x 10^{-8} permanent frequency offset due to the 10 g_{rms} random vibration input, whereas the other two EMXOs appear to have been relatively unaffected by that same input. Thus, it appears that the internal structural characteristics of a crystal oscillator may change when exposed to vibration above a certain level, which is unit dependent.

Vibration Alone

Prior to performing the vibration tests, a special fixture was designed and built. The six EMXOs were mounted on the fixture, which was fitted to the shake table head. The vibration tests were conducted with random vibration levels ranging 0.1 g_{rms} to 13.5 g_{rms}. The frequency of the oscillators was recorded as a function of the applied vibration levels. The vibration stimuli were removed for a period of one hour between the end of one vibration test and the application of a different vibration level for the next test.

The results of the vibration alone tests for three of the EMXOs are depicted in Figure 8. The frequency of EMXO number one was extremely sensitive to high levels of the random vibration, exhibiting peak-to-peak frequency fluctuations of up to 9 x 10^{-8} at 13.5 g_{rms}. In an oscillator screening process, that particular oscillator should be discarded. The other two oscillators exhibited much smaller changes in frequency (less than 1.5 x 10^{-9}) to the same set of vibration stimuli. The changes in frequency of EMXOs numbers two and three due to vibration could easily be compensated for in a Kalman Filter/Neural Network model.

Vibration and Temperature Combined

Following completion of the vibration only testing, a sequence of combined vibration-temperature tests was performed in which both vibration and temperature

stimuli were simultaneously applied. The vibration shake table with the six mounted EMXOs was placed in the Thermotron heating chamber. The temperature soaking cycle was programmed into the heating chamber controller, and the vibration levels applied were the same as in the previous vibration only tests. Frequency drift curves were generated for each of the set of test conditions.

Figure 9 shows the results of the second test performed on 16 June 1993, where a vibration level of 5 g_{rms} was combined with a temperature cycle profile. The frequency outputs of EMXOs numbers two and three exhibit a clear dependence on the temperature profile. Also, those two oscillators exhibited short-term frequency stabilities within $1x10^{-9}$. On the other hand, EMXO number one exhibited a short-term stability of greater than $2x10^{-8}$. Again, as noted in the discussion of the vibration only test results, in an oscillator screening process, EMXO number one should be discarded.

Figure 10 shows the results of the tests performed on 17 June 1993, where a vibration level of 1 g_{rms} (rather than the 5 g_{rms} of the previous day) was combined with the temperature cycle profile. The results for EMXOs numbers two and three are very similar to those shown in Figure 9 for the 5 g_{rms} vibration level. However EMXO number one, when exposed to the less severe vibration level of 1 g_{rms}, showed a significant improvement of greater than an order of magnitude in its short-term frequency stability (1×10^{-9} versus 2×10^{-8} for the 5 g_{rms} vibration level).

Test Results Summary

The EMXO performance measurements made during the EMXO evaluation phase have confirmed some undesirable performance characteristics seen in earlier tests. In particular, the EMXO output frequency can change suddenly when under vibration, and also for some period of time while at rest following removal of the vibration stimuli. The test results seem to indicate that the units which are the most sensitive to vibration are the same units that exhibited sudden frequency changes in a benign laboratory environment. The size of the frequency jumps on some of the units was several times larger than the long-term frequency drift accumulated over one year due to oscillator aging.

The relatively large frequency jumps appearing at random and the frequency stabilization during the first few hours of operation after power turn-on made it difficult to precisely determine the oscillator aging. A further complication in verifying aging performance is that the oscillator data sheets specify aging for continuous operation, while most of the EMXO measurements were made under intermittent operation. However, the measured EMXO aging appears to be well within specification.

The frequency stability measured over temperature was in close agreement with the EMXO data sheets. Also, visual interpretation of the test results did not reveal any change in the frequency versus temperature profile when the oscillators were exposed to vibration, except for the frequency jumps and instabilities that appeared to occur at random during the environmental stress testing.

This EMXO evaluation phase considered only the average EMXO output frequency stability for time-keeping applications. Neither was the spectral purity measured nor was the instrumentation calibrated to precisely measure frequency stability below 20-second averaging times.

Conclusions

A description was provided above of both the unstressed and stressed tests performed on six EMXOs. The environmental stressed conditions consisted of cold

storage, hot storage, vibration, temperature cycling, and combined vibration/temperature.

The experimental data collected during these tests represent an extensive database, which will be useful in characterizing crystal oscillators (EMXOs) under stressed environmental conditions. This database forms the basis of the "training" effort for our Kalman Filter/Neural Network model in order to compensate for the aging, for the stressed environmental effects, and possibly the retrace effects of the oscillator.

Both the frequency fluctuations during stabilization and the retrace effects experienced by the oscillators as a result of the cold storage measurements appear to be repeatable and similar in terms of magnitude to those observed at 25°C. On the other hand, the frequency fluctuations during stabilization following hot storage conditions were much larger in amplitude, but still repeatable.

There is a strong correlation between the output frequency of the oscillators and the stressed input conditions of temperature cycling and combined vibration/temperature. The frequency vs. temperature data for vibration levels less than 5 g_{rms} barely shows any dependence on the level of random vibration. We believe that for the case of temperature effects on the oscillators, we can, with a very high degree of confidence, compensate for the frequency fluctuations.

Our data shows that those oscillators which experienced significant frequency changes as a result of the higher vibration levels were the same ones that experienced similar frequency sensitivity at benign conditions. This will be extremely valuable in designing screen-while-train networks that automatically discard oscillators with excessive random frequency jumps, which would then be followed by stressed-level input training of the remaining "accepted" units.

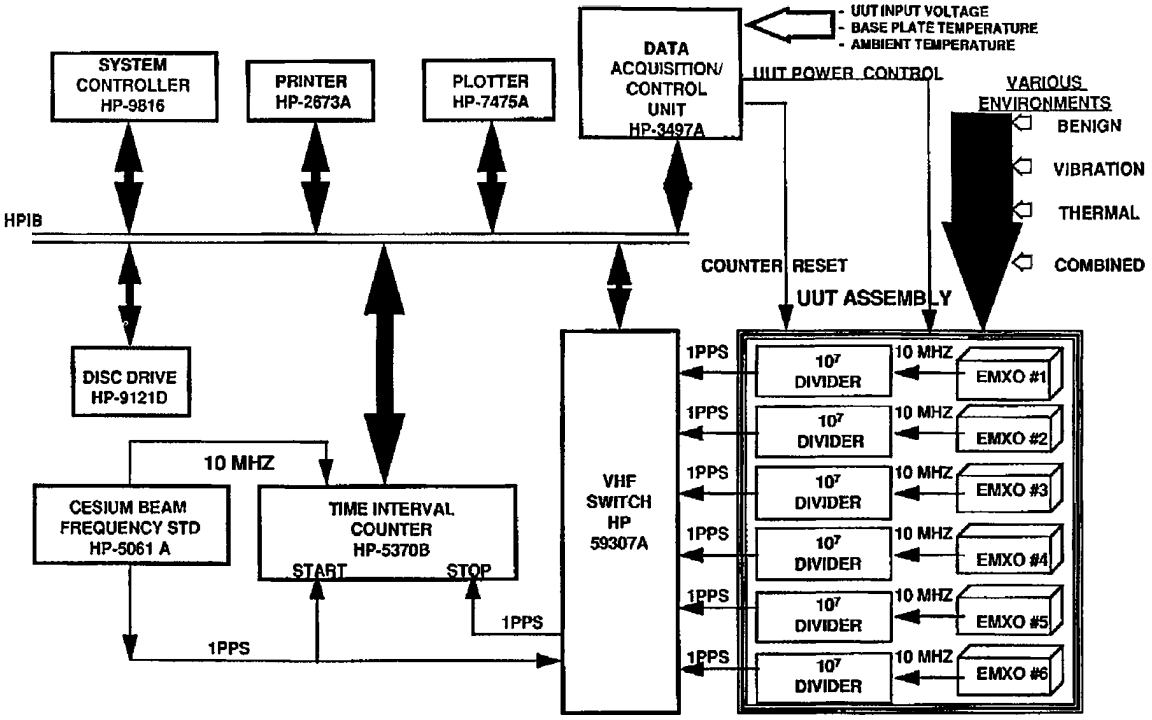


Figure 1. Test Equipment Configuration Indicating the Measurement Setup with the Six UUTS

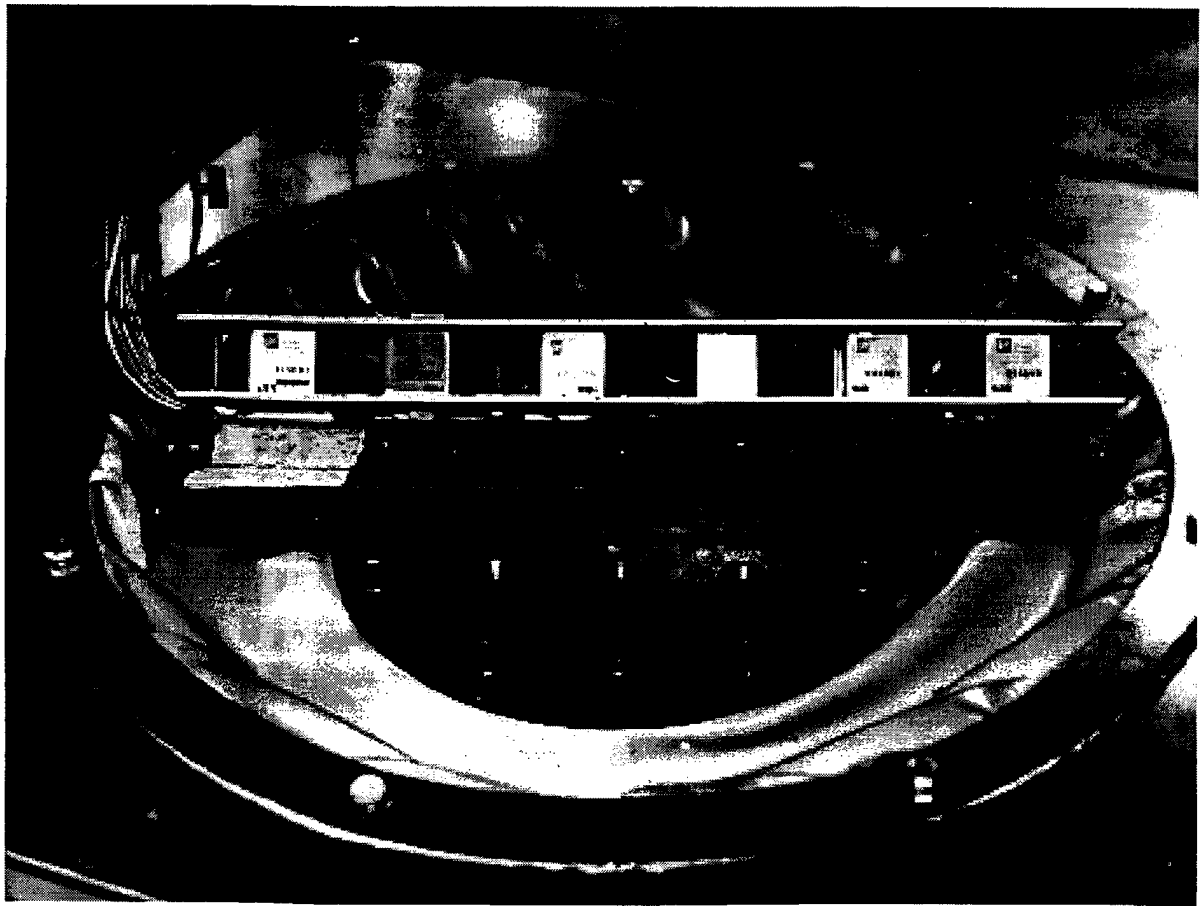


Figure 2. Test Bed Configuration

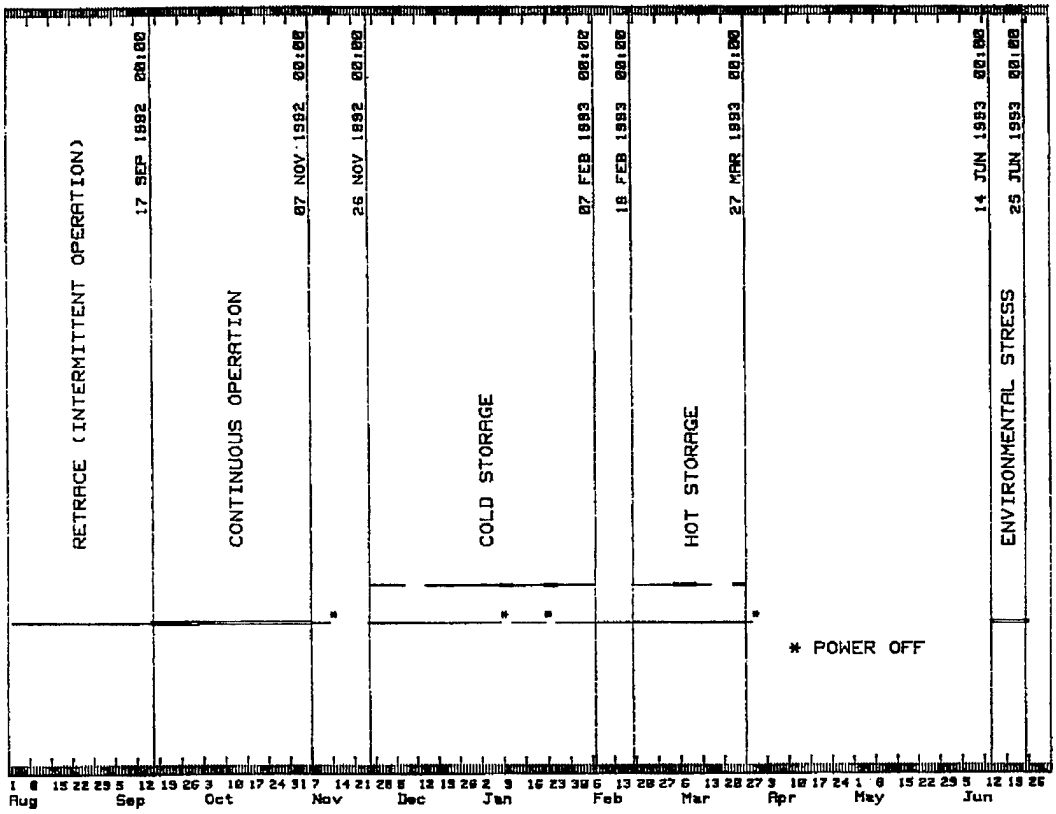


Figure 3. Test Sequence

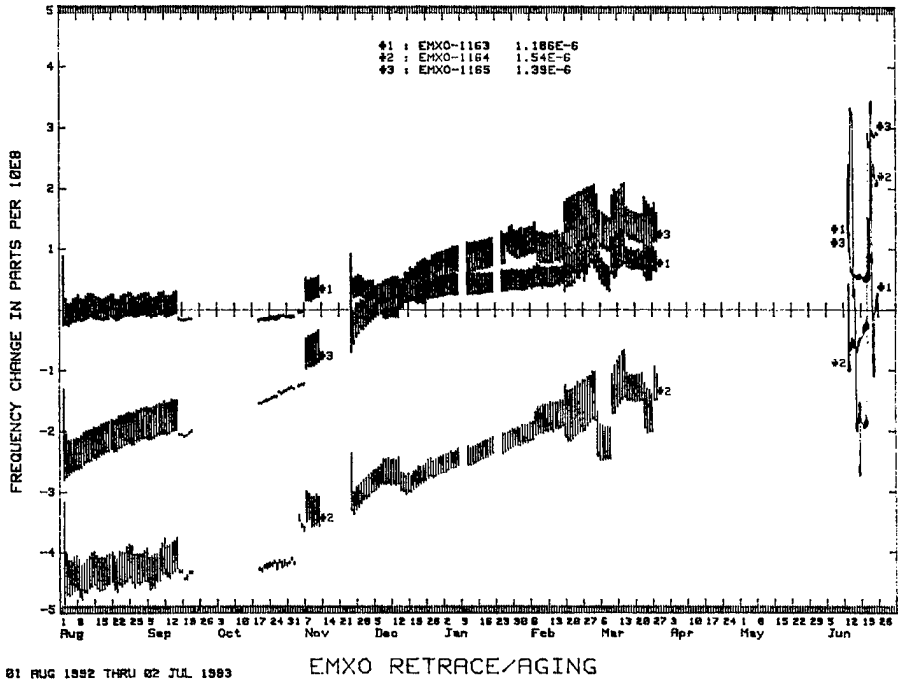


Figure 4. Long-Term Frequency Stability

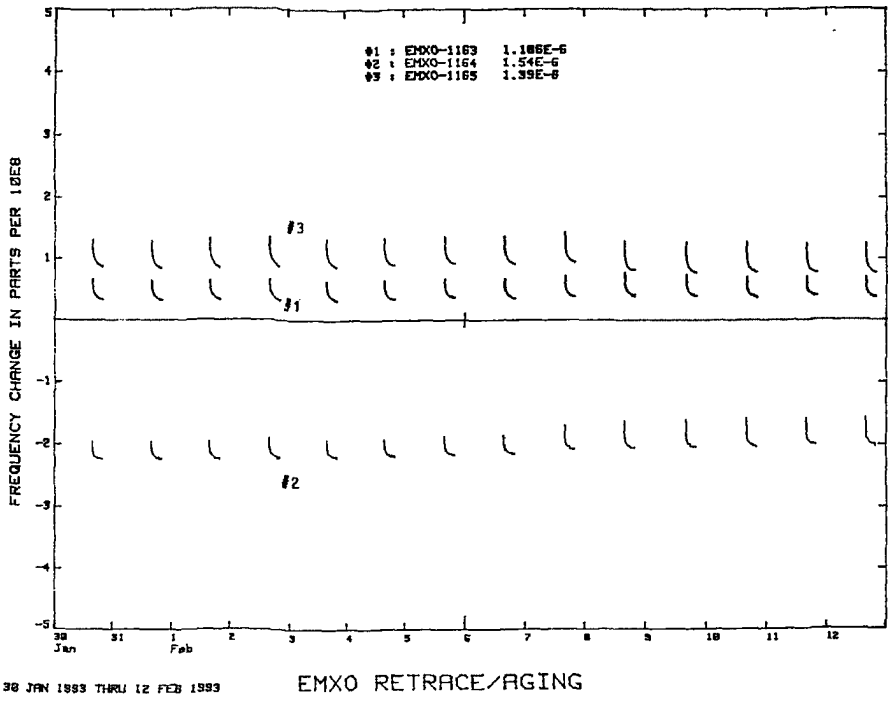


Figure 5. Retrace Oscillator Performance after Cold Storage

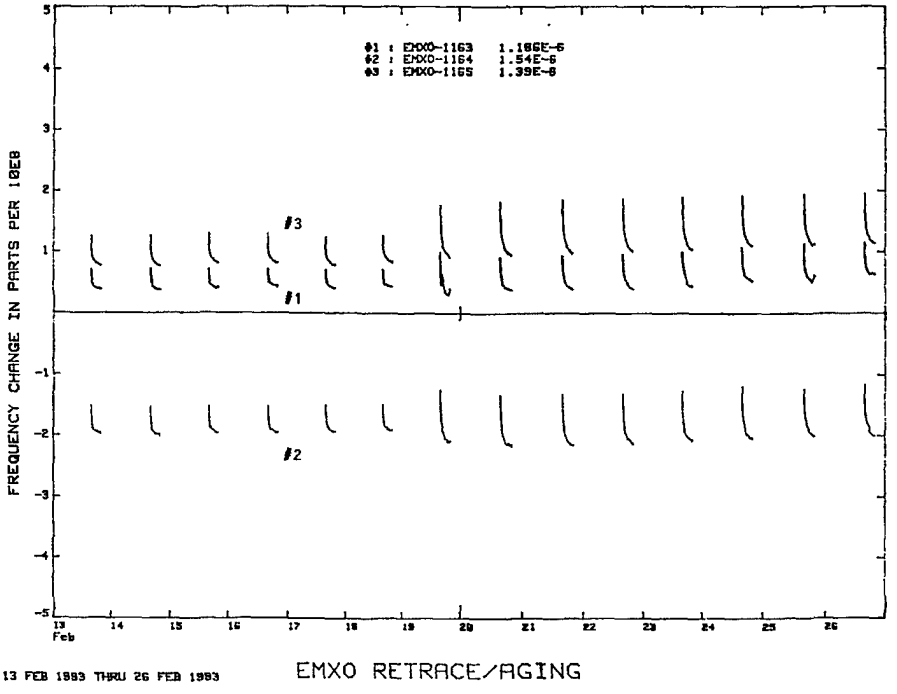


Figure 6. Retrace Oscillator Performance after Hot Storage

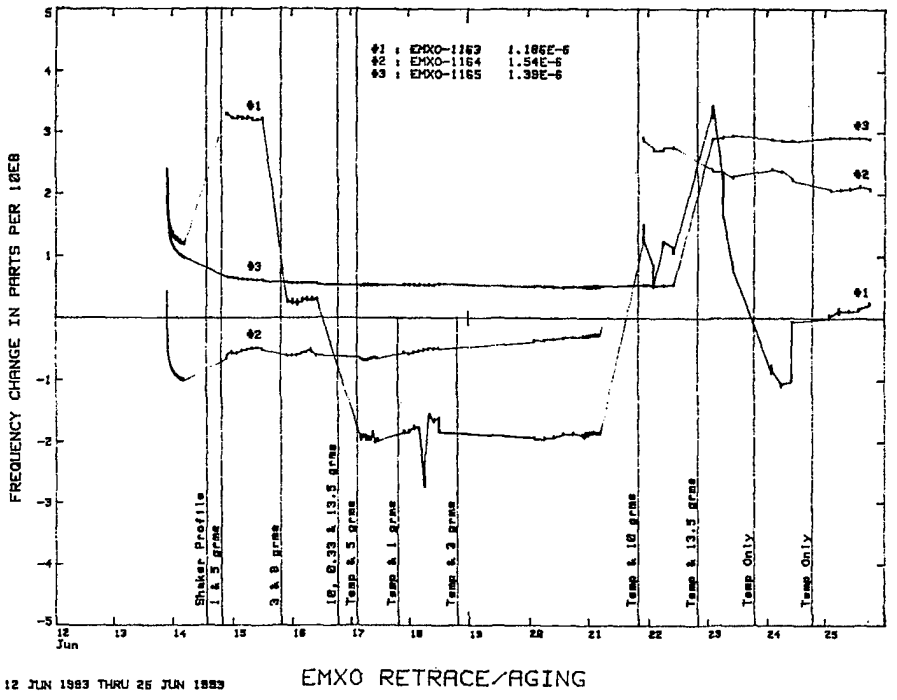


Figure 7. Frequency at Rest between Environmental Stress

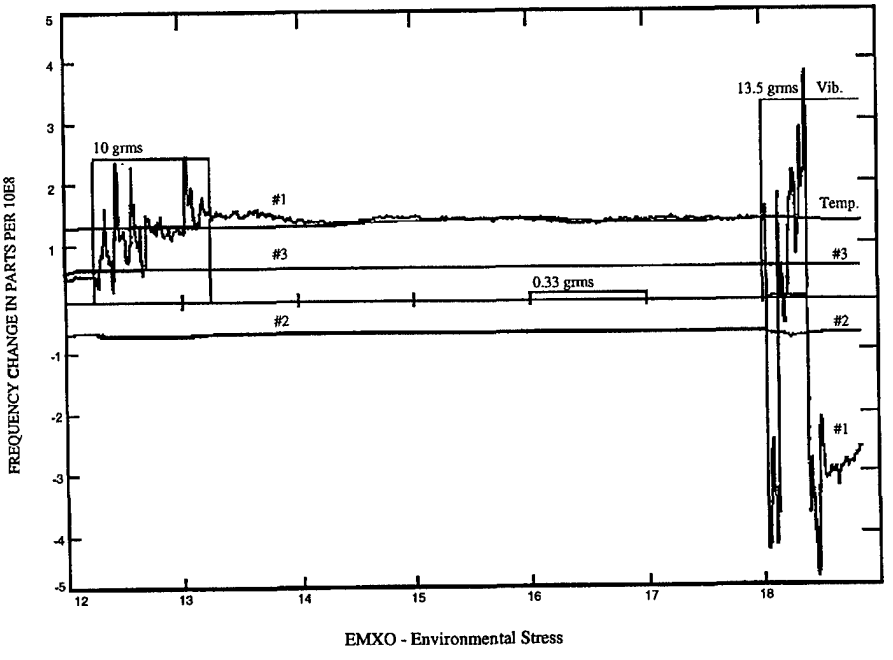


Figure 8. Frequency vs Vibration

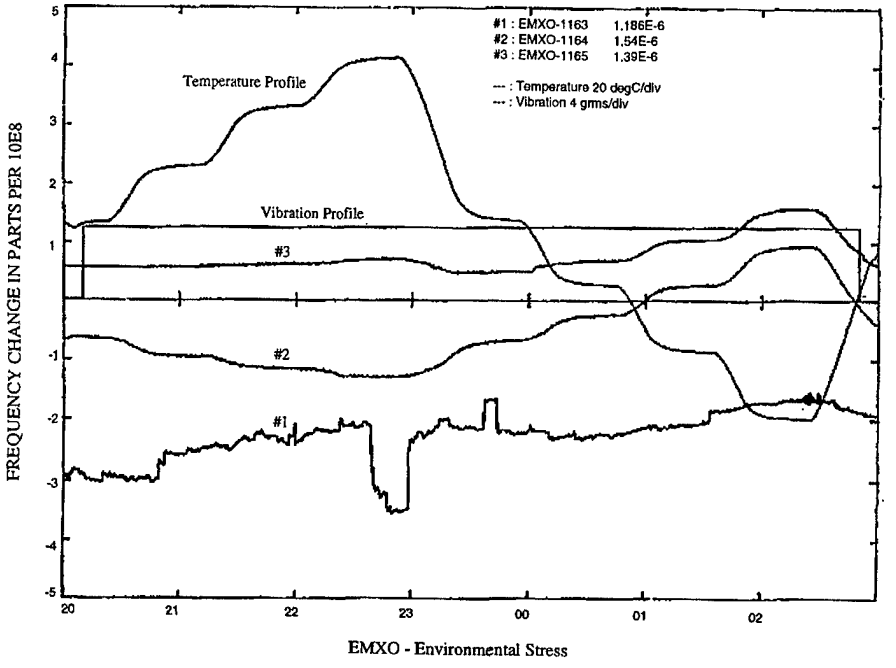


Figure 9 Frequency vs Temperature at 5 g_{rms} Vibration Level

16 June 1993 20:00:00 to 03:00:00

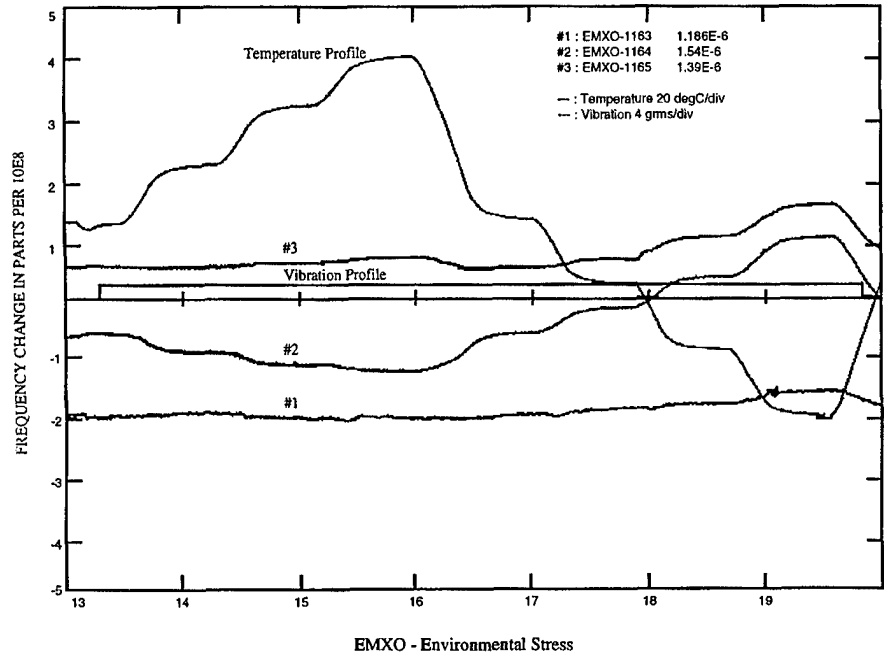


Figure 10. Frequency vs Temperature at 1 g_{rms}

17 June 1993 13:00:00 to 20:00:00

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

AN ULTRA-MINIATURE RUBIDIUM FREQUENCY STANDARD WITH TWO-CELL SCHEME

Y. Koyama, H. Matsuura(*), K. Atsumi(*), Y. Nakajima, and K. Chiba

FUJITSU LIMITED, FUJITSU TOHOKU DIGITAL TECHNOLOGY LTD(*).

1015 Kamikodanaka, Nakahara-KU, Kawasaki, 211, JAPAN

Summary

An ultra-miniature, general-purpose rubidium gas cell frequency standard with a two-cell scheme has been successfully developed, with the aim of developing one of the smallest atomic frequency standards in the world.

The new rubidium frequency standard embodies new electrical and structural design features, such as a miniature physics package incorporating a cylindrical TE111 microwave cavity in which an integrated filter-resonance cell is entirely loaded, a 10MHz ultra-miniature slave oscillator using voltage controlled TCXO(VC-TCXO) technologies established in the mobile telecommunication field, state of-the-art electronic circuit/assembly technologies, and a simplified, fabrication-oriented structure.

As a result of these features, the new rubidium standard has an extremely small volume(310 cc), light weight(700 grams), and a wide operating temperature range (-10 degree C to +60 degree C).

A series of measurements have been made to characterize the performance of the new rubidium frequency standard. In spite of the substantial reduction in size, it exhibits excellent performance, comparable to that of existing high performance models of commercial rubidium frequency standards, including previous models which Fujitsu has developed.

Presented in this paper are the details of the design concepts and electronic performance of this newly-developed frequency standard.

Introduction

Rubidium frequency standards have already found many applications in the field and systems involving precision frequencies. Fujitsu has been producing several models of rubidium frequency standards for about 25 years. Since our miniature, high performance rubidium frequency standard, designated 5407A(now designated FC6011A), was successfully developed in 1981[1], a great many rubidium frequency standards have been rapidly introduced as practical frequency sources in a wide range of applications, such as TV broadcasting, telecommunications, and precision measurement systems in Japan. In TV broadcasting systems, frequency stability of transmission carrier and color sub-carrier has been required to use carrier bandwidth effectively, or to prevent carriers from disturbing one another in Japan. The development of our rubidium frequency standard has enabled TV systems to employ rubidium frequency standards as frequency sources. In communication systems, digital synchronization networks have been established, utilizing cesium frequency standards as the source of the reference clock. When the reference clock signals are supplied to each system through the clock path in this network, the quality of electrical performance, especially clock jitters or wander degrades by disturbance. To compensate for this poor performance, a network synchronization equipment using highly-stable digital processing Phase Lock Loop(PLL) has been introduced in each terminal office. Rubidium frequency standards play an important role as a high-performance slave oscillator in the PLL[2][3].

As rubidium frequency standards begin to become more prevalent, customers demand that they be more cost-effective and perform better than ever before. Fujitsu has continued to develop rubidium frequency standards, to produce the reductions in size and weight, the cost-saving designs, and

0-7803-2500-1/95 $4.00 © 1995 IEEE

improvements in electrical performance[4][5], needed to meet these customer demands.

In order to achieve a compact design, which is one of preconditions to realize an attractive model, it was necessary to reduce the size of the physics package, which occupies a major part of the volume. For this reason, the excitation mode of the microwave cavity and the loading glass cell (resonance cell) located inside the cavity, which acts as a dielectric, were studied.

Simultaneously, electronic circuits were redesigned to achieve a low parts count, simplified, cost saving design, without sacrificing performance.

In this study, the Fujitsu general-purpose rubidium gas cell frequency standard, which is one of smallest in the world, will be discussed emphasizing the reduction in size and the improvement in electrical performance which has been achieved.

Design

Figure 1 is a simplified block diagram of the new rubidium frequency standard, designated model FC6031A. The system discussed in this study consists of three sub-assemblies:

1. The physics package, which serves as a very high-Q frequency discriminator to produce an error signal corresponding to the difference between the input microwave frequency and the intrinsic resonance frequency of rubidium atoms.
2. The LF circuit pack, including the servo-control circuit which produces the control voltage for steering the slave oscillator by analyzing the error signal from the physics package, slave oscillator, and a voltage regulator.
3. The HF circuit pack, including a frequency multiplier which generates a VHF signal from the output of the slave oscillator mounted on the LF circuit pack, and a frequency synthesizer which produces a fractional frequency necessary for synthesizing the interrogation frequency.

The mechanical structure has been simplified to achieve ease of assembly. On the physics package, both the LF and HF circuits are mounted closely to achieve high electrical performance. The physics package is also mounted on a base plate(heat spreader) to remove heat generated by the voltage regulator of the LF circuit. The LF circuit, the HF circuit, and the physics package are simply connected using a small number of wires, instead of conventional connectors. This simple physical design facilitates the assembly and test of the circuits.

Physics package

A microwave cavity containing the resonance cell accounts for a large part of the volume of the physics package. In order to realize a compact design of the physics package, the size of the microwave cavity itself shall be reduced.

First, we studied the excitation mode of the microwave cavity and selected TE111 mode because it was the most effective in a cylindrical microwave cavity and met our cavity specification in terms of minimum size. In Figure 2, the relationship between diameter and the resonance length of a cavity is depicted, which is excited in TE111 mode at 6.8 GHz, or a rubidium resonance frequency, assuming that the average dielectric constant is E 1.4. E 1.4 can be determined by a resonance cell composed of a glass in a cavity. The relationship between diameter and resonance length for the less desirable TM010 mode, which is actually independent of resonance length, as well as data calculated for cavity volumes of 10cc, 20cc, and 30cc, are also plotted in Figure 2. The hatched areas, depicting a cavity diameter of 32mm and resonance length of 26mm and a cavity diameter of 24mm, with a resonance length of 44mm, show that the cavity volume which can be realized is be about 20cc. The dimension of a cavity should be determined taking the following factors into consideration.

A. A diameter of a cavity shall be designed separately, not using the TM010 excitation mode.

B. The resonance length shall be long enough to obtain a relatively longer resonance region with a light from a rubidium lamp, resulting in a sufficient resonance signal.

We selected a cavity diameter of 24mm initially, in order to meet condition A above.

Next, our concern was the size of the resonance cell and its location in the cavity, because the amplitude and SNR of a resonance signal depend greatly on these factors. Considering size, glass thickness, and mounting location of the resonance cell as parameters, the amplitude and SNR of the resonance signal were discussed and experimented with. The size of the resonance cell, including glass thickness and resonance length

of the cavity were optimized, in addition to using a movable plate for frequency tuning of the cavity. Brass plated with aurum was selected as the cavity material, because it has high electric conductivity for the microwave cavity and has moderate heat capacity for the temperature oven to be able to prevent its temperature change from environmental temperature change.

Initial evaluations of the cavity have proved that the loaded Q(quality factor) is about 400, including the resonance cell and a photo cell with a hole for light transmission.

After a series of studies and experiments, a 24cc cavity was finally achieved, incorporating slight changes in the diameter and length. The resulting size of the physics package was reduced to 44 mm(W) x 82 mm(L) x 44 mm(H) (160 cc), which is 1/5 the size of our conventional package. It also includes a rubidium lamp exciter and a preamplifier, (which have also been reduced in size or parts count), heat insulators between the cavity and the lamp exciter, and a duplicated mu-metal magnetic shield.

Some important characteristics of the physics package were evaluated. The excellent SNR of the resonance signal of 64 dB per square root Hz, and the line width of 440 Hz at 6834 MHz have been obtained under conditions representing at reasonable operating conditions for the rubidium lamp and resonance cell, while connected to the electronic circuits described later.

Figure 3 is an exploded view of the physics package, along with the LF and HF circuit pack.

Electronic Circuits

The electronic circuits employ state of-the-art technologies.

(1)The LF circuit pack

The LF circuit pack includes a servo-control circuit, a voltage regulator, and a slave oscillator, which is usually mounted on the HF circuit pack, because of its small space margins.

One of the prominent features is the introduction of a 10 MHz ultra-miniature VC-TCXO(0.4 cc) as a slave oscillator, the technology of which has been established in the mobile telecommunication area. The original frequency stability of a VC-TCXO is not as good as that of a conventional high stability VCXO using a highly-stable quartz crystal unit. However, the frequency sweeper function in the LF circuit can permit frequency lock acquisition, resulting in the introduction of the VC-TCXO as a slave oscillator.

The servo-control circuit utilizes the combination of a synchronous filter and a synchronous DC restorer, which is based on sample-hold technology and serves as a filter having very sharp frequency-selective characteristics, without requiring an active filter composed of bulky electronic components. This technology proved that it has satisfactory frequency discrimination sensitivity and sufficient loop gain and has enabled the significant reduction in parts count of this circuit.

(2)The HF circuit pack

The HF circuit pack includes a multiplier chain and a frequency synthesizer on a printed circuit board.

The frequency multiplier chain, in spite of its compactness, synthesizes a high power(+ 18 dBm into 50 ohm), low noise VHF(90 MHz) signal multiplying 3 x 3 times the output frequency of the slave oscillator mounted on the LF circuit pack, with phase-modulation at low frequency(155 Hz) at the 30 MHz stage.

The frequency synthesizer uses a sampling phase-lock loop composed of a miniature VCXO(1.5 cc), and a programmable counter composed of a small size of gate-array. It permits the coarse frequency adjustment for this frequency standard.

(3)Other circuits

A preamplifier circuit board is contained inside the physics package. Its configuration, whose input impedance is extremely low, contributes to the low noise detection of the photo cell output, because this configuration derives a short circuit current of the photo cell without any forward bias voltage which is the source of shot noise.

All circuit packs have utilized surface mount devices as much as possible to reduce component mounting areas, not employing any new LSIC or hybrid ICs, except for a small sized gate-array in the HF circuit.

Packaging

Figure 4 shows the exterior view of the rubidium frequency standard. The 10 MHz output is on the rear panel. C field adjustment, which enables the fine frequency adjustment of the frequency standard, is accessible from the top panel.

Performance

Several important evaluation have been made to characterize the performance of the new rubidium standard.

Short-term Stability

Figure 5 shows the short-term stability of the FC6031A. The stability was measured by referring to a low-noise crystal oscillator for tau's of 10E-3 to 100 seconds, and by comparing two of the FC6031A's to each other for tau's longer than 100 seconds. The measurement was made in a benign environment.

The short-term frequency stability, which is one of the most important performance, depends on both the SNR and the line Q of the rubidium resonance[4]. Using the values described above, excellent short-term stability of 9.1 x 10E-12 per square root tau was calculated and 1.3 x 10E-11 was obtained experimentally. The difference between them is regarded as negligible.

Long-term Stability

The long-term frequency stability of a rubidium frequency standard is mainly dominated by that of the resonance cell.

Figure 6 shows the long-term stability plots for about 40 days. The aging rate is less than 2 x 10E-11 per month, which is as highly stable as that of conventional frequency standards, like the FC6021A, because high-vacuum processing techniques for the manufacture of a highly stable resonance cell have already been established.

Environmental Stability

Rubidium frequency standards are more or less sensitive to environmental changes. This sensitivity must be minimized to take advantage of the intrinsic stability of the standard. The FC6031A was subjected to changes in ambient temperature and magnetic field.

(A) Temperature

The temperature coefficients of both the resonance cell and the lamp cell can be easily controlled by changing the ratio of the mixture of buffer gases and the total buffer gas pressure in the resonance cell. In this study, a resonance cell and a lamp cell, made with conventional buffer gas mixtures and pressure, were utilized.

Figure 7 shows the effect of ambient temperature on the frequency of the FC6031A. We achieved a frequency stability of less than +/- 2 x 10E-10 at a temperature range of -10 to + 60 degree C, based upon the application of highly stable temperature-control technology and analysis to the temperature insulation between two cells and to improve heat removal. Further temperature stability will be achieved by optimizing the mixture ratio of buffer gases and pressure.

(B) Magnetic Field

Figure 8 shows the effect of the ambient magnetic field on the frequency of the FC6031A . The FC6031A was subjected to various orientations in the earth's field. The frequency change was less than 4 x 10E-11 for any orientation ,i.e., equal to about 1 gauss change.

Warm-up Characteristics

Figure 9 shows the results of warm-up and frequency retrace characteristics measurements of the FC6031A after a 24 hours period of off-time at 25 degree C. The control loop is established a approximately 6 minutes after being turned on subsequent to this 24-hour off-time. The frequency difference before and after power-on is within +/- 3 x 10E-10 at 10 minutes, +/- 3x 10E-11 at 30 minutes, respectively. These excellent characteristics were achieved by the heat reduction capacity of the microwave cavity, and the lamp house and by the heat insulated structural design of the physics package, all without increasing heating power at warm-up.

Supplied voltage Characteristics

Figure 10 shows the effect of changes in supplied voltage on the frequency . The FC6031A has a voltage regulator to reduce this effect, and no discernible frequency change was observed for +/- 10 % changes in the supplied voltage.

Technical Data

A summary of the technical data on the FC6031A is shown in Table 1.

Conclusion

In this paper, the details of the design of an ultra miniature rubidium frequency standard have been presented as well as the performance test results.

Many electrical and structural features have been incorporated into the design, which have resulted in an extremely compact, light weight product, with a wide operating temperature range, and low-power consumption.

The long-term stability was better than 2 x 10E-11 / month. The short-term stability measurement has shown that sigma Y(tau) = 1.3 x 10E-12 for tau = 100 seconds. The frequency stability is less than +/- 2 x 10E-10 at the temperature range of -10 to +60 degree C. The effect of the ambient magnetic field is less than 4 x 10E-11 for any orientation in the earth's magnetic field.

These results demonstrate that no performance has been sacrificed for the small volume achieved, and that the new rubidium frequency standard is promising in a wide range of applications in the field.

Acknowledgment

The authors would like to thank Messrs. A. Fukuoka, and H. Tanizawa of FUJITSU LIMITED, and Z. Osawa of FUJITSU TOHOKU DIGITAL TECHNOLOGY LTD. for their guidance and encouragement throughout the work.

References

[1] T.Hashi, et al., "A miniature high-performance rubidium frequency standard", Proc. 35th Annual Symposium on Frequency Control, pp.646-650, 1981.

[2] H.Tsuji, et al., "Clock supply module using rubidium oscillator", National Convention Record of IECE of Japan (in Japanese), No.B-336 autumn, 1989.

[3] K.Chiba, "Atomic frequency standards indispensable for digital communication networks", Ultrasonic technology (in Japanese), No.10, Vol.7, pp.44-52, 1992

[4] K.Chiba, et al., "An Ultra-miniature Rubidium Frequency Standard", Proc. 39th Annual Symposium on Frequency Control, pp.54-58, 1985.

[5] Y.Koyama, et al., "High performance miniature rubidium frequency standard", National Convention Record of IECE of Japan (in Japanese), No.B-958 spring, 1995.

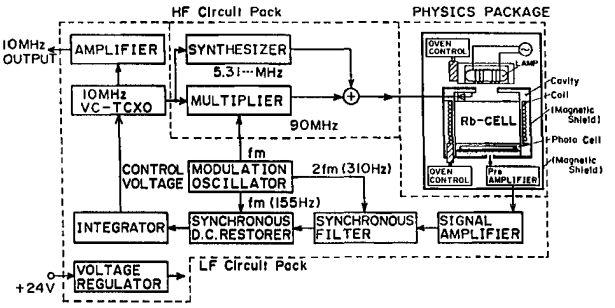


Fig. 1 Block diagram of the FC6031A

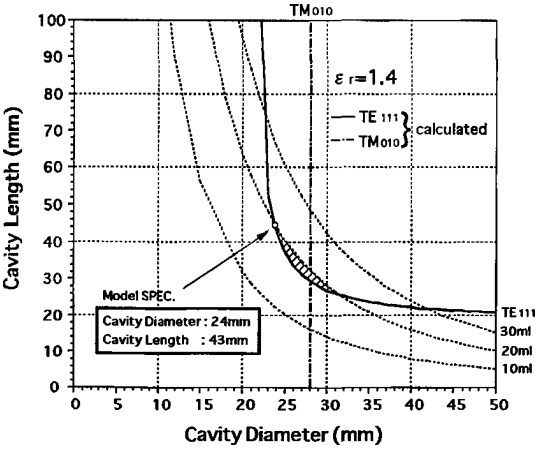


Fig. 2 Relationship between diameter and length of cavity

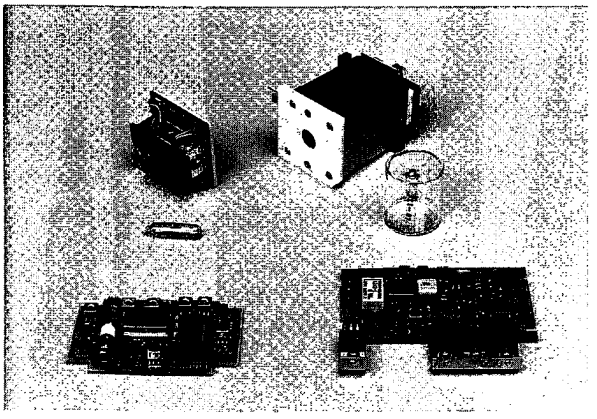


Fig. 3 Exploded view of the physics package with the LF and HF circuit pack

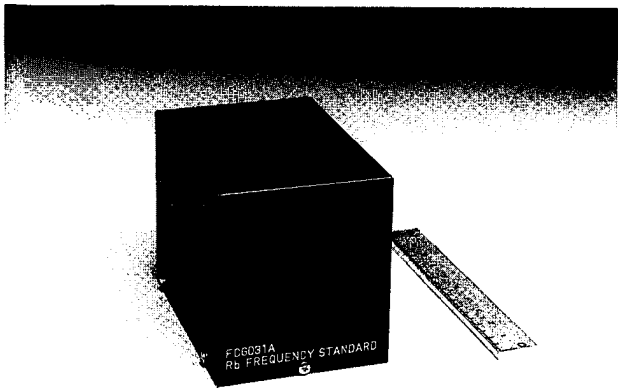


Fig. 4 Exterior view of the rubidium frequency standard

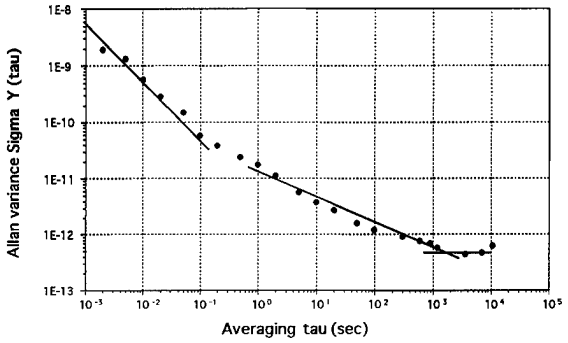


Fig. 5 Short-term stability

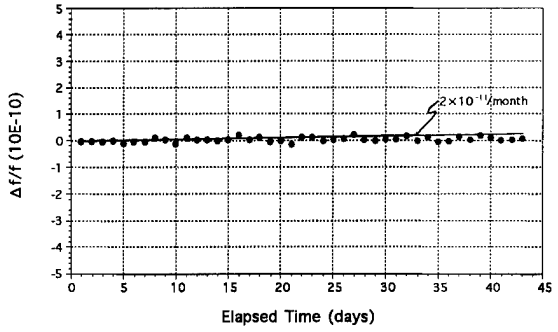


Fig. 6 Long-term stability

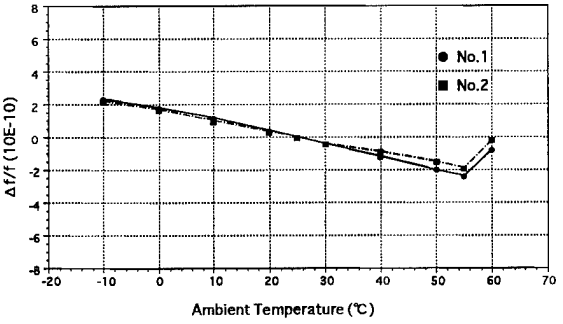


Fig. 7 Frequency versus temperature

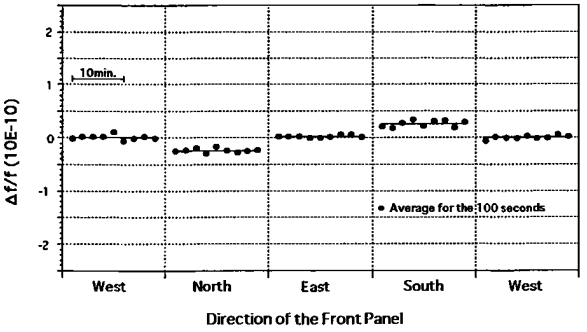


Fig. 8 Frequency versus orientation

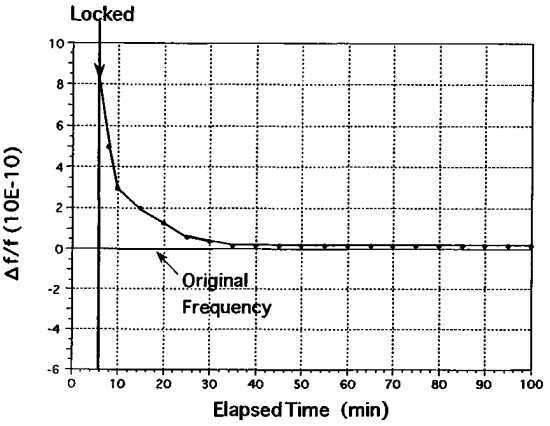


Fig. 9 Warm-up and frequency retrace characteristics

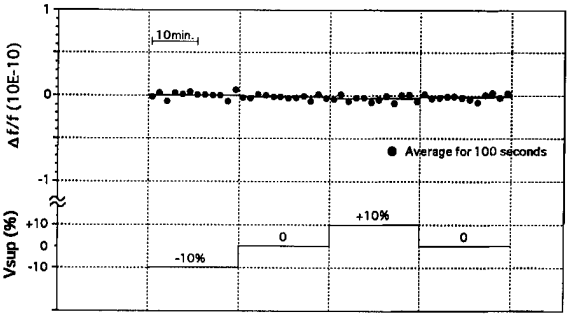


Fig. 10 Frequency versus supply voltage

1.Output

a)Frequency	: 10 MHz (sine wave)
b)Level	: 0.5 Vrms into 50 ohm
c)Frequency stability	
- Long-Term	: < 3 x 10E-11/month
- Short-Term	: Sigma Y(tau) < 1.3 X 10E-11/(root tau) for tau 3 sec. < tau < 100sec.
- Temperature	: < 2.5 x 10E-10/(25 +/- 35 degree C)
- Retrace	: < 3 x 10E-11 at 30 minutes after a turn-on at 25degree C
- Input Voltage	: < 1 x 10E-11(for +24 V +/- 10 %)
d)Time to lock	: 6 minutes(approximately)

2.Input

a)Voltage	: +24 V
b)Power	: 8 W (steady at 25 degree C) 22 W (during warm-up)

3.Size	: 58 mm(W) x 84 mm(D) x 64 mm(H)
4.Weight	: 677 grams (1.5 lbs)

Table 1. Technical data for the FC6031A

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SUBMINIATURE RUBIDIUM FREQUENCY STANDARD: MANUFACTURABILITY AND PERFORMANCE RESULTS FROM PRODUCTION UNITS

T. McClelland, I. Pascaru, I. Shtaerman, C. Stone, C. Szekely, J. Zacharski
Frequency Electronics, Inc.
55 Charles Lindbergh Blvd., Mitchel Field, NY 11553

N.D. Bhaskar
The Aerospace Corporation
P.O. Box 92957, Los Angeles, CA 90009

Abstract

The FE-5650 subminiature rubidium frequency standard is a 216 cc (13 cubic inch) device capable of producing any output frequency between 1 Hz and 15 MHz, with a setting resolution of 1×10^{-14}. This device was designed to provide a flexible, low cost means of achieving atomic frequency stability.

Typical applications for this type of frequency standard are presented, with a discussion of appropriate cost-performance trade offs to be considered. Design for manufacturability considerations which ensure the production of reliable frequency standards at the lowest possible cost are discussed. Performance results are presented, with a particular emphasis on the statistical distribution of key performance measures based on production samples. The data and analysis are obtained from measurements made at FEI, as well as at NIST, in Boulder, CO, and The Aerospace Corporation.

Introduction

At the 1993 Frequency Control Symposium, Frequency Electronics, Inc. presented a paper describing the design of a new family of rubidium vapor atomic frequency standards.[1] This family of devices, model FE-5650A, was designed to replace high precision quartz oscillators in non-military applications.

The basic design strategy behind the FE-5650A was to provide value to the user. Previous designs (primarily for military applications) emphasized the achievement of specific performance goals, independent of cost. With the FE-5650A, the approach was to provide atomic standard performance under conditions which would allow for the lowest possible cost, and smallest possible size.

The FE-5650A is shown in Figure 1. It measures 3.7 x 7.7 x 7.6 cm (1.44 x 3.03 x 3.00 inches), for a total volume of 216 cc (13.1 cu. in.). It weighs 338 grams (12 oz.), and typically dissipates 6.5 watts, steady state. It is available, at the time of this writing, for $1595 each, in quantity. An optional version of the FE-5650A, designed to operate in military environments (-55 to +71°C, airborne vibration), is also available. A detailed description of the FE-5650A design is provided in the 1993 FCS paper, referenced above.

In this paper, performance results of manufactured FE-5650A units are presented. First, the targeted performance niche for this type of frequency standard is discussed. This provides a context in which to evaluate the actual performance of manufactured units. Then, actual performance data measured on production units is presented. As much as possible, statistical data are provided, in an effort to highlight the performance range to be expected from production devices.

Performance Goals

For many applications, performance just a little better than a good quartz oscillator is required. Sometimes this can be provided by adding extra ovens to a quartz oscillator, or standby power to existing ovens in order to minimize the effect of thermal transient stabilization. However, this becomes increasingly difficult as stabilities in the 10^{-11} range are demanded; and the associated costs begin to increase exponentially. Yet even crude Rb vapor frequency standards achieve $\sim 10^{-11}$ performance. Therefore, a low cost Rb standard as a replacement for a high cost quartz oscillator makes sense.

The FE-5650A is designed to warm-up within 4 minutes, at which point the aging should be less than 2×10^{-11} per

0-7803-2500-1/95 $4.00 © 1995 IEEE

day, and the short term stability better than $3 \times 10^{-11}/\sqrt{\tau}$, out to averaging times of at least 100 seconds.

The availability of the GPS satellite navigation system has made it possible to obtain atomic frequency stability for the price of a GPS receiver. A very inexpensive (usually quartz) frequency source in the receiver can be disciplined by the GPS satellite clock ensemble. For time intervals between satellite updates, the receiver frequency source is free running, but can be corrected, if necessary, each time new satellite clock information is received. For many applications requiring precise time or frequency, the stability of a quartz oscillator is adequate during the time intervals between satellite updates. However, for some applications, the better short term stability (for 1 to 10,000 second averaging times) of a Rb atomic standard is necessary to achieve adequate performance during the time intervals between satellite updates. Critical applications, such as precision aircraft approaches, which must function in extenuating circumstances (such as the loss of satellite communication), can be greatly enhanced by improved short term stability of the receiver frequency source.

The FE-5650A is well suited to such critical applications. The short term stability specification of the FE-5650A, $3 \times 10^{-11}/\sqrt{\tau}$, approaches that of the GPS satellite ensemble. (As is shown in the following section, the actual short term stability performance of manufactured FE-5650A units is $<4 \times 10^{-12}/\sqrt{\tau}$; almost an order of magnitude better than this specification.) Therefore, the FE-5650A makes an excellent flywheel, providing stand-alone performance for short times, and being disciplined to the much better stability of GPS satellite frequency standards for long times. How often a disciplined FE-5650A needs to be updated by GPS depends on the application, of course; but also on its Allan variance performance at longer averaging times. Ideally, a disciplined frequency standard should be updated after intervals such that its free running performance just begins to be significantly degraded relative to the GPS satellite clock ensemble. Therefore, improved FE-5650A Allan variance at longer averaging times translates directly into longer free running periods before GPS updates are required, for a given level of required stability.

Performance of Manufactured Units

In this section performance of production FE-5650A units shall be presented. Data are presented for:

1. Frequency vs temperature
2. Allan Variance
3. Drift/aging
4. Phase noise

The Allan variance and aging data are the result of measurements made at the National Institute of Standards and Technology (NIST), Boulder Co.; The Aerospace Corporation, Los Angeles, Ca.; as well as Frequency Electronics, Inc., Mitchel Field, NY.

Frequency vs Temperature

Stability of frequency in the presence of significant temperature variations is a weak area for most commercially available Rb frequency standards. This is unfortunate since many commercial applications require the frequency standard to operate at remote sites with little or no temperature control. Typically, combined frequency variations due to random fluctuations, drift, and other environmental effects can be kept to $< 2 \times 10^{-11}$/day in a Rb standard, whereas the frequency variations due to common working environment temperature variations can easily contribute $\sim 1 \times 10^{-10}$/day. Control of frequency variations due to temperature is especially difficult in very small Rb standards, due to the inherently close proximity of many interacting elements in the physics package.

For these reasons, special effort has gone into the design of the FE-5650A in order to control frequency variations due to temperature. Furthermore, actual performance of manufactured units has been closely monitored; first of all so as to assure performance within quoted specifications ($\pm 3 \times 10^{-10}$ of nominal for $-5°C \leq T \leq 50°C$), and second of all as a first step toward continued performance improvements.

Measurements have been made on a sample of 236 production FE-5650A units, in 8 production lots. The mean shift in frequency due to a -5 to +50 temperature change is:

$$\langle \Delta f \rangle = -0.1 \times 10^{-10}, \quad (1.)$$

with a sample standard deviation of:

$$\sigma(\Delta f) = 2.4 \times 10^{-10}. \quad (2.)$$

Clearly, the standard deviation, $\sigma(\Delta f)$, is of much more relevance than the mean, $\langle \Delta f \rangle$ (which is essentially 0).

From a manufacturing point of view, half of the job consists of reducing the mean, $<\Delta f>$, to zero. Practically speaking, this result has been achieved, as evidenced by Equation (1.). The other half of the job, which is an ongoing process for the FE-5650A, consists of reducing $\sigma(\Delta f)$ to the smallest possible value. Needless to say, a great deal of process control effort is presently focused on reducing $\sigma(\Delta f)$, for the FE-5650A.

Frequency vs temperature performance for a particular FE-5650A unit is shown in Figure 2. The Figure shows actual data of frequency vs temperature, as well as a least squares fit of the data to a linear function. (It is difficult to make out the fit line in the Figure, since it is super-imposed almost exactly on top of the actual data - indicating a "good" fit.) The overall frequency shift is clearly not as good as some of the units, as indicated by Equation (1.), but is qualitatively representative of the performance of the manufactured units. In particular, the linearity of frequency vs temperature is typical of the FE-5650A units.

The linearity of frequency vs temperature suggests that internal electronic compensation might be feasible. In fact, a modified FE-5650A has been developed which accomplishes just this; specifically, reduction of the total frequency shift due to temperature to $\sim 1 \times 10^{-11}$. Prototype tests of this approach are in process, and an FE-5650A option incorporating this design will be available soon.

Allan Variance

As stated above, the original design concept for the FE-5650A was as a replacement for quartz oscillators. With this in mind, the Allan variance specification for the FE-5650A was established at:

$$\sigma_y(\tau) = 3 \times 10^{-11}/\sqrt{\tau}, \qquad (3.)$$

for $\tau \leq 100$ seconds. This requirement is typical of commercially available Rb standards.

It has turned out that Allan variance performance almost an order of magnitude better than this specification is routinely achievable with the FE-5650A units. This poses certain measurement difficulties, since very few commercially available frequency standards of any kind are capable of such performance. Therefore, measurements showing the true performance of the FE-5650A must be made using a carefully chosen reference, with adequate performance.

For this reason, arrangements were made to measure the FE-5650A at NIST, where the measurement system includes a hydrogen maser as the reference. The results are shown in Figure 3. As is evident from the Figure, the data are approximated by $\sigma_y(\tau) \cong 3 \times 10^{-12}/\sqrt{\tau} + 3 \times 10^{-15}\sqrt{\tau}$. It should be noted that the measurements at NIST were made with the FE-5650A in a laboratory environment (nominal temperature 25 °C, with ±3 °C variations).

Measurements were made on two additional units at the Aerospace Corporation. For these measurements, the FE-5650A units were compared to a Hewlett Packard 5061B Cs beam standard, as reference. These measurements were also made in a laboratory environment (nominal temperature 25 °C, with ±3 °C variations). The results are shown in Figures 4 and 5. Figure 4 shows raw frequency as well as Allan variance data for SN 1420, measured over a 4 day period of time. Figure 5 shows raw frequency and Allan variance data for SN 1438 measured over a 10 day period of time. The short term Allan variance results in these figures are limited by the performance of the HP5061B reference, which is specified at $\sigma_y(\tau) \cong 8 \times 10^{-12}/\sqrt{\tau}$. The longer term performance ($\tau \geq 1000$ seconds) is a true representation of the FE-5650A, since the Cs reference is much better than the Rb standard in this region.

From Figures 3 through 5 it is clear that the Allan variance exhibits a $1/\sqrt{\tau}$ dependence (white noise) for averaging times out to ~ 1000 seconds, and approximately a $\sqrt{\tau}$ dependence (random walk fm?) for averaging times > 1000 seconds.

Additional Allan variance measurements (not shown) have been made at Frequency Electronics, Inc. By sequentially using one FE-5650A unit as reference, and comparing to other FE-5650A units, it is possible to obtain a reasonable measure of the Allan variance, with the reference contribution eliminated (3 cornered hat). Using these measurements, plus the Allan variance results from NIST, (total sample size of 5 units), it is found that the white noise component of the Allan variance has a mean value of:

$$<\sigma_y(\tau = 1 \text{ sec.})> = 3.29 \times 10^{-12}, \qquad (4.)$$

with a standard deviation of $\pm 0.31 \times 10^{-12}$. Similarly, using the NIST measurement, the two measurements from The Aerospace Corporation, and one additional long term measurement at Frequency Electronics, Inc.

(total sample size of 4 units), it is found that the "random walk" component of the Allan variance has a mean value of:

$$<\sigma_y(\tau)> = 3.85 \times 10^{-15}\sqrt{\tau}, \quad (4.)$$

with a standard deviation of $\pm 0.35 \times 10^{-15}\sqrt{\tau}$. This result is based on analysis of the measurements at 12,800 seconds, and the assumption that the $\sqrt{\tau}$ dependence is correct.

Drift/Aging

Drift usually refers to the linear, monotonic component of frequency change measured over long (greater than 1 day) periods of time. This can be a somewhat vague concept with Rb frequency standards, since typically, a linear fit of frequency data gives different results depending on the length of the data, as well as the absolute time at which the data are taken.

In this paper, we consider frequency changes occurring over a time period of approximately one month. Typical performance for the FE-5650A is shown in Figure 6, for a 40 day period of time. (The data in the Figure are taken in a laboratory environment, with normal temperature and barometric pressure variations.) Similar data taken at the Aerospace Corporation (on a different unit) are shown in Figure 7. These data are also taken in a laboratory environment. As is evident from either Figure, per day drift rates vary significantly during the test period.

A useful measure of frequency aging is obtained by taking the simple difference between the frequency at the end of a ~1 month period of time and the frequency at the beginning of the period, divided by the length of the time period.

This type of analysis has been performed on 9 FE-5650A units (each measured for > 1 month). The result is:

$$<df/dt> = -4.1 \times 10^{-12}/\text{day}, \quad (5.)$$

with a standard deviation of $\pm 5.8 \times 10^{-12}$/day. This result shows that the FE-5650A units typically (although not always) age negatively.

Phase Noise

The FE-5650A uses a direct digital synthesizer (DDS), clocked to a direct sub-multiple of the Rb atomic frequency, to generate the RF output. This has the advantage of allowing virtually any frequency between 1 Hz and 15 MHz to be generated, with the full stability of the Rb atomic system. However, this makes the phase noise of the RF output dependent on the output frequency; and therefore it is not possible to provide a single plot representative of the FE-5650A phase noise performance. With this caveat, a phase noise plot of a 10 MHz FE-5650A production unit is shown in Figure 8. The phase noise specification for the FE-5650A at 10 MHz is:

10 Hz	-90 dBc
100	-125
1000	-145

Summary

The FE-5650A rubidium atomic frequency standard is a very small, inexpensive, precision frequency source. It was designed as an economical alternative to high precision quartz crystal oscillators. In this paper, actual performance of production FE-5650A units is discussed. It is found that short term stability, as measured by the Allan variance, is generally much better than the design specification. In fact, for averaging times out to ~1000 seconds, the performance is as good as the best satellite Rb standards currently available, but at around 1% of the cost of such devices! This excellent short term stability makes the FE-5650A an attractive choice for applications requiring a GPS disciplined atomic frequency standard, such as clock aided satellite navigation for precision aircraft approaches. In such applications, a disciplined FE-5650A can maintain frequency accuracy of $< 3 \times 10^{-13}$

Acknowledgments

The authors gratefully acknowledge the assistance of The National Institute of Standards and Technology, (NIST), Boulder, Co., and in particular Dr. Robert E. Drullinger, in making Allan variance measurements reported above.

References

[1] M. Bloch, et. al, "Subminiature Rubidium Frequency Standard for Commercial Applications," in Proceedings, 1993 IEEE International Frequency Control Symposium, 1993 p. 164

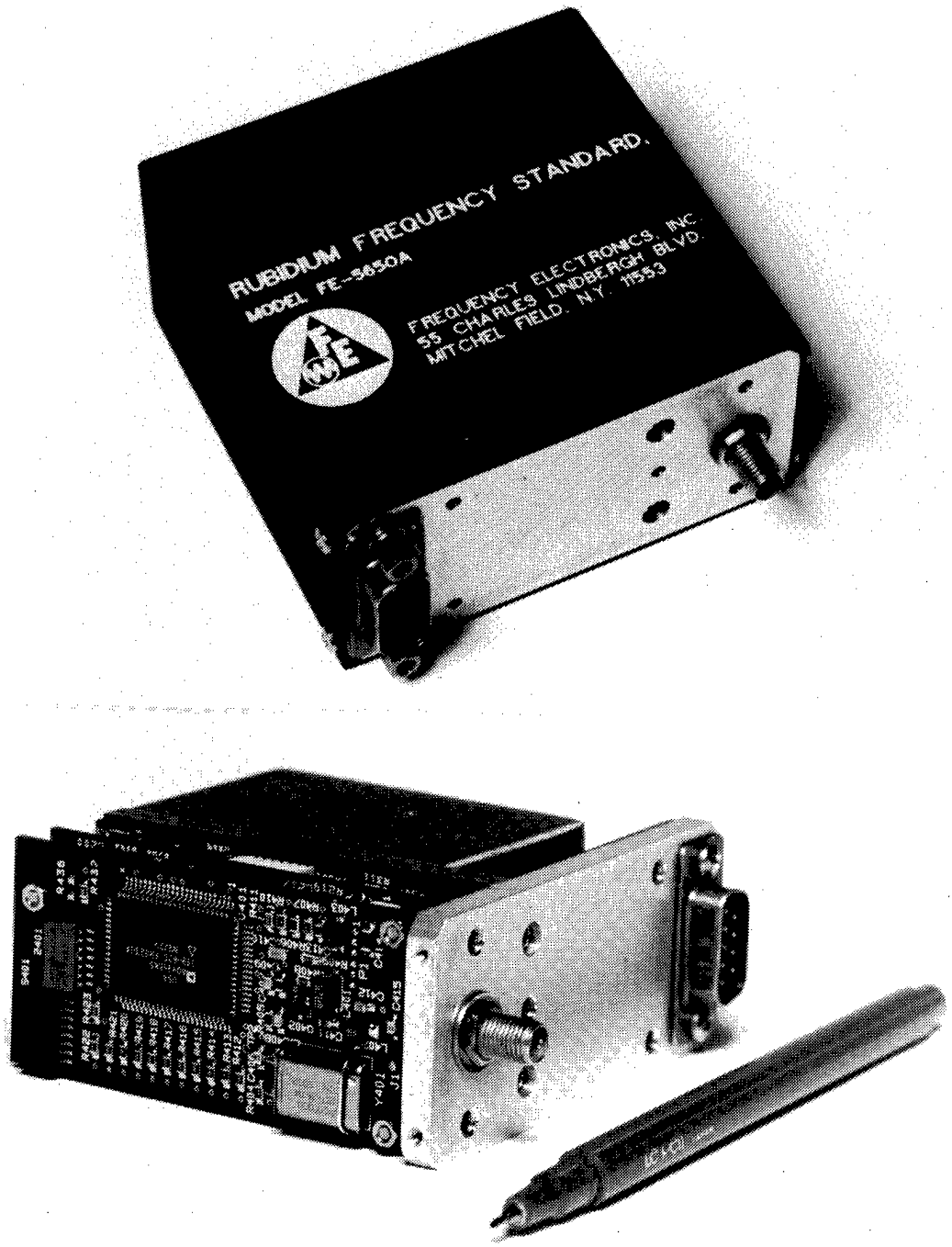


Figure 1
Photographs of FE-5650A Rb atomic frequency standard

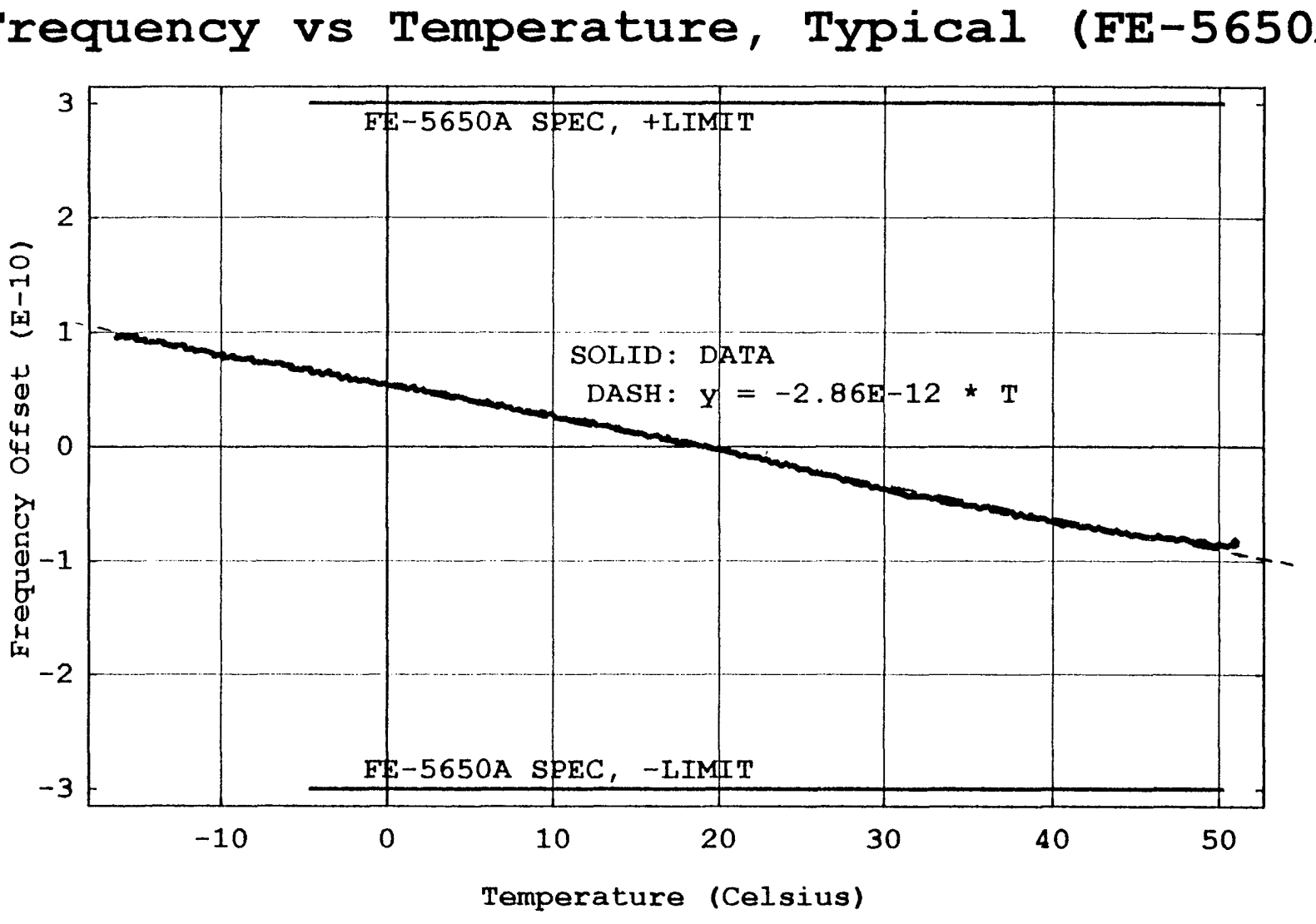


Figure 2
Frequency vs Temperature, FE-5650A

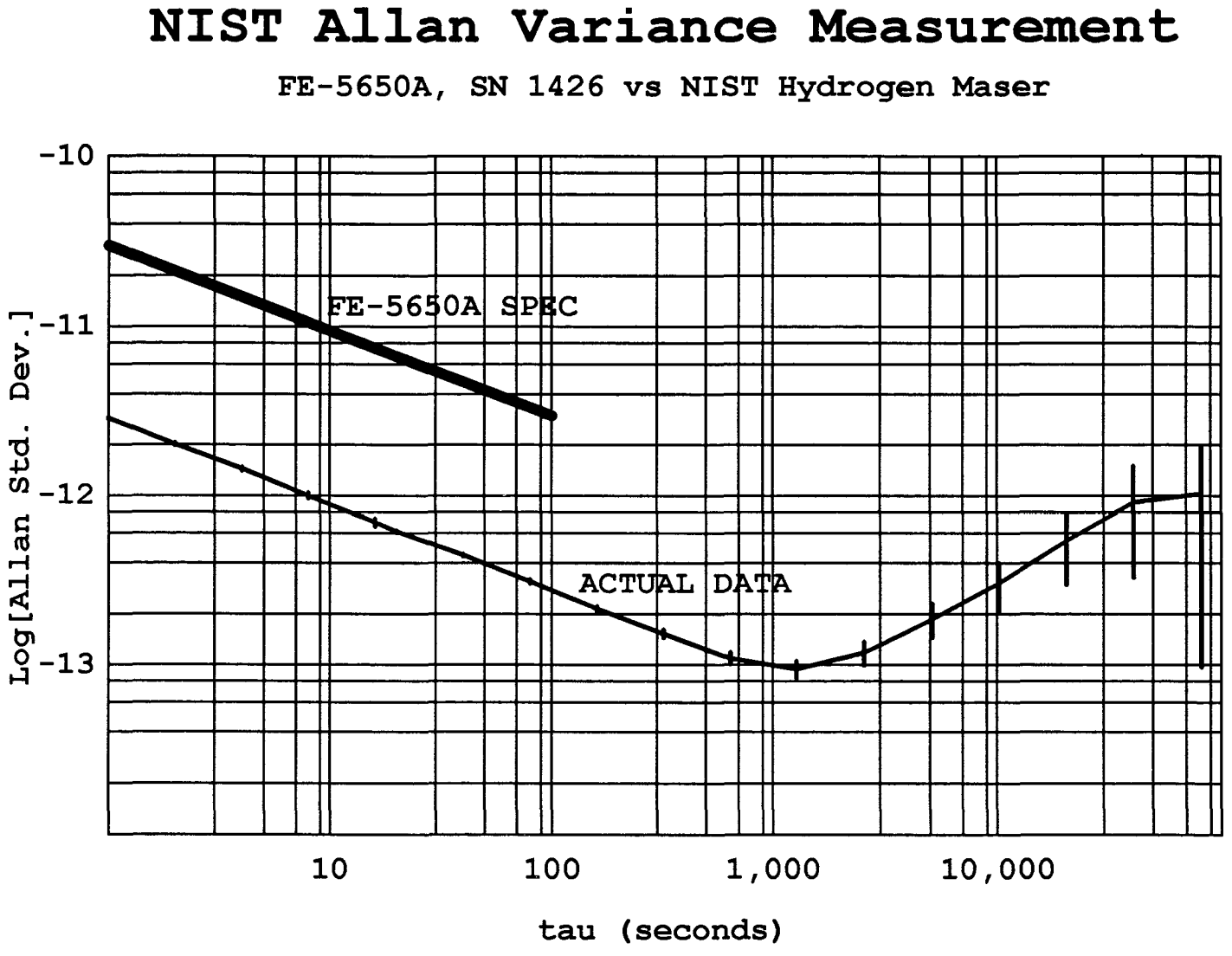


Figure 3
Allan variance data, FE-5650A vs NIST hydrogen maser

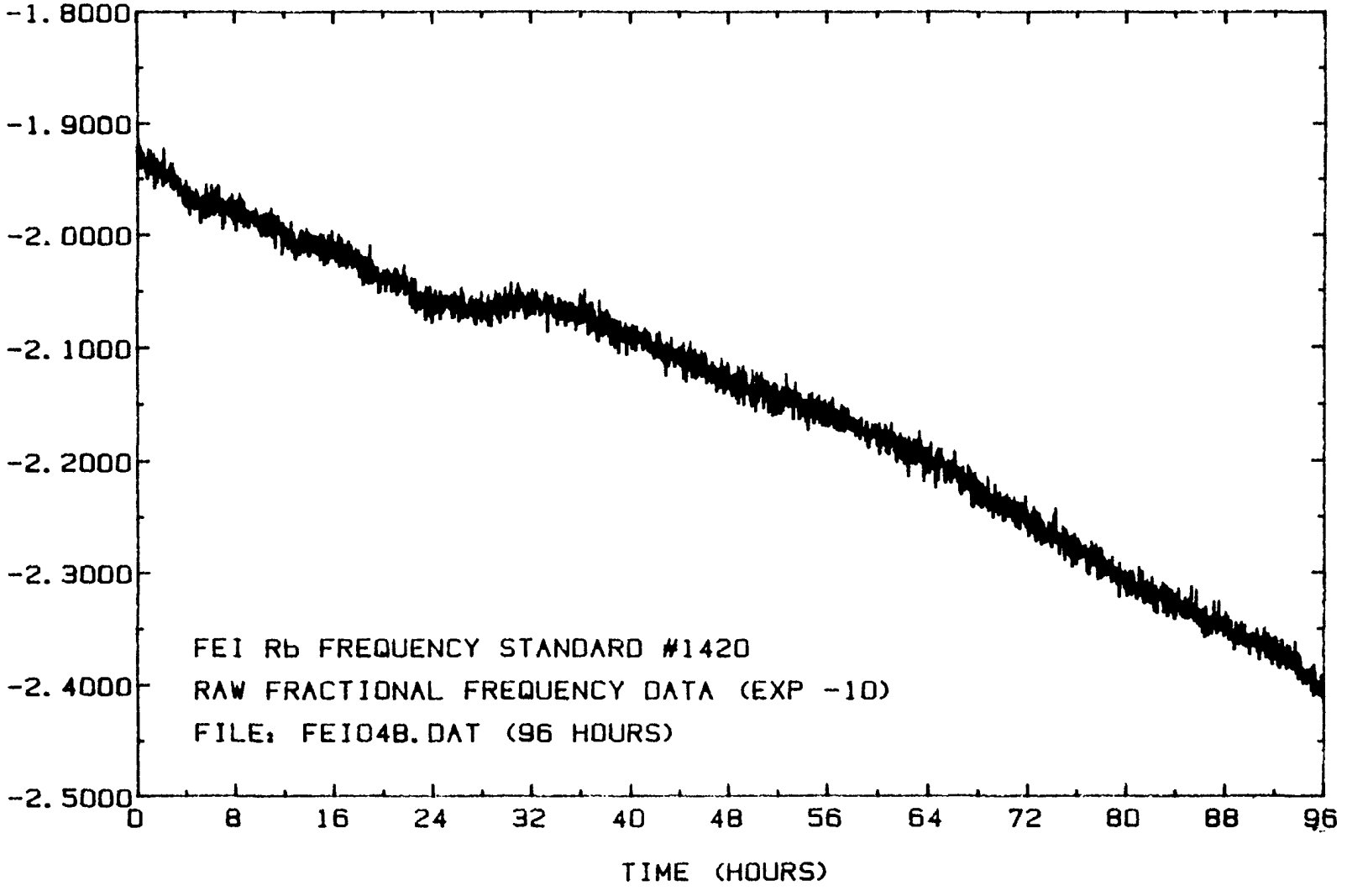


Figure 4
a. Frequency offset vs time, SN 1420

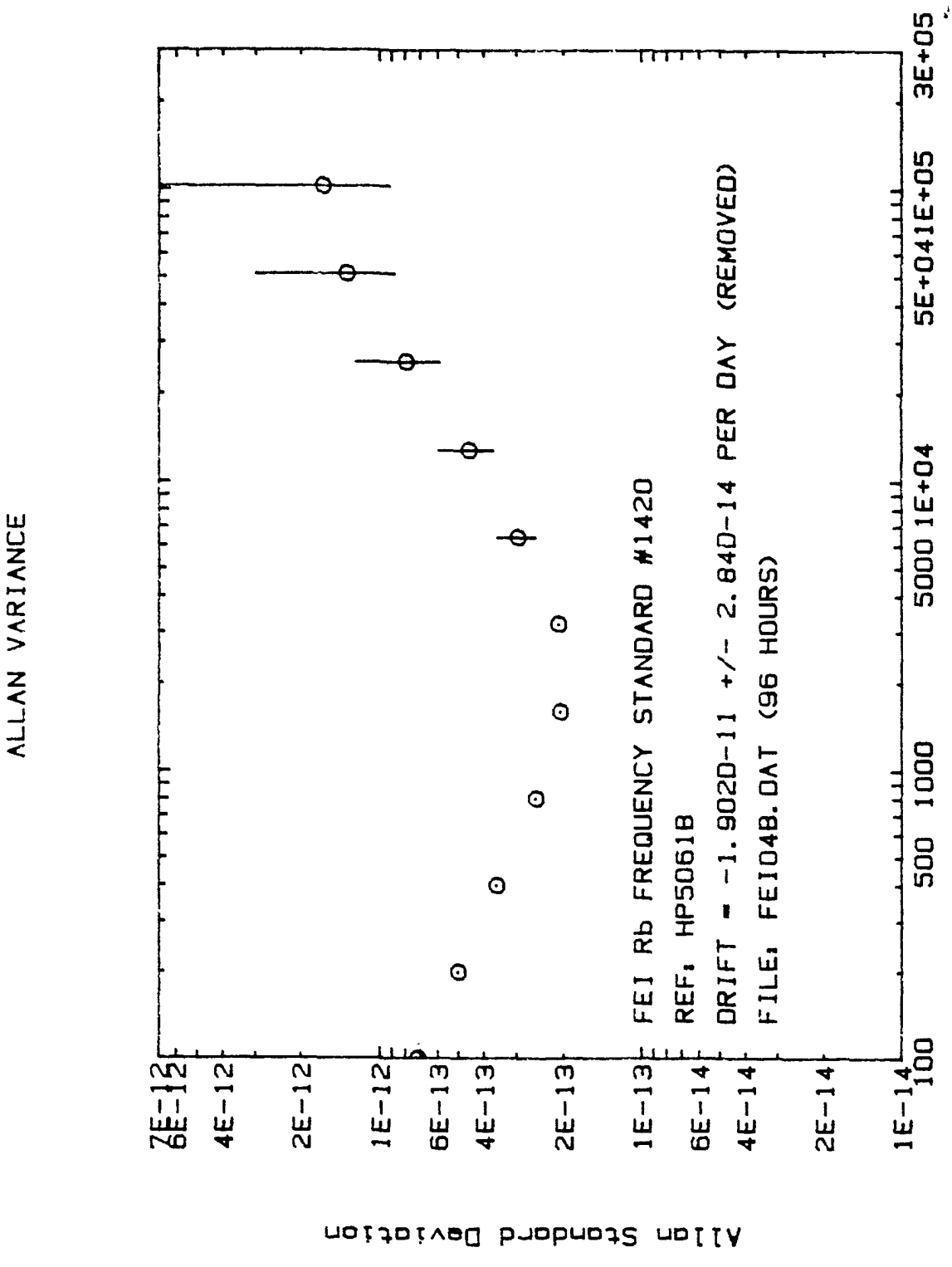


Figure 4
b. Allan standard deviation calculated from data in (a.).

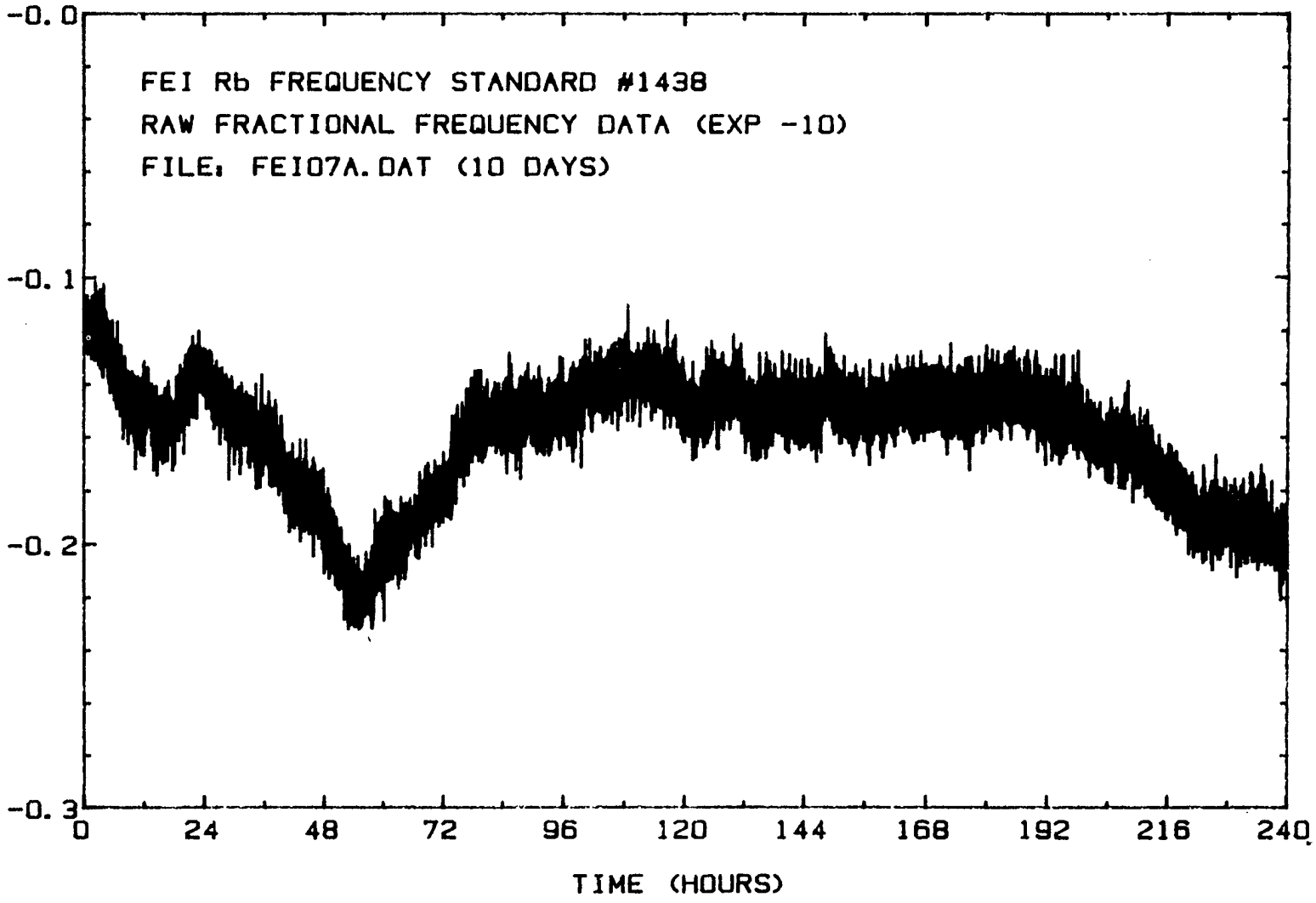


Figure 5
a. Frequency offset vs time, SN 1438.

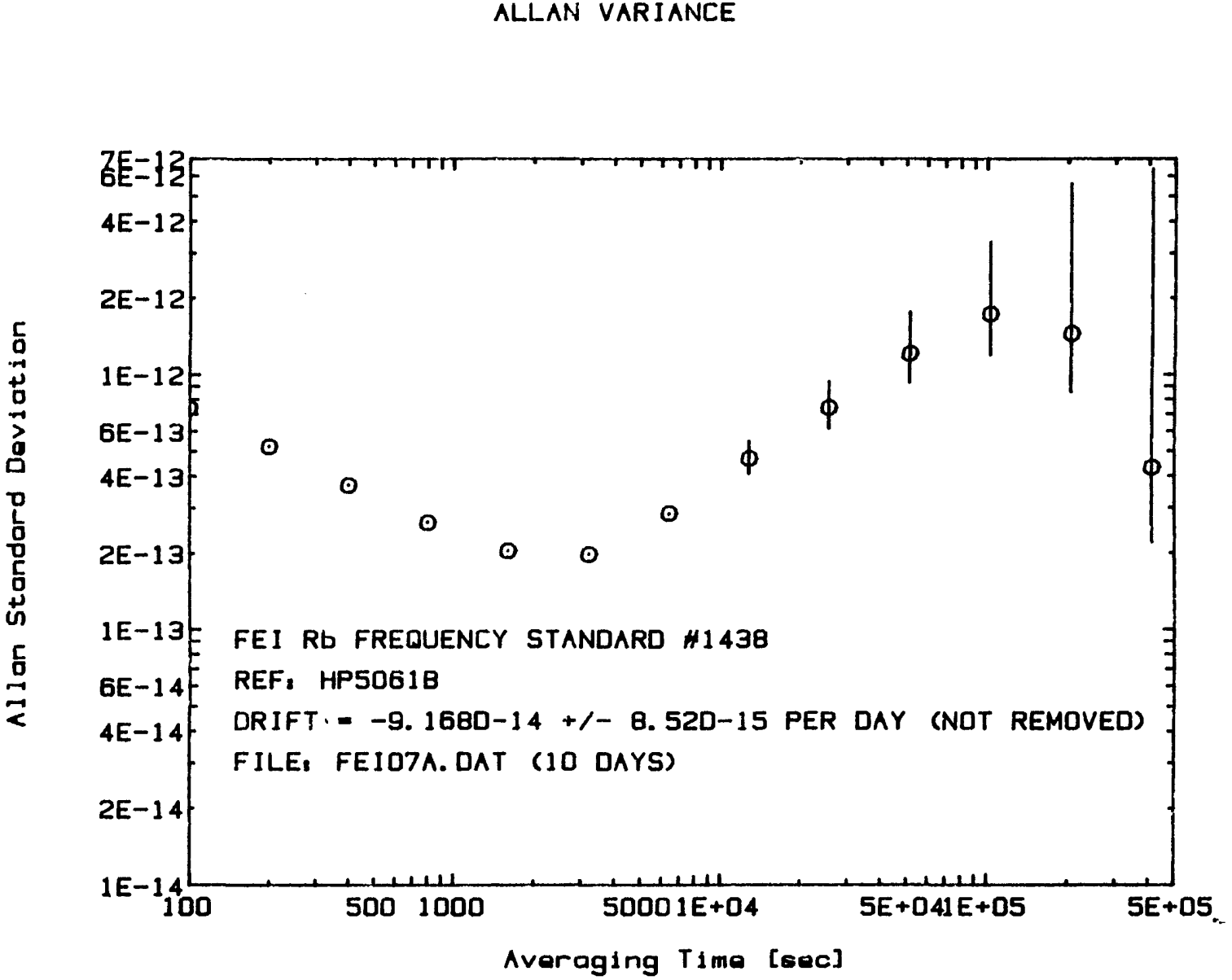


Figure 5
b. Allan standard deviation calculated from data in (a.).

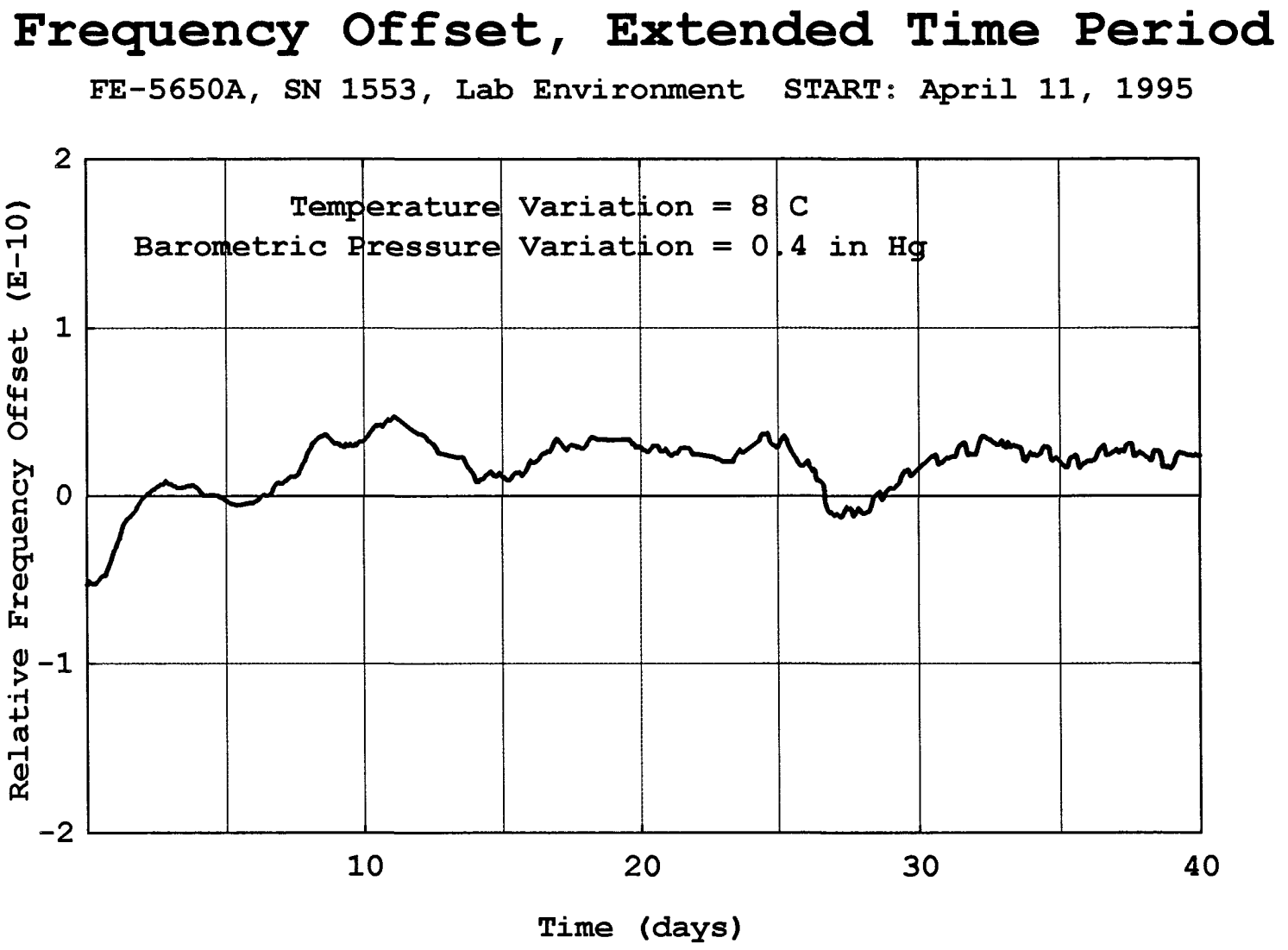


Figure 6
Frequency offset vs time. 40 day continuous monitoring of SN 1553

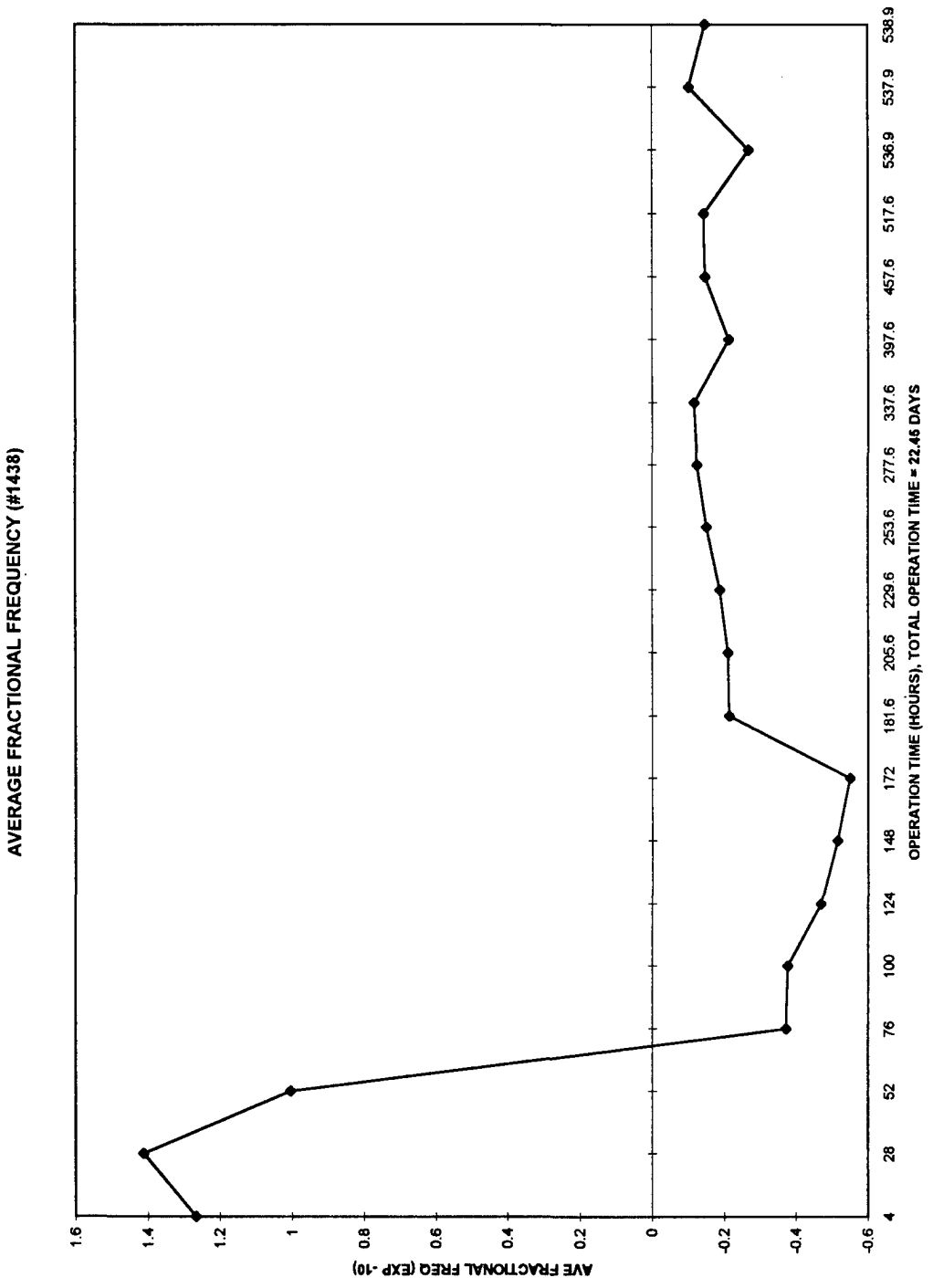


Figure 7
Frequency offset vs time. 20 day continuous monitoring of SN 1438, at The Aerospace Corporation.

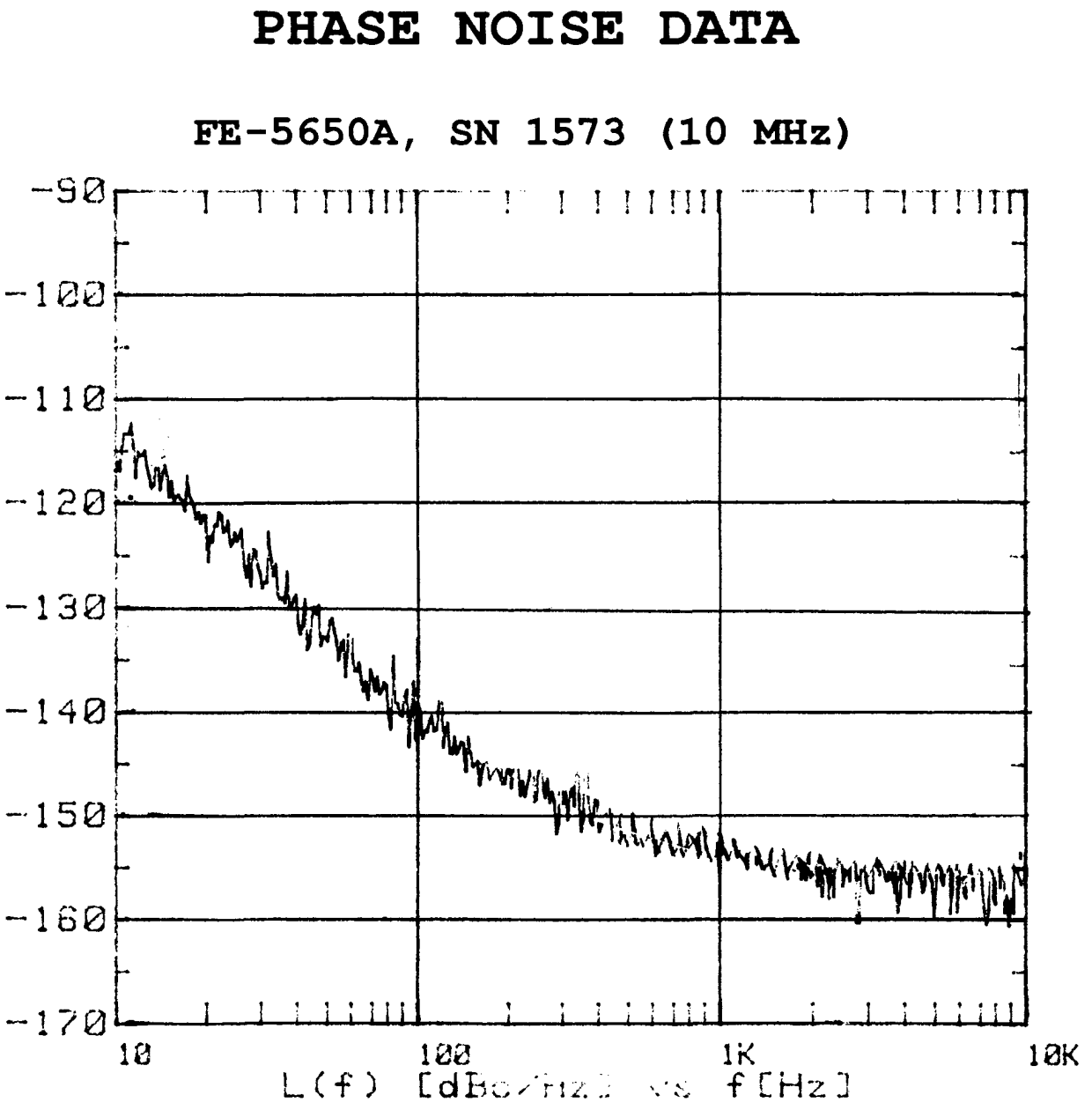


Figure 8
Phase noise data at 10 MHz (SN 1573)

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

MINIATURIZED RUBIDIUM CLOCKS FOR SPACE AND INDUSTRIAL APPLICATIONS

Claire Couplet, Pascal Rochat *, Gaetano Mileti, Hartmut Schweda, Pierre Thomann, Giovanni Busca
Observatoire Cantonal, Rue de l'Observatoire 58
CH-2000 Neuchâtel, Switzerland

** with Tekelec Neuchâtel Time SA , Mail 59*
CH-2000 Neuchâtel, Switzerland since March 1995

Abstract

This paper describes current developments of miniature rubidium clocks based upon a magnetron cavity performed by the Neuchâtel Observatory. Two variants of clocks for ground and space applications, respectively, have been developed. A brief description of both magnetron-type cavity and physics package basic characteristics is presented. The short-term stability obtained is discussed and compared with the theoretical value. The use of this cavity resonator, surrounding the absorption cell gives a repeatable short-term stability of 3 to 6 x 10^{-12} x $\tau^{-1/2}$ for both ground and space version.

Other parameters like temperature coefficient and long-term stability have been optimised and tested. Good performances have been obtained through a careful balance between the compensation of the light shift and the buffer gas shift effects.

The two versions of the clocks are as follows:

1. The small industrial version for commercial applications has a volume of 0.25 l and a typical power consumption of 7 W. This version is now in production.

2. The space-borne RUSO is designed for various missions. This model is now fully developed and a series of 6 flight models will be fully manufactured and tested before the end of 1995.

INTRODUCTION

The development of a miniaturized industrial Rubidium Ultra-Stable Oscillator (RUSO) at the Observatory of Neuchâtel has been completed in 1994 and the industrial RUSO activities have been shifted to the industry. The industrial production of several hundred of pieces of the miniature RUSOs started in March 1995. The development of a space compatible rubidium standard which started in October 1992 in the framework of ESA's In-Orbit Technology Demonstration Programme (TDP) is terminated now. Two qualification models have been delivered to Astro-Space Center in Moscow and are presently used for the tests of the space radiotelescope RADIOASTRON I including VLBI observations /1/.

This paper presents :

- The latest test results we obtain on conventional lamp pumped RUSO based upon magnetron-type cavity including the improvements of the short-term stability.
- Test results on various engineering and qualification models of space RUSO.
- Preliminary studies and results on a laser pumped rubidium oscillator using the magnetron cavity.
- Measurements of rubidium consumption in the lamps of RUSOs.

INDUSTRIAL MINIATURIZED RUBIDIUM OSCILLATOR

The principal characteristics of this unit are its very small volume (0.25 l) and the built-in data interface usable for monitoring and control.
The typical performances of this Rubidium standard are

- 1.5×10^{-10} over - 20 to 65°C op. temp. range
- 10 MHz output, adjustable by step of 2×10^{-10} with internal synthesizer
- fine adjustment by analog or RS232 command in step of 1×10^{-11}

0-7803-2500-1/95 $4.00 © 1995 IEEE

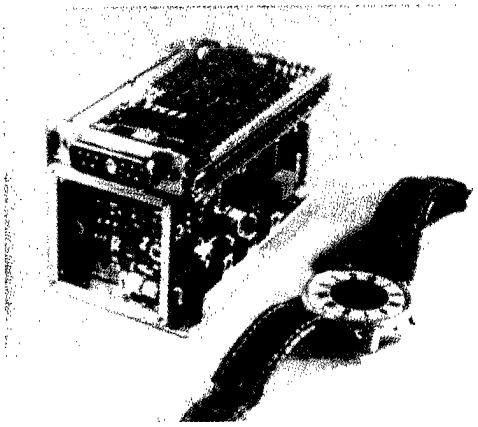


Fig. 1: 0.25 l industrial Rb clock

- $4 \times 10^{-12} \times \tau^{-1/2}$ typical short-term stability
- 15W (25W optional) power consumption during warm-up
- <8W@25°C steady state power consumption
- < 5 min. warm-time to reach a frequency stability of $< 5 \times 10^{-10}$
- working parameters stored in EEPROM memory

Intensive tests have been performed on 10 prototype units, starting with long-term stability measurements. The short-term stability has been improved by optimising the microwave interrogation (see Fig. 1).

Physics Package

The principal characteristics of this miniature physics package are the following:

- The use of integrated filter technique (IFT)
- The use of a magnetron microwave resonator

The magnetron resonator is a cylindrical cavity loaded with a concentric capacitive-inductive structure (annular metal electrodes). It allows smaller cavity dimensions and concentrates the microwave H-field at the end region of the cell. Note that in the IFT the entrance region of the cell is used for spectral filtering. This concentration offers a very good efficiency of the Rb atoms hyperfine transition interrogation process.

With this configuration we obtain a signal-to-noise level comparable to TE_{011} or TE_{111} cavity resonance modes of standard rubidium oscillators and in consequence a typical short-term stability of 3 to $4 \times 10^{-12} \times \tau^{-1/2}$. This level of performance is comparable with ultra-stable rubidium standards using separated filter technique (see Fig. 2).

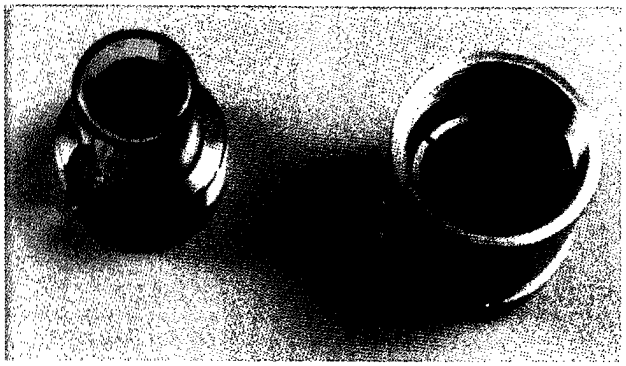


Fig. 2: Magnetron Cavity Assembly

Electronics package

The interrogation signal generation starts with a fundamental mode 20 MHz crystal multiplied by 9 in order to produce 180 MHz which is phase modulated by a 5.3 MHz synthesizer. The output frequency is directly derived from the 20 MHz VCXO.

The 10 MHz center frequency adjustment is made by changing the 5.3 MHz synthesizer frequency. The step size of frequency change is normalized to 2×10^{-10}/step at 10 MHz level.

The controller board contains a microcomputer unit which is associated with multi-channel D/A converters. The multi-channel D/A converter is used for parameters setting of the instrument. The microcomputer also controls the interrogation and synchronous detection sequence as well as Rb absorption "dip" search.

The built-in serial interface allows an automatic parameter adjustment procedure during the manufacturing process as well as a coarse and fine adjustment of the center frequency. All working parameters can be stored in a built EPROM memory. In addition an 8 channels A/D converter is used for monitoring principal internal signals like: light level, signal level , thermostats voltages, RF power etc. (see Fig. 3).

Fig. 4 gives a typical frequency behaviour (Y=1.5E-10) and power consumption (PIN) characteristics (7.5 W at 25°C) versus ambient temperature.
Fig. 5 shows a typical warm-up characteristic with 25W input power. We reach 3×10^{-10} frequency stability after 4 min. of operation.

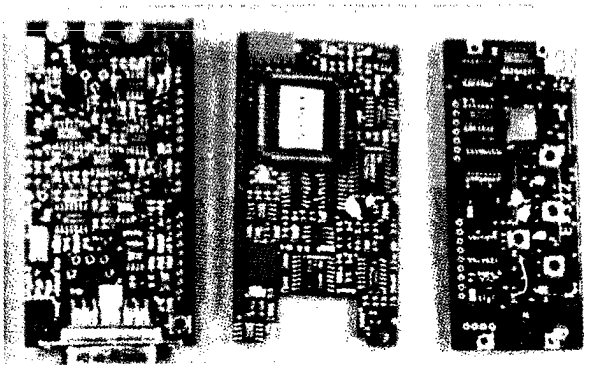


Fig 3: Electronics package

Tests results on miniature Rubidium:

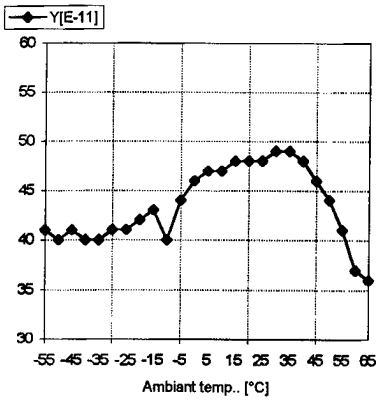


Fig. 4: Frequency versus ambient temperature characteristics

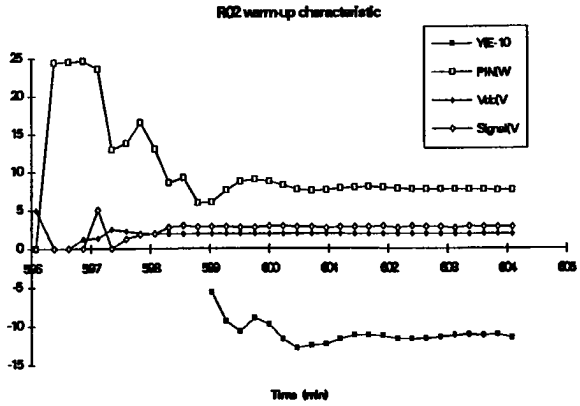


Fig. 5: Warm-up characteristic

For the final industrial version the input power will be limited to 15W. In this case the warm-up time will be 5 to 6 min. instead of 4 min.

Y [E-10]	: relative frequency in 10^{-10} units
PIN	: input power in W
VDC	: light level intensity in V
Signal	: 2nd harmonic signal level

Long-term frequency behaviour

Fig. 6 represents the long-term frequency behaviour under normal laboratory conditions of two industrial units R05 and R08. After few months of operation, stability better than 1×10^{-11}/month where observed.

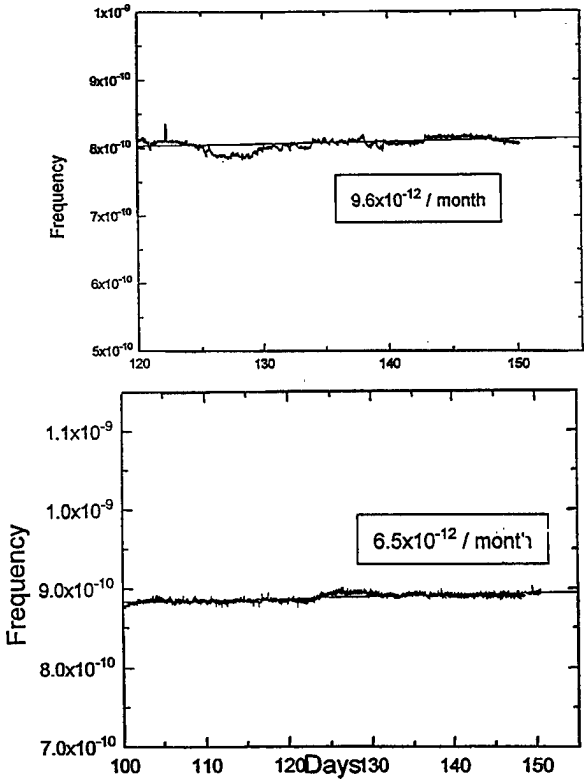


Fig. 6: Long-term characteristics

Short-term stability:

Fig. 7 shows typical results obtained for the phase noise at 10 MHz measured down to 1mHz Fourier frequency. Fig. 8 shows the equivalent short-term stability against a selected BVA quartz crystal oscillator for observation times up to 200 sec. Since this measurement was limited by the stability of the BVA oscillator, it was needed to measure the RUSO also against H-maser reference. Fig. 9 shows the short-term stability of the RUSO against H-maser as reference. We observe from this figure a flicker floor of ~1e-13 for an observation time greater than 1000 sec. (without drift remove !!).

Note: This measurement was made in stable laboratory condition with very low temperature drift. The drift observed from this figure corresponds to a very low temperature drift since the tested long-term stability of this particular unit was better than 3e-13/day.

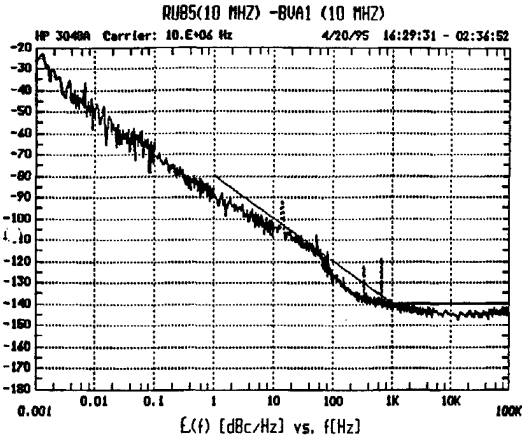


Fig. 7: Phase noise

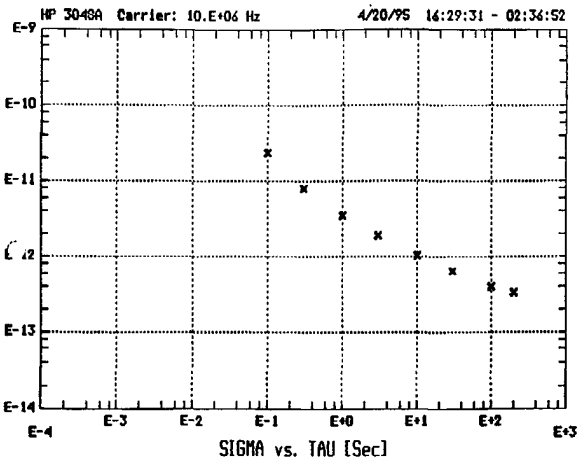


Fig. 8: Short-term stability with BVA osc. as reference

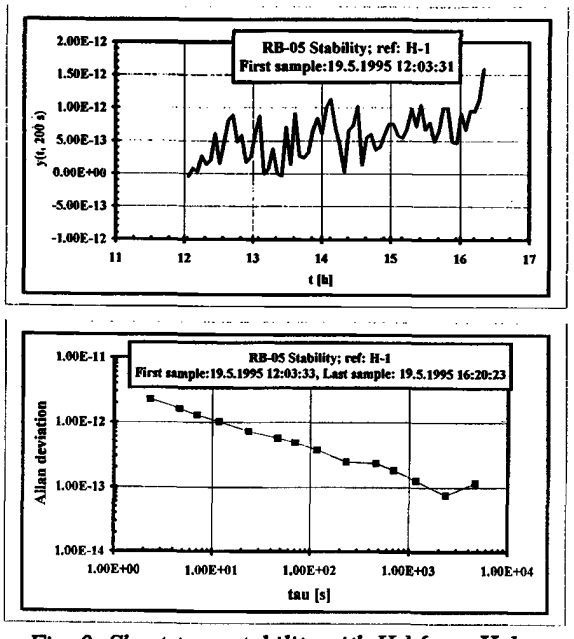


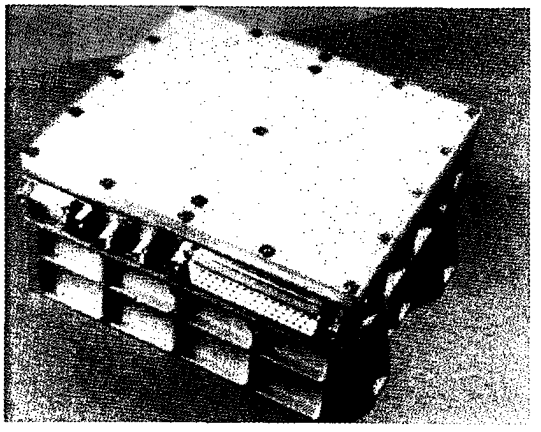
Fig. 9: Short-term stability with H-Maser H-1 as reference

SPACE RUSO

Physics package

The principle design characteristics of the physics package are its low power consumption, small size and mass, along with minimal environmental sensitivities and mechanical ruggedness.

A unique feature of the design is that all parts of the physics package are incorporated in a sealed titanium tube. Inside this tube, lamp and cell sections form two separate blocks which operate at well-defined but different temperatures. The cylindrical tube is filled with xenon for low thermal conductivity. Each block is connected to one of the end plates by a thin-walled tubular titanium spacer. This configuration greatly reduces the thermal flow between the blocks and to the tube envelope. It allows a very compact design with low power consumption, short warm-up time and minimal environmental sensitivities. Apart from the improved stability upon outside temperature variations, it is obvious that the sealed container makes the physics package virtually insensitive to barometric influences and allows RUSO operations in pressurised or vacuum environment.



Test results

The signal parameters measured on the engineering model are practically the same as the industrial unit and are the following :

Absorption "dip" amplitude	: 1 μA
Photo cell current	: 150 μA
Shot noise	: 8 pA/√Hz
Total noise	: ~10 pA/√Hz
Line Q	: 3.6×10^6
Discriminator slope	: 0.56 nA/Hz
Predicted stability (with total noise and 50% dead time)	: $3.94 \times 10^{-12} \times \tau^{-1/2}$
Measured stability (Pre-Engineering Model)	: $\sim 4 \times 10^{-12} \times \tau^{-1/2}$

The following table is a condensation of the specifications and test results on the RUSO qualification model.

PARAMETER	ESA spec.	TEST Results
Long-term frequency stability ($\Delta f/f_0$)	**< 4e-11/m**	**< 1e-11/m**
Short-term stability $\sigma(\tau^{-1/2})$ for $1 \leq \tau \leq 100$s	**< 1e-11 $\tau^{-1/2}$**	**4e-12 $\tau^{-1/2}$**
Temperature sens. ($\Delta f/f_0$) in - 0 to + 40°C range	**< 4e-12/°C**	**< 3e-12/°C**
Static g-sensitivity	**< 4e-12/g**	**3.5e-12/g**
Magnetic sensitivity	**< 2e-11/G**	**< 1e-11/G**
SSB phase noise (L (f_m) (f_0 = 5 MHz) in dB		
f_m = 1 Hz	**-85**	**-90**
= 10 Hz	**-105**	**-110**
= 100 Hz	**-125**	**-129**
= 1.000 Hz	**-145**	**-148**
= 10.000 Hz	**-145**	**-148**
Warm-up time to reach stability < 5 x 10^{-10}	**< 6 min**	**< 3e-10 / 6min**
Retrace ($\Delta f/f$)	**< 2e-11**	**< 1.2e-11**
Power consumption during warm-up	**< 15W**	**14.5W**
Power consumption during normal operation	**< 8W**	**7.5W (25°C)**
Volume	**< 1 liter**	**1.2 liter**
Mass	**< 1kg**	**1.3kg**
DC power supply voltage range	**22 to 43V**	**16 to 45V**

PRELIMINARY STUDIES AND RESULTS ON LASER PUMPED GAS CELL CLOCKS

In parallel with the development of conventional rubidium frequency standards, ON is investigating on the potential of using a laser diode instead of the traditional discharge lamp /3/-/6/. So far, we have used the same microwave resonator in both cases, in order to compare quantitatively the results. With an accurate study of the phenomenon of light-shift saturation, we demonstrated its geometrical origin /5/. With a simple theoretical analysis the observed effects and their dependence on the laser beam parameters - such as frequency, intensity and geometry - could be interpreted in terms of basic physical phenomena.

The understanding of these aspects is important in order to optimise the geometry of the experimental set-up, but also in order to perform quantitative realistic predictions of the achievable performances.

We performed a simple calculation based on the existing models /7/ that allows realistic quantitative double resonance signal calculations. The resonance signal produced by the atoms in a given geometry was calculated. We could verify that - in the case of a miniature device using a buffer gas - the theoretical results correspond to our experimental values within 10%.

Typically, the optimal double resonance signal obtained with 1-2 cm^3 of Rb 87 vapour produces a discriminator slope of 1 nA/Hz, and the double resonance dip has an amplitude equal to 20 to 30 percent of the DC level (which produces a shot noise of 1 pA/$\sqrt{Hz}$). These numbers lead to an achievable short-term stability (shot noise limited Allan variance) $\sigma y = 1$ *to* $2 \times 10^{-13}\ \tau^{1/2}$ /6/.

From these results, the characteristics of the signal in larger volumes - and consequently the achievable stability - can be extrapolated, and the design of a future device can be tailored to the required performances.

Contrary to the discharge lamp case, in which the shot noise limit can be achieved and the spectral properties of the light are not fluctuating rapidly, the use of laser diodes requires a more severe control of the light frequency and intensity. This problem is largely dependent on laser diode manufacture technology. For our purpose, we investigated low-cost commercial diodes used in CD players, and studied the effect of the intensity noise and of the frequency noise.

Regarding the intensity noise, two contributions to the clock instability can be considered : a) direct S/N reduction, b) clock frequency fluctuations induced by laser intensity fluctuations through the intensity dependence of light-shift. Our measurements showed that it is possible to operate such a laser at a level producing - with a 1,5 cm^3 vapour volume - a resonance signal as high as described in the previous paragraph, and an intensity noise contribution comparable to the normal shot noise. This demands a Relative Intensity Noise (RIN) of 10^{-12} - 10^{-13} / Hz. At this level, if the laser detuning with respect to the zero light-shift frequency is below 100 MHz, contribution b) is negligible. Thus, the intensity noise limited short-term stability with commercial laser diodes is *2 to 4 x* $10^{-13}\ \tau^{1/2}$ */6/.*

Regarding the frequency noise, different processes may alter the stability of the clock, and one question is to determine the optimal servo-scheme of the laser frequency stabilisation, and whether it is necessary or even convenient to use a separate absorption cell to lock the laser /8/. Our light-shift measurements and our frequency noise measurements showed that a fast servo loop can reduce the frequency noise of commercial laser diodes emitting at 780 nm below 10 kHz/$\sqrt{Hz}$. At this level, the laser frequency noise directly converted to clock frequency noise via the light-shift is still compatible with an Allan variance at the $10^{-13}\ \tau^{1/2}$ level. However, we observed that some indirect effects of the laser frequency fluctuations might alter in a non-negligible way the short-term stability of the clock. In fact, as the laser frequency is tuned across the absorption line of the interrogated atoms, the frequency fluctuations are converted into intensity fluctuations at a level which depends on the exact operating frequency. In principle, it should be convenient to choose this frequency as close as possible to the maximum of the absorption profile. However, this might not be the optimal choice, especially for two reasons. The first is due to the fact that the frequency noise and the intensity noise are correlated, so that a well adjusted detuning might offer a useful cancellation. The second reason is related to the long-term performance. In fact, as it is theoretically expectable, and as we measured experimentally, the particular broadened hyperfine structure of the rubidium atoms in the vapour involve a detuning of some tens of MHz between the maximal absorption laser frequency and the zero light-shift frequency. Taking into account all these effects, the laser frequency noise limited short-term stability for a miniature standard is *1 to 3 x* $10^{-13}\ \tau^{1/2}$ */6/.*

Further investigations have to be made in order to verify these predictions, and determine which solution is the best compromise. As a conclusion for our short-term stability study, we summarise our results in Table 1.

CALORIMETER DETERMINATION OF RB CONSUMPTION MEASUREMENTS

The problem of aging of Rb lamps due the diffusion of Rb metal into the glass walls of its bulbs has been studied before /2/. Since Differential Scanning Calorimetry (DSC) has proven to be very successful in the past, we applied this method for the determination of the Rb content of two Rb lamps as function of its operation time.

Unlike the lamp bulbs used in the space RUSO which are made of special highly alkaline-resistant PHILIPS glass we studied lamps made of Schott glass type 8436 under two different operation conditions.

Type of light Source	DC level of photocurrent	Experimental absorption dip	Experimental absorption dip ampl./FWHM	Theor. shot noise limited stability	Theor. AM-FM limited stability	Experimental short-term stability
Discharge lamp	150 mA	0,5-1 % of DC current	0,8 nA/Hz	2 to 3 x $10^{-12}\ \tau^{1/2}$	-	2,8 to 4 x $10^{-12}\ \tau^{1/2}$
Commercial laser diode	4 mA	20-30 % of DC current	1,2 nA/Hz	1 to 2 x $10^{-13}\ \tau^{1/2}$	3 to 5 x $10^{-13}\tau^{1/2}$	-

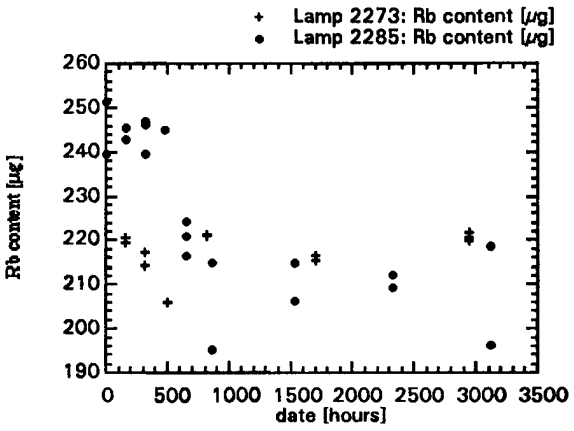
Table 1 *Summary of theoretical and experimental results concerning the performance of a passive cell clock based on an optically pumped Rb vapour of 1-2 cm^3 volume with buffer gas. The laser diode must have a line width, a RIN and a frequency noise @ 300 Hz below respectively 100 MHz, 10^{-12}Hz, and 10 kHz/$\sqrt{Hz}$.*

1) lamp #2273 operated under 5 Watts discharge,
2) lamp #2285 operated under standard discharge as in industrial RUSO.

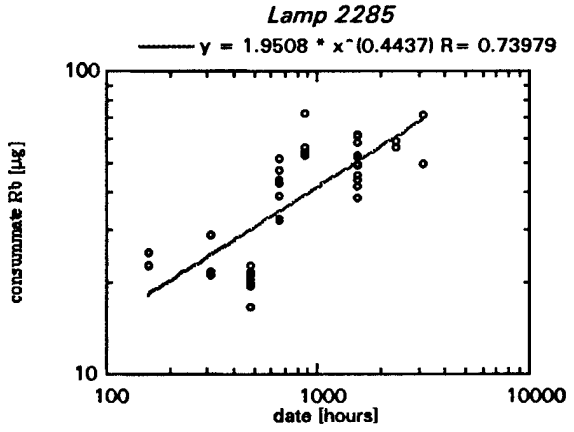
A SETARAM Micro-DSC was used in the study of these lamps.

Fig. 11 displays the different behaviour of the two lamps. The lamp #2273 which was operated at an increased discharge level shows an almost constant Rb content over the whole operation time of several months while lamp #2285 experienced an exponential loss in the first 30 days of operation. As a conclusion a burn-in at elevated discharge levels is highly recommended for such lamps.

Fig. 12 shows ht evolution of Rb consumption as function of operation time in lamp #2285.



Rubidium content of lamp as a function of consumption time



Consumption of Rb in the lamp #2285

References:

/1/ P. Rochat, H. Schweda, G. Mileti and G. Busca, IEEE 48th Frequency Control Symposium, Boston 1994, pp 716-723.

/2/ C.H. Volk et al., 38th Ann. Frequency Control Symposium 1984, pp 387-400.

/3/ J. T. Liu, P. Thomann, L. Zhang, G. Busca, (1990), Proc. 4th EFTF, Neuchâtel, pp 157/59.

/4/ P. Willemin, G. Busca, J. T. Liu, P. Thomann, (1991), Proc. 5th EFTF, Besançon, pp 181/86.

/5/ G. Mileti, P. Thomann, (1994), Proc. 8th EFTF, Münich, pp 377-384.

/6/ G. Mileti, P. Thomann, (1995), Proc. 9th EFTF, Besançon, to be published.

/7/ J. Vanier, C. Audoin, (1989), The Quantum Physics of Frequency Standards, Hilger.

/8/ S. Kinugawa, T. Imamura, H. Suga, Y. Ohuchi, Y. Koga, Y. Saburi, (1993), Proc. International Symposium on Atomic Frequency Standards and Coherent Quantum Electronics, Nara, pp 81/82.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

RECENT RESULTS OF THE LPTF CESIUM FOUNTAIN PRIMARY FREQUENCY STANDARD

G. Santarelli, S. Ghezali, Ph. Laurent, S. N. Lea, M. Bahoura, K. Szymaniec, E. Simon and A. Clairon

LPTF/BNM, Observatoire de Paris, 61 avenue de l'Observatoire, F-75014 Paris, France

Abstract-**A Cs atomic fountain has been in operation at LPTF for more than one year. Ramsey fringes as narrow as 0.7 Hz are obtained. The short term fractional frequency stability is $2\,10^{-13}\,\tau^{1/2}$, limited by the frequency noise of the local oscillator. Most of the classical sources of inaccuracy are considerably reduced in an atomic fountain, most being reduced to the 10^{-16} level (in molasses operation). The new limiting effect is the spin exchange frequency shift due to collisions between cold atoms. The measurements we have performed show that it is possible to have a shift of $2\,10^{-13}$ with a potential stability of $10^{-13}\tau^{-1/2}$. An evaluation of the major shifts and of the frequency stability limits in our fountain primary frequency standard are presented.**

I INTRODUCTION

The Ramsey separated oscillatory field method which is used in cesium frequency standards leads to a resonance line width inversely proportional to the time between the two microwave interactions. With sub-Doppler laser cooling, the atomic r.m.s velocity is reduced to 2 cm/s (few μK temperature); when using a fountain configuration about 0.6 m in height above the microwave cavity we have less than one Hertz line width [1,2,3]. The quality factor of the transition is increased by a factor 100 with respect to conventional Cs frequency standards. The stability of the cesium fountain is about 3 times better than optically pumped standards. The envisaged accuracy of a few 10^{-16} is yet to be demonstrated, but after less than two years' operation we are close to being able to quote an accuracy better that of any existing standard.

The fountain is not far off being a simple and reliable experimental standard. Advantage will be taken of the rapid development in modern laser diodes and optical fibers, to evolve from the present complex set-up to a simplified and more easily reproducible device.

* Present address National Physical Laboratory, Queens Road, Teddington, TW110LW UK

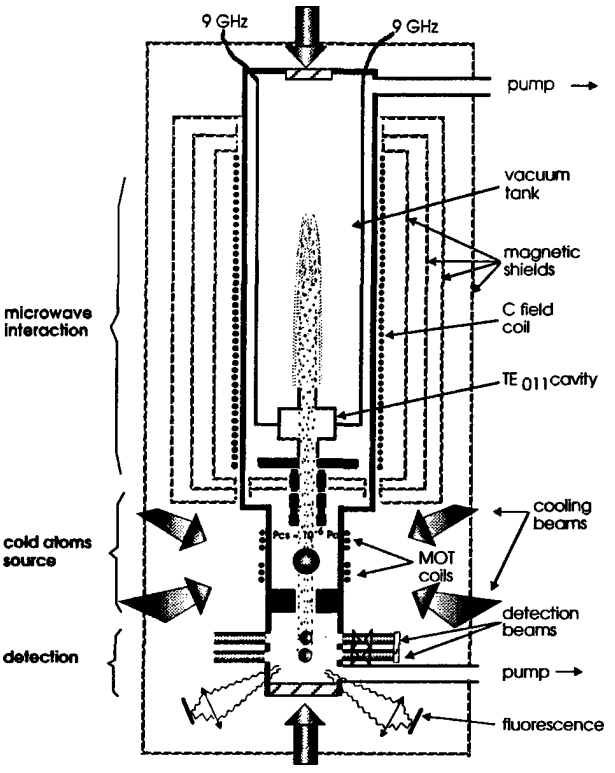


Figure 1. The Cesium atomic fountain prototype.

II FOUNTAIN DESCRIPTION

The device shown in Fig. 1 is the cesium atomic fountain operating at the LPTF. The fountain has been described in the CPEM 95 [3]. We recall the basic operation of the standard. Six independent laser beams tuned to the red of the cyclic F=4-F'=5 transition capture and cool atoms to a few μK in a low pressure cesium vapor cell (10^{-6} Pa), at room temperature. From 10^5 up to few 10^8 atoms are cooled down to 2.5 μK in a lin-per-lin optical molasses or a σ^+–σ^- magneto optical trap (M.O.T.). The loading step is followed by launching which is performed using a moving optical molasses obtained by detuning the vertical beams symmetrically by a given quantity $\Delta\nu$, creating a standing wave moving at the

0-7803-2500-1/95 $4.00 © 1995 IEEE

velocity $v_z=\Delta v\lambda$ in the laboratory frame. The intensity and frequency of the laser beams are controlled via acousto-optic modulators. At the end of the launching phase, all the beams are blocked by mechanical shutters to avoid light shifts due to any residual light transmitted through the modulators.

The launched atoms perform a ballistic flight under the effect of gravity passing twice through a cylindrical TE_{011} copper cavity in which they undergo a $\pi/2$ microwave pulse, During the flight the atoms experience a highly homogeneous magnetic field (C-field) produced by a solenoid and compensation coils. Additional compensation coils ensure the continuity of the magnetic field along the atoms' path.

The C-field and the microwave region are surrounded by a set of three μ-metal magnetic shields in order to reduce the axial magnetic fluctuations and to improve the homogeneity of the magnetic field. The axial shielding factor is about 10^4. The whole set-up is surrounded by a fourth μ-metal shield. To improve the stability of the axial magnetic field an active servo loop is used (only in molasses operation). In the upper part of the device, the pressure is kept low ($<10^{-7}$ Pa) in order to reduce collisions with residual background gases. Graphite tubes between the different regions reduce the Cs pressure in the microwave and detection the regions.

Under typical conditions only 10% of the launched atoms are detected. The population of both hyperfine levels are measured by fluorescence. The Ramsey signal is given by $N_{F=4}/(N_{F=3}+N_{F=4})$, where $N_{F=4}$ and $N_{F=3}$ are the number of atoms in the F=4 and F=3 levels respectively. The normalization enables the signal variations due to shot-to-shot atom number fluctuations, which are of the order of 1%, to be reduced. The collection efficiency is about 0.7% and the total detection noise (photodiode and transimpedance amplifier) is equivalent to few tens of atoms per shot. The operation of the fountain is pulsed with a cycle period of about 1.1s.

In order to tune and stabilize the microwave cavity the core of the fountain (C-field zone and cavity), is heated. In normal operation the temperature can be controlled to within 0.5 K by a digital servo system.

The 9.1927..GHz clock transition frequency is generated by a new frequency multiplication chain[4]. Fig. 2 depicts the microwave generator; first a 100 MHz oscillator is phase locked on a high stability commercial 10 MHz BVA quartz with a bandwidth of about 100 Hz. The 100 MHz output is frequency doubled and routed to the local oscillator input of a sampling mixer. This device mixes the 46th harmonic of the 200 MHz signal with the microwave output of a dielectric resonator oscillator (DRO) at 9.192..GHz. The beat note signal at 7.3..MHz is amplified and phase compared with the output of a commercial synthesizer in order to phase lock the DRO

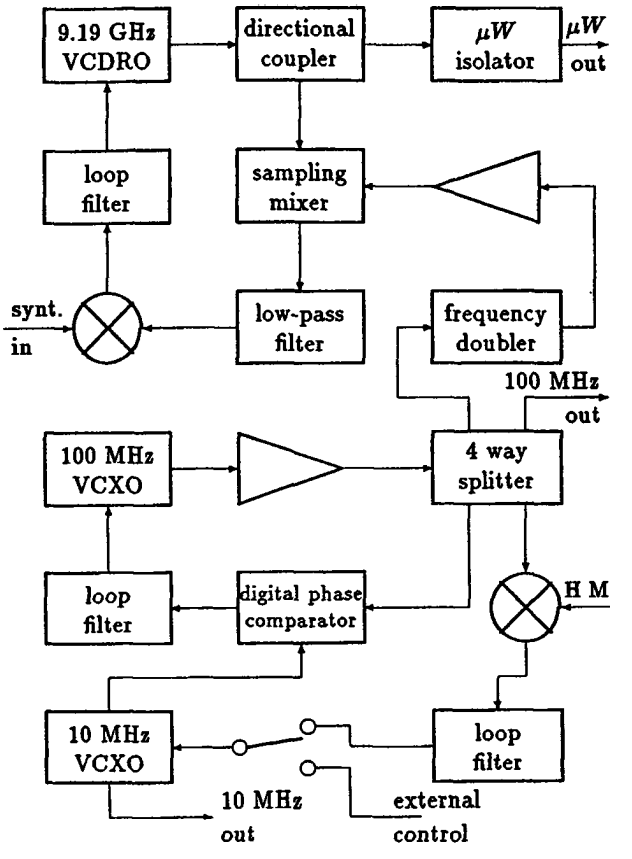


Figure 2: Block diagram of the frequency multiplication chain.

with about 100 kHz bandwidth. With this configuration we can modulate and the fine-tune the frequency of the microwave signal by changing the synthesizer frequency. This enables the frequency to be locked at the atomic resonance.

Special attention has been given to the spectral purity and the phase stability of the microwave chain output signal. The final goals of 10^{-16} stability per day and an accuracy of the same level need a phase stability of 10 ps/day and a spurious level lower than -60 dB with respect to the carrier. We have built two chains which show a fractional frequency stability of a few 10^{-17}/day.

Fig. 3 shows typical Ramsey fringes obtained for a launching height of 0.3 m above the cavity. The Rabi envelope is about 60 Hz wide corresponding to an interaction time in the cavity τ of about 15 ms. The high number of fringes is due to the narrow velocity distribution. The relative width of the velocity distribution of the atoms is only 1%; this induces a small variation of the fringes' contrast. The width of the narrowest fringe that

we can realize in our fountain, is about 700 mHz, a factor 100 less than that obtained in the best thermal beam device.

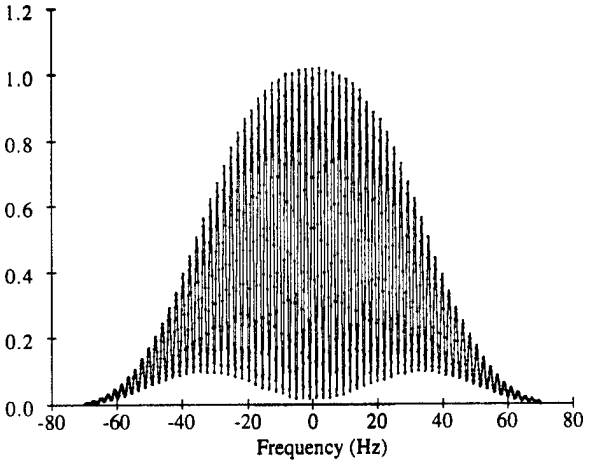


Figure 3. Ramsey fringes obtained when launching the atoms 0.3 m above the microwave cavity.

III FREQUENCY STABILITY

The frequency stability can be evaluated in terms of the two sample Allan standard deviation. For the fountain we expect a fractional frequency stability given by:

$$\sigma_y(\tau) = \frac{\delta\nu}{\pi\nu_0 S/N}\sqrt{\frac{T_c}{\tau}} \qquad (1)$$

where T_c is the fountain cycle time, τ the integration time, ν_0 the clock transition frequency and $\delta\nu$ the line width of the Ramsey fringe. The signal-to-noise ratio S/N is the relative Allan standard deviation of the shot-to-shot fluorescence signal measured on the top of the fringe (around 300 in molasses operation and up to 1000 in using a M.O.T.). Fig. 4 shows the measured fractional frequency stability of the fountain against the H-maser. The expected stability is about $7\ 10^{-14}\ \tau^{1/2}$, whereas we measure only $2\ 10^{-13}\ \tau^{1/2}$. As pointed out by G. J. Dick [5], the local oscillator frequency noise is down-converted by an aliasing phenomena under pulsed operation, leading to a degradation of the detected signal to noise. The noise that is involved in this process is the local oscillator frequency noise located around multiples of the cycle frequency. The degradation is directly related to the unavoidable dead time between two interrogation periods, due to the loading time, the detection and the flights between source and microwave cavity and interrogation zone to the detection area. In our fountain, we have typically 500 ms of ballistic flight above the microwave cavity and 600 ms for the rest of the operations and it is quite difficult to improve this duty cycle. To realize the potential stability of the fountain we need a local oscillator which is beyond the current technology for room-temperature oscillators. A further improvement of the short-term frequency stability by a factor two can be achieved by using the best available quartz oscillator. It is important to observe the key role played by the stability in the accuracy evaluation at the 10^{-16} level. With a stability of $2\ 10^{-13}\ \tau^{1/2}$, effects at the few 10^{-16} level are easily evaluated in one day of measurements, if two similar devices are compared.

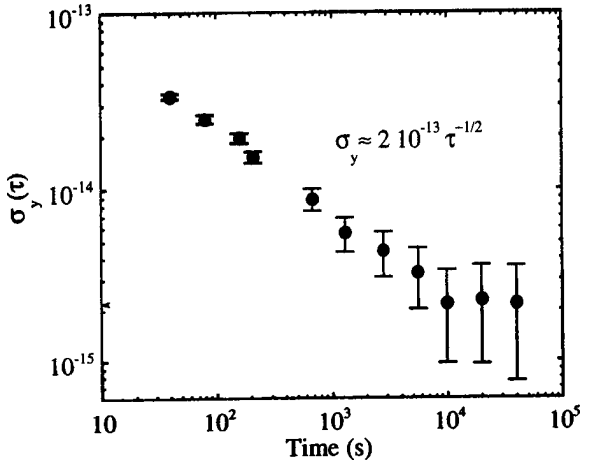


Figure 4. Fractional frequency stability of the atomic fountain vs the H-maser.

VI FREQUENCY SHIFTS

Most of the frequency shifts which limit the accuracy of Cs atomic frequency standards are considerably reduced when using cold atoms and long interrogation time. For instance, the relativistic frequency shifts, which are limiting effects in conventional cesium beam standards, are less than 10^{-17}.

Cavity related effects, which depend on the width of the central Ramsey fringe, are also considerably reduced. Due to the low atom number, and a new microwave cavity with reduced loaded quality factor Q_{cav} of about 10000, we are far from maser oscillation and the maximum cavity pulling can be calculated as in a passive cesium beam standard [6] by:

$$\frac{\Delta\nu}{\nu_0} = \frac{\Delta\nu_{cav}}{\nu_0}\left(\frac{Q_{cav}}{Q_{at}}\right)^2 \qquad (2)$$

where $\Delta\nu_{cav}$ is the cavity frequency detuning and, Q_{at} is the atomic quality factor. With a temperature stability of

0.5 K, the offset of the cavity frequency is always less than 100 kHz and the consequent relative uncertainty is only 10^{-17}.

The residual first order Doppler shift is due to the phase gradient of the microwave field due to the losses in the copper walls.

$$\frac{\Delta\nu}{\nu_0}=\frac{\phi_1-\phi_2}{\pi Q_{at}} \qquad (3)$$

$$\phi_i = \iint_s \rho_i(\bar{x})\,\phi(\bar{x})\,d\bar{x} \quad i=1,2$$

where $\rho_1(\bar{x})$ and $\rho_2(\bar{x})$ are the normalized atom distributions over the cavity hole (10 mm diameter) during the first and second pass respectively and $\phi(\bar{x})$ is the phase gradient. In molasses operation, it is reasonable to assume that, at the first pass, the atomic cloud has a gaussian distribution with a standard deviation of about 4.5 mm. We allow for possible misalignment in the launching direction of 1 mrad (about 1 mm) of the atomic cloud with respect to the center of the cavity. For the second pass we assume a uniform atomic distribution, due to the expansion of the atomic cloud. For a high Q cylindrical TE_{011} cavity A. De Marchi and coworkers [7] have calculated the phase distribution in the 2D case. Using the resulting phase map we have calculated an average phase shift of 4 μrad, leading to a first order Doppler shift of about 10^{-16}. This result is quite optimistic because the phase distribution also depends on the cavity end caps and the effective conductance of the copper walls. From the same work, it seems to be possible to improve the phase field flatness by feeding the microwave cavity symmetrically, the resulting phase gradient being strongly reduced. A residual first order Doppler shift of less than 10^{-15} appears feasible.

The black body shift, which is related to the Stark shift induced by the thermal radiation, is 1.78 10^{-14} at 300 K and it can be calculated at a few 10^{-16} level, if the radiation temperature is known within 0.5 K [8]. Unlike classical Cs beam devices, the thermal spectrum inside the fountain is probably a black body radiation. The C-field tube and the microwave cavity are both made of copper and are at the same homogeneous temperature; external apertures account only for 0.2% of the total surface. The accuracy of all Cs standards depends on this important effect; the fountain is well suited as a device to perform a measurement of it.

As a consequence of the narrow line-width, the C-Field in a fountain can be reduced, leading to a reduction of the sensitivity to magnetic field fluctuations and inhomogeneities. A field of 190 nT (cf. about 6 μT in a conventional beam standard) is sufficient to separate the seven $\Delta m_F=0$ microwave transitions by 1.4 kHz. Under these conditions, the overlap of neighboring transitions is completely negligible. A further advantage of a fountain is the possibility of mapping the axial static magnetic field by launching atoms to different heights [9]. This technique gives a direct measurement of the local magnetic field with a resolution better than 0.1 nT and a spatial resolution of about 1 cm^3.

The Ramsey fringes for magnetic field sensitive transitions have a contrast better than 50%: this confirms that the averaged magnetic field over the free flight above the cavity differs by less than 0.2 nT for different atomic trajectories. This allows the average magnetic field to be measured.

The axial magnetic field variations, above the cavity, are less than 0.1 nT. This measurement allows us to calculate the average of the quadratic magnetic field with an equivalent frequency uncertainty of less than 3 10^{-16}.

The Ramsey and Rabi shifts are estimated to be negligible at the few 10^{-15} accuracy level; these two shifts can be eliminated by using atoms in $m_F=0$.

The effect of spurious lines in the microwave spectrum has been considered. We use the equation derived by Audoin et al. [10] for monokinetic atoms

$$\frac{\Delta\nu}{\nu_0}=f(\Delta,T,\tau,B,B_1) \qquad (4)$$

where Δ is the frequency difference between the central atomic resonance frequency and the spurious line frequency, T is the interrogation time and B, B_1 are respectively the microwave and the spurious field amplitudes. Considering the equation for typical fountain parameters we have a maximum shift of 2.5 10^{-16} for a spurious line 60 dB under the carrier at the optimum power ($B\tau=\pi/2$).

At very low temperature, collisions between atoms cause a large spin exchange frequency shift. Theoretical evaluation has been performed by Tiesinga et al. with a quantum calculation [11]. The collisional frequency shift is mainly proportional to the atomic density averaged over the free flight above the cavity n_{ave}. It is given by:

$$\frac{\Delta\nu}{\nu_0}\propto\frac{k\,N_i}{2(2\pi)^{3/2}\sigma_i^{\,2}\sigma_f}\propto n_{ave} \text{ if } \sigma_i<<\sigma_f \qquad (5)$$

where N_i is the initial atom number, σ_i and σ_f are the initial and final r.m.s atom cloud sizes in the cavity.

A first measurement of this shift has been performed by S. Chu and coworkers [12]. Measurements have also been made at LPTF [13]. For atoms almost equally

distributed among all Zeeman sub levels of the hyperfine level F=4 of the ground state, the measured relative collisional frequency shift is about $-10^{-21}\ n_{ave}$.(in M.O.T. operation).

Other measurements have been made with atoms prepared in F=3, giving a relative collisional shift of $\sim 5.9\ 10^{-22}\ n_{ave}$. Further investigation has to be done if the effect of the collisions is to be stated accurately, especially in molasses operation. Nevertheless we can estimate a shift of about $2\ 10^{-15}$ for the present fountain configuration in molasses operation.

We have also considered the effect of microwave leakage. Particular attention has been devoted to reducing leakage from connectors and cables. The new cavity has an improved design to reduce the leakage field by at least 100 dB. The frequency shift due to the microwave leakage depends on the microwave power. Because the atom cloud is monokinetic, the fountain can be operated with $B\tau = (2n+1)\pi / 2$, so that a relatively high microwave field can be used to estimate the level of the leakage.

V DISCUSSIONS AND PERSPECTIVES

At present, when using optical molasses, collisions constitute a limitation to the accuracy at $2\ 10^{-15}$, because we cannot measure the effect more precisely, the resolution being limited by the H-maser stability. The short term frequency stability is limited by the local oscillator frequency noise.

For the long term, some of the effects which play a role on the accuracy budget, will, at the same time limit the ultimate frequency stability of the standard. For example, the collisional frequency shift can limit the stability to 10^{-16} in two ways. The atom number fluctuations are about 10% in the long-term, additionally cavity drift changes the atomic population distribution (m_F=0) changing the collisional shift. Long term fluctuations in the average magnetic field, which are of the order of 0.1 nT, will also limit the stability to 10^{-16}.

Future improvements are possible on the accuracy side. The collisional frequency shift can be reduced to a few 10^{-16} by ejecting from the fountain the atoms which contribute to collisions but not to the detected signal. These atoms are those which are in the wings of the transverse velocity distribution. This rejection can be realized using 2D velocity selective Raman transitions [14,15].

The evaluation of our device at the few 10^{-16} level requires comparison with another standard with similar performance. A second fountain, with 2D Raman velocity selection capability, is being built in our laboratory.

ACKNOWLEDGMENTS

This work forms a part of the programme of scientific developments of the Bureau National de Metrologie (BNM). The assistance of A. H. Gerard, M Dequin, M. Lours is gratefully acknowledged. Useful discussions with A. De Marchi and G.D. Rovera are acknowledged. One of us (S. N. Lea) thanks the EC for the financial support by a Human Capital and Mobility grant. The research programme has the financial and logistic support of the French Space Agency (CNES) via the Pharao project. We also acknowledge the financial participation of the Conseil Général de la Région Ile de France.

REFERENCES

[1] K. Gibble and S. Chu, "Future Slow-Atom Frequency Standards," Metrologia, vol. 29, pp. 201-212, 1992

[2] M. Kasevich, E. Riis, S. Chu and R. De Voe, "rf spectroscopy in an atomic fountain," Phys. Rev. Lett., vol. 63, pp. 612-615, 1989.

[3] A. Clairon, Ph. Laurent, G. Santarelli, S. Ghezali, S.N. Lea, and M. Bahoura, "A cesium fountain frequency standard: preliminary results," *IEEE Trans Instrum. Meas.*, vol. IM 44, pp. 128-131, April, 1995.

[4] G.D. Rovera, G. Santarelli and A. Clairon, "Frequency multiplication chain for the atomic fountain," submitted for publication IEEE *Trans. Ultr. Ferr. Freq. Contr.*

[5] G. J. Dick, "Local oscillator induced instabilities in trapped ion frequency standards," in Proc. of *Precise Time and Time Interval*, Redendo Beach, 1987, pp.133-147.

[6] J. Viennet, C. Audoin and M. Desaintfuscien, "Cavity pulling in passive frequency standards," IEEE *Trans. Instrum. Meas.*, vol. 21, pp. 204-209, August, 1972.

[7] W.M. Itano, L. Lewis and D. Wineland, "Shift of the $^2S_{1/2}$ hyperfine splitting due to the black body radiation," Phys. Rev. A, vol. 45, pp. 1233-1235, February, 1982.

[8] A. Khursheed, G. Vecchi, and A. De Marchi, "Spatial variations of field polarization and phase in microwave cavities: application to the Cesium fountain cavity," IEEE *Trans. Ultr. Ferr. Freq. Contr.*

[9] S. Ghezali, Ph. Laurent, S.N. Lea, G. Santarelli, M. Bahoura, K. Szymaniec and A. Clairon, "A Cesium fountain frequency standard: present status and future improvements," in Proc. of 10th European Time and Frequency Forum, Besancon, 1995.

[10] C. Audoin, S. Lacey, L.S. Cutler, "Frequency offset due to spurious line in cesium beam standards," *IEEE Trans Instrum. Meas.*, vol. IM 27, pp. 325, Dec., 1978.

[11] E. Tiesinga, B. J. Verhaar, H. C. T. Stoof and D. Van Bragt, "Spin exchange frequency shift in a cesium atomic fountain," Phys. Rev. A, vol. 45, pp. R2671-R2673, March, 1992.

[12] K. Gibble and S. Chu, "A Laser-Cooled Cs Frequency Standard and a Measurement of the Frequency Shift due to the Ultra-Cold Collisions," Phys. Rev. Lett., vol. 70, pp. 1771-1774, March, 1993.

[13] S. Ghezali et al. to be published.

[14] M. Kasevich, D. Weiss, E. Riis, K. Moler, S. Kasapi, and S. Chu, "Atomic velocity selection using stimulated Raman transitions," Phys. Rev. Lett., vol. 66, pp. 2297-2300, May 1991.

[15] S.N. Lea, A. Clairon, Ch. Salomon, Ph. Laurent, B. Lounis, J. Reichel, A. Nadir and G. Santarelli, "Laser cooling and trapping of atoms: new tools for ultra-stable caesium clocks," Physica Scripta, vol. T51, pp.78-84, June, 1994.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

FURTHER INVESTIGATION OF A PROTOTYPE MICROWAVE FREQUENCY STANDARD BASED ON TRAPPED $^{171}Yb^+$ IONS

Matthew J. Sellars, Peter T.H. Fisk, Malcolm A. Lawn, Colin Coles
National Measurement Laboratory
CSIRO Division of Applied Physics
PO Box 218 Lindfield NSW 2070
Sydney, Australia

Abstract

We compare the frequency of the $^{171}Yb^+$ 12.6 GHz $M_F=0\rightarrow0$ ground state hyperfine 'clock' transition in buffer gas-cooled ion clouds confined in two similar, but not identical, linear Paul traps. After correction for the known differences between the two ion traps, including significantly different second-order Doppler shifts, the frequencies agree within an uncertainty of less than 2 parts in 10^{13}. The frequency of the clock transition in an isolated ion at zero temperature, velocity, electric field and magnetic field, was found to be
12 642 812 118.468 ± 0.0016 Hz.

Introduction

Investigations of the suitability as microwave frequency standards of ground-state hyperfine transitions ('clock' transitions) in clouds of buffer gas-cooled $^{199}Hg^+$ and $^{171}Yb^+$ ions confined in radio frequency (RF) Paul traps have raised questions concerning the frequency offsets arising from the Paul trap environment [1-9]. Some of these offsets can be relatively large, and also difficult to quantify, and may therefore limit the accuracy of a trapped-ion frequency standard. For example, the quadratic Stark shift of the 12.6 GHz ground state hyperfine transition of buffer gas-cooled $^{171}Yb^+$ ions due to the confining RF fields in a hyperbolic Paul trap has been measured as -7 parts in 10^{13} [1], with an uncertainty of approximately 50%. In the same system, the magnitude of the second-order Doppler shift is greater than 1 part in 10^{12}, depending on the operating conditions of the trap, again with an uncertainty of approximately 50%. Analogous measurements [10] on buffer gas-cooled $^{199}Hg^+$ ions in a hyperbolic Paul trap have indicated a limiting accuracy of 2.5 parts in 10^{13}.

In the case of a hyperbolic Paul trap, the node of the RF electric field, which is the bottom of the confining pseudo-potential well, is a geometric point. Due to the mutual Coulomb repulsion of the ions, an ion cloud must necessarily extend into the region where the confining RF electric field is non-zero. This results in the observed significant quadratic Stark shifts, and induces micro-motion of the ions at the RF frequency which contributes to the second-order Doppler shift as well as RF heating of the ion cloud. The linear ion trap configuration is advantageous [11] for frequency standards purposes because the node of the confining RF field is a line. This results in a substantially smaller influence of the confining RF fields on a cloud of a given number of ions than in a hyperbolic trap.

In both the hyperbolic and linear trap types the magnitudes of the second-order Doppler shift and the quadratic Stark shift depend on the physical properties of the ion cloud, which are difficult to measure directly. The calculation of the resulting frequency offsets must therefore rely on a model of the ion cloud which is based on measurable parameters, such as the trap dimensions, the temperature of the ion cloud and the frequencies (secular frequencies) at which the ions oscillate in the confining pseudo-potential well. The validity of such a model may ultimately prove to be a limiting factor for the accuracy of a trapped ion frequency standard.

We have previously reported [5,6] some aspects of the performance of a prototype frequency standard (IT-1) based on the 12.6 GHz ground state hyperfine 'clock' transition in $^{171}Yb^+$ ions confined in a linear Paul trap. In this paper we report results from a second prototype standard (IT-2), similar to IT-1 but with significant improvements in magnetic shielding and optical detection efficiency. Comparing the measured clock transition frequencies of the two traps under conditions where they contain different numbers of ions, and consequently ion clouds of different sizes, has provided a test of a model which gives values for the (different) second-order Doppler shifts in the two traps.

Experimental

The ion trap system IT-1 (fig. 1) has been described previously [12,13,5,6].

0-7803-2500-1/95 $4.00 © 1995 IEEE

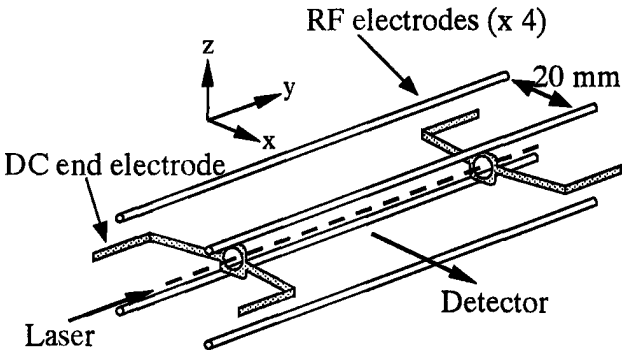


Figure 1: Schematic electrode structure of the linear ion traps IT-1 and IT-2. In this paper we refer to motion in the x-z plane as transverse, and motion in the y-z plane as longitudinal. The diameter of the RF electrodes is 2.3 mm.

For the present work the only change to IT-1 concerns the DC end electrodes, which were replaced with electrodes which approximated more closely a 6 mm ring centered on the longitudinal axis of the ion trap. An unexpected and as yet not understood side-effect of this change was a reduction in the ion storage capacity of IT-1 by a factor of approximately 2, observed in both the fluorescence intensity from the ion cloud and also in the calculated ion number. The separation of the end electrodes was unchanged.

The components of the new IT-2 system (fig. 2) inside the vacuum chamber are essentially identical to those of IT-1, with the exception of the end electrodes which are slightly different in shape. IT-2 differs significantly from IT-1 in three areas:

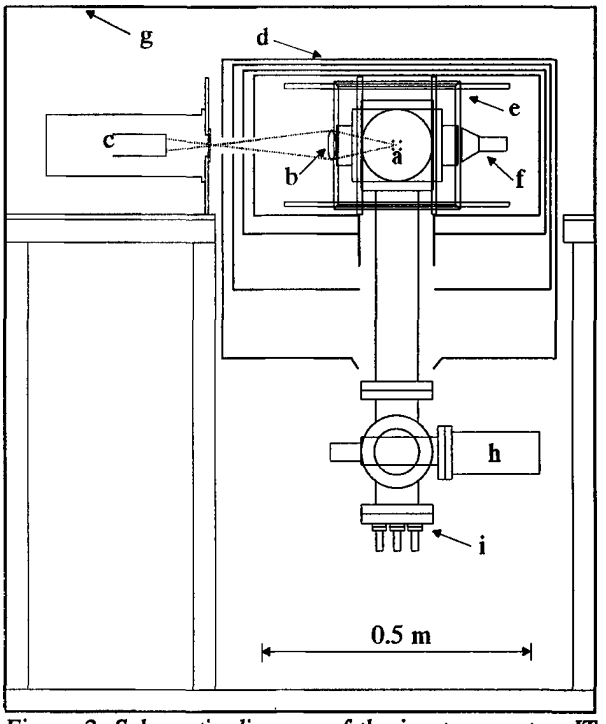


Figure 2: Schematic diagram of the ion trap system IT-2. a) Trap RF electrodes, b) Fluorescence focussing lens, c) Photomultiplier tube, d) Magnetic shielding (4 layers, see text), e) Helmholtz coils, f) Microwave horn, g) Outer magnetic shield, h) Turbo-molecular pump, i) Electrical vacuum feedthroughs.

	IT-1	**IT-2**
RF electrode separation	20 mm	20 mm
RF electrode diameter	2.3 mm	2.3 mm
End electrode separation	60 mm	60 mm
RF frequency	500 kHz	510 kHz
RF amplitude	253±1 V_{p-p}	251±1 V_{p-p}
End electrode voltage	20.0 V (stability ±0.001 V)	20.0 V (stability ±0.001 V)
He pressure	3.0±0.6 x 10^{-4} Pa	3.0±0.6 x 10^{-4} Pa

Table 1: Operational parameters applied to the ion traps for the measurements described in this paper. The uncertainty in the helium pressure is the variation observed between the readings of two independent gauge heads and gauge controllers connected to a single vacuum system.

	IT-1	IT-2
Ion cloud length	not measured	0.024 m
Ion cloud temperature	346±20 K	359±20 K
Secular frequencies (trap loaded)		
longitudinal	3.8±1 kHz	6.2±1 kHz
transverse	12±1 kHz	14±1 kHz
Helium pressure shift	1.0±0.3 mHz	1.0±0.3 mHz
Larmor frequency	74475±30 Hz	81331±20 Hz
Ramsey fringe amplitudes		
$M_F=0\rightarrow 0$ transition	3000 photon counts	19000 photon counts
$M_F=0\rightarrow -1$ transition	1400 photon counts	7000 photon counts
Laser scatter background	1000 photon counts	8000 photon counts
Raw clock frequency	12 642 812 119.3456±0.0002 Hz	12 642 812 119.5165±0.0005 Hz

Table 2: Measured trap parameters. The ion cloud length was measured using a UV-sensitive CCD camera, and the ion cloud temperature was calculated from the measured Doppler width of the optical transition [5,6]. The pressure shift due to helium was taken from a previous measurement on IT-1 [5,6]. The raw clock frequency is the uncorrected frequency of the 12.6 GHz $^{171}Yb^+$ $M_F=0\rightarrow 0$ ground-state hyperfine resonance.

	IT-1	IT-2
Trap potential well depth	0.9 eV	0.9 eV
Transverse secular frequency (empty trap)	20.3 kHz	19.8 kHz
Ion cloud radius (1/e of central density)	2.4 mm	2.1 mm
Ion number	~ 5 X 10^5	~3 X 10^5
Second-order Doppler shift	-10.3±1.2 mHz	-8.6±0.9 mHz
Stark shift	-0.15±0.07 mHz	-0.11±0.05 mHz

Table 3: Parameters calculated from the data in tables 1 and 2 according to a theoretical model of the ion cloud. The values given for the number of ions in the traps are estimates only, based on the linear ion density calculated using equation A.4, and the observed length of the ion cloud in trap 2.

1) Magnetic shielding

IT-2 has five layers of shielding consisting of an inner layer of silicon steel, surrounded by two layers of Co-Netic magnetic shielding material, surrounded in turn by another layer of silicon steel. The entire apparatus, including the photomultiplier housing, turbo-molecular pump and electrical feed-throughs, is surrounded by an outer layer of silicon steel, to form a box with edge dimensions 1.1 m x 1.4 m x 1.6 m (h). The total shielding factor of this system was measured as approximately 250.

2) Optical collection efficiency

The optical collection efficiency in IT-2 was improved over that of IT-1 by a factor of approximately 12, by using a larger diameter lens (0.01 m diameter, 0.01 m focal length) (fig. 2) to focus the fluorescence from the ion cloud onto the photomultiplier's photocathode.

3) Optical delivery

The 369 nm laser light used to probe the $^{171}Yb^+$ optical resonance transition was delivered to IT-2 via a 200 μm core diameter multi-mode optical fibre. This allowed IT-2 to be located in a different room from the laser system (frequency doubled titanium:sapphire laser pumped by an argon ion laser), and its associated major magnetic influence. Approximately 50 μW of laser light entered the ion trap.

The loading, operation and microwave interrogation sequence of both traps were essentially the same as described previously [5,6], and the operational parameters are shown in table 1. The 12.6 GHz microwave signal for each ion trap was synthesised separately (fig. 3) from a common source [5,6], based on a sapphire-loaded superconducting resonator [14]. As in the previous work, the microwave $\pi/2$ pulse duration was 400 ms, with the pulse centers separated by 25 s, yielding Ramsey fringes with a period of 40 mHz.

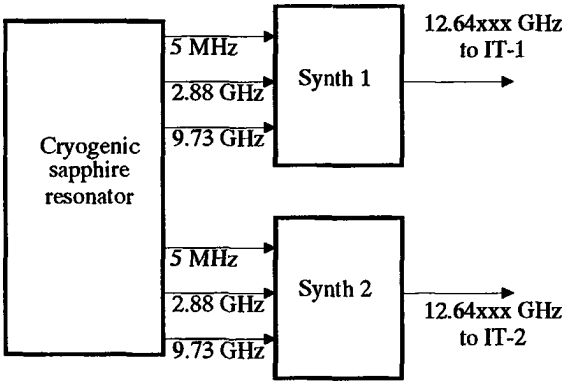


Figure 3: Schematic diagram of the system used to generate the microwave interrogation signals for the two ion traps. Both 12.6 GHz signals were gated simultaneously to generate the Ramsey interrogation pulse sequence.

The laser system, which generates the 369.4 nm radiation used to probe the populations of the $^{171}Yb^{+}$ ground state hyperfine levels following the microwave interrogation pulse sequence, is common to both ion trap systems. The laser radiation is not expected to produce any long-term offsets, correlated or otherwise, in the clock frequencies of the two systems, since the laser light is blocked during the microwave interrogation sequence [5,6]. However, despite corrections made in real-time for laser power fluctuations [5,6], it is possible that the residual effects of such power fluctuations may appear as correlated short-term frequency fluctuations. These short term fluctuations will average out over a few measurement cycles, and are not expected to affect the frequency results presented in this paper, which are averages of 400 or more measurement cycles.

Isolation of IT-1 from the microwave interrogation radiation applied to IT-2, and vice-versa, was better than 50 dB. At the present level of precision, the effects of microwave cross-talk between the two systems are not measurable. There are also a number of sidebands on the 12.6 GHz signal, due to the mixing scheme used in its generation [5,6]. The maximum shift in the measured clock frequency induced by these sidebands was calculated as 2 parts in 10^{15}, which is currently negligible.

Results

The results of measurements made on the two ion traps are shown in table 2.

The frequency of the 12.6 GHz $^{171}Yb^{+}$ $M_F=0\rightarrow 0$ hyperfine resonance in each ion trap was simultaneously measured by locking [5,6] the two 12.6 GHz signals to the Ramsey fringe pattern from each ion trap over a period of approximately 8 hours. The cycle time of the experiment was such that one frequency measurement was made on both traps every 72 s. During each measurement cycle the Larmor frequency in each trap was obtained by measuring the frequency separation of the $M_F=0\rightarrow 0$ and $M_F=0\rightarrow -1$ Zeeman components of the 12.6 GHz hyperfine transition [5,6]. The Larmor frequency in each ion trap varied by less than 80 Hz during the 8 hour period of the measurement, and the results for both the Larmor frequency and the clock frequency presented in table 2 are averaged over this period.

The longitudinal and transverse secular frequencies were measured from the microwave excitation spectra from each trap, an example of which is shown in figure 4. The different ratios of longitudinal to transverse secular frequency in the two traps are not yet understood, but are assumed to be associated with the supposedly minor differences in the shape of the end electrodes fitted to the two traps.

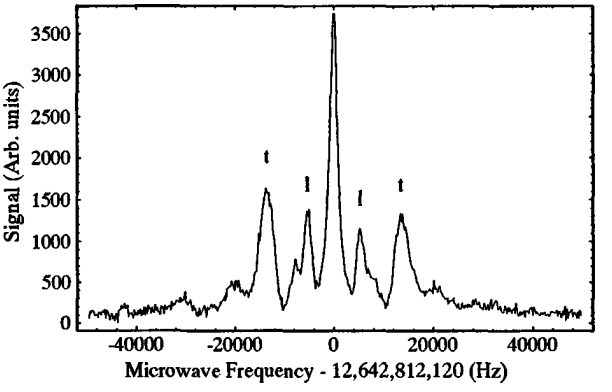


Figure 4: Sideband structure on the 12.6 GHz clock transition in IT-2 due to longitudinal (l) and transverse (t) motion of the ions .

The stability, characterised by the Allan deviation $\sigma_y(\tau)$, of the 12.6 GHz $^{171}Yb^{+}$ clock transition in IT-2 with respect to the cryogenic sapphire resonator is shown in fig. 5. For integrating times between 500 s and 5000 s the system exhibits a white noise-limited performance $\sigma_y(\tau) = 1.4 \times 10^{-13}\,\tau^{-1/2}$. For integrating times shorter than 500 s the gain of the digital servo locking the 12.6 GHz signal to the clock transition is small, and beyond 5000 s the measured performance is probably limited by the stability of the cryogenic sapphire oscillator. Direct stability comparison with a reference hydrogen maser is not yet possible at this performance level due to problems with noise acquired in the transmission (over approximately 30 m) of the hydrogen maser reference signal to the laboratory containing the ion trap and associated microwave system.

The expected performance of the system is given by [15]

$$\sigma_y(\tau) = \frac{\delta f}{\pi f_0}\left(\frac{N}{S}\right)\sqrt{\frac{T_c}{\tau}} \qquad (1)$$

where δf is the width of the central Ramsey fringe, f_0 is the frequency of the clock transition, N/S is the single

measurement cycle noise-to-signal ratio and T_c is the measurement cycle time. Using the data presented in table 2 for IT-2, a limiting performance of $\sigma_y(\tau)=8\times10^{-14}\tau^{-1/2}$ is calculated. The reason for the slightly poorer than expected performance has not yet been determined.

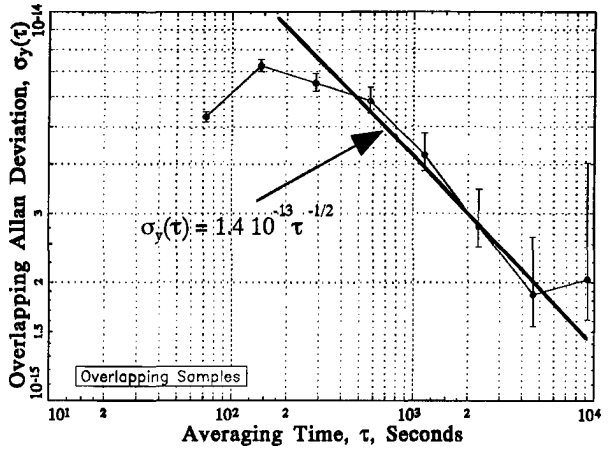


Figure 5: Fractional frequency stability, characterised by the Allan deviation $\sigma_y(\tau)$, of the 12.6 GHz $^{171}Yb^+$ clock transition in IT-2, with respect to the cryogenic sapphire resonator.

Accurate determination of the clock frequency

We now calculate the frequency, referenced to TAI, of the 12.6 GHz $^{171}Yb^+$ $M_F=0\rightarrow0$ ground state hyperfine transition in a stationary ion at zero magnetic field in a perfect vacuum.

Meis et al. [16] have presented a theoretical model of an ion cloud in a spherically symmetric potential in a hyperbolic Paul trap. In the appendix, we present a corresponding model for a linear trap. In this section we outline the use of this model, with some of the parameters in table 2, to calculate the radial density profile of the ion clouds in IT-1 and IT-2, and consequently the second-order Doppler shift and the quadratic Stark shift in the two ion trap systems.

The results of these calculations are presented in table 3. The depth of the trap potential was calculated from an analytical expression [17] for the ponderomotive potential of a infinitely long linear Paul trap. The depth depends on the geometry of the trap as well as frequency and voltage of the RF trap drive. Close to the axis of the trap the potential is approximately harmonic in the transverse direction, and can be parameterised by ω_S, the transverse secular frequency of the empty trap (equation A.1). When ions are introduced into the trap the Coulomb repulsion between them will modify the potential seen by each ion. The difference between the observed secular frequencies (table 2) and theoretical value for an empty trap (table 3) reflect this change in the potential.

The second order Doppler shift can be separated into two components, that due to the thermal motion of the ions and that due to the micromotion induced by the RF

	IT-1	IT-2
Raw clock frequency	12 642 812 119.3456±0.0002 Hz	12 642 812 119.5165±0.0005 Hz
Corrections (measured directly)		
C field (2nd order Zeeman)	-0.8783±0.0007 Hz	-1.0474±0.0005 Hz
He pressure shift	-0.001±0.0003 Hz	-0.001±0.0003 Hz
Total measured corrections	-0.8793±0.0008 Hz	-1.0484± 0.0006
Corrections (calculated according to model)		
Second order Doppler shift	0.0103±0.0012 Hz	0.0086±0.0009 Hz
Stark shift	0.00015±0.00007 Hz	0.00011±0.00005 Hz
Total calculated corrections	0.0105±0.0012 Hz	0.0087±0.0009 Hz
Correction for reference maser rate vs TAI	-0.0088±0.001 Hz	-0.0088±0.001 Hz
Corrected clock frequency (vs TAI)	12 642 812 118.4680±0.0018 Hz	12 642 812 118.4680±0.0016Hz

Table 4: Corrections applied to the raw clock frequencies.

trapping electric field. The thermal motion component is given by

$$\left.\frac{\Delta f}{f}\right|_{secular} = -\frac{3}{2}\frac{kT}{mc^2} \qquad (2)$$

where k is Boltzmann's constant, T is the kinetic temperature characterising the secular motion of the ions, m is the mass of a $^{171}Yb^+$ ion and c is the speed of light.

The micromotion component depends on the spatial distribution of the cloud in the RF trapping field. Consequently, the radial density profile (equation A.7) of the ion cloud was determined by using the measured values of the temperature and the observed transverse secular frequency ω_m, assuming an infinitely long trap, a cylindrically symmetric harmonic potential and a Maxwellian secular velocity distribution. The ion cloud radii given in table 3 are those at which the density of the cloud falls to 1/e of the central density. The micromotion contribution to the second order Doppler shift is calculated by averaging the shift over the profile of the cloud, with the result that the total second order Doppler shift is given by

$$\left.\frac{\Delta f}{f}\right|_{\substack{total \\ Second-order\ Doppler}} = -\frac{3}{2}\frac{kT}{mc^2}\left(1+\frac{2}{3}\left(\frac{\omega_s}{\omega_m}\right)^2\right), \qquad (3)$$

a result identical to that of Meis et al. The derivation of this expression therefore depends on the same assumptions used to derive the radial density profile. The uncertainties in the second order Doppler shift stated in table 3 reflect only the uncertainties in the temperature and secular frequency measurements.

The quadratic Stark shift is given by

$$\left.\frac{\Delta f}{f}\right|_{Stark} = -k_s\langle|E|^2\rangle \qquad (4)$$

where the triangular brackets denote averaging over a cycle of the RF trapping field and over the calculated profile of the ion cloud. The experimentally determined value of k_S for $^{171}Yb^+$ is $2 \pm 1 \times 10^{-21}$ m^2 V^{-2} [1]. This value for k_S may also be used to calculate [18] the fractional frequency shift due to ambient blackbody radiation. The resulting fractional shift, -1.4×10^{-15}, is presently negligible.

The corrections applied to the raw clock frequencies are detailed in table 4.

It is evident that despite the significantly different corrections for the second order Doppler shift applied to the raw frequency from each ion trap, the two final frequency values agree to within 1 part in 10^{14}, although the uncertainties in the two values are somewhat larger. Further work will be required to determine if this unexpectedly close agreement despite larger uncertainties is coincidental. Nevertheless, the present agreement is encouraging, and indicates that it will be worth testing the ability of the model of the ion cloud to predict the second-order Doppler shift over a wider range of ion cloud sizes and trap parameters.

Our measurement of the clock frequency is in good agreement with that of Tamm et al. [1], 12 642 812 118.471±0.009 Hz.

Conclusion

We have measured the frequency of the $^{171}Yb^+$ 12.6 GHz $M_F=0\rightarrow0$ ground state hyperfine 'clock' transition in buffer gas-cooled ion clouds confined in two similar, but not identical, linear Paul traps. After correction for the known differences between the two ion traps, including significantly different second-order Doppler shifts, the frequencies agree within 1 part in 10^{14}, with a statistical uncertainty in each value of about 1 part in 10^{13}. It is possible that some of the statistical uncertainties in the parameters used in the frequency-correction calculations may have been estimated too conservatively.

Our best value for the frequency of the 12.6 GHz $M_F=0\rightarrow0$ clock transition in an isolated $^{171}Yb^+$ ion at zero temperature, velocity, electric field and magnetic field, is 12 642 812 118.468 ± 0.0016 Hz. It must be stressed that the quoted uncertainty is based on the assumption that the second-order Doppler shift is given by equation 3, that is, that the ion clouds behave according to the model presented in the appendix. The good agreement between the clock frequencies in the two ion trap systems is significant, but not conclusive, evidence that this assumption is valid to within the uncertainty in the measured parameters.

Appendix
Calculation of the Second Order Doppler Shift

The calculation presented here of the second order Doppler shift for a linear ion trap is an extension to the linear ion trap case of previous work [16] on a spherically symmetric potential in a hyperbolic Paul trap.

The pondermotive trapping potential experienced by an ion of mass m close to the longitudinal axis of an infinitely long linear Paul trap can be written in the form

$$\Psi(r) = m\omega_s^2 r^2/2 \qquad (A.1)$$

where $\omega_s/2\pi$ is the transverse secular frequency and r is the distance from the longitudinal axis of the trap. In the limit where the radius r_e of the RF electrodes is small compared with their separation a, the expression for ω_s is [17]

$$\omega_s = \frac{4eV_0}{\sqrt{2}m\Omega a^2 \ln\left(\frac{a}{r_e}\right)}, \tag{A.2}$$

where V_0 is the peak voltage on the RF electrodes and $\Omega/2\pi$ is the frequency of the RF field.

Since the RF cycle average of the energy associated with the pondermotive potential is equal to the RF cycle average of the kinetic energy of the micromotion of the ions, given by $m\langle v^2\rangle/2$, the second-order Doppler shift due to the micromotion is given by

$$\left.\frac{\Delta f(r)}{f}\right|_{micro} = -\frac{1}{2}\frac{\langle v^2\rangle}{c^2} = -\frac{\Psi(r)}{mc^2} = -\frac{\omega_s^2 r^2}{2c^2} . \tag{A.3}$$

To determine the total second order Doppler shift due to the micromotion, equation A.3 must be integrated over the density profile of the ion cloud. The spatial distribution of the ion density in the cloud at a temperature T, assuming a Maxwellian distribution, is given by [19]

$$n(r) = C\exp\left(\frac{-V(r)}{kT}\right) \tag{A.4}$$

where C is a normalisation constant. V(r) is the total radial potential seen by the ion, given by

$$V(r) = \frac{m\omega_s^2 r^2}{2} + \Phi(r) = \frac{m\omega_m^2(r) r^2}{2} + \Phi(0) \tag{A.5}$$

where $\Phi(r)$ is the space charge potential due to the other ions in the trap and $\omega_m(r)/2\pi$ is the loaded-trap secular frequency, in which the radial dependence of $\Phi(r)$ has been included. If the potential is assumed to remain harmonic in the presence of the space charge, that is, if the ion cloud is not so cold that its dimensions are determined mainly by the mutual Coulomb repulsion of the ions, $\omega_m(r)$ may be approximated by the constant value ω_m, so that

$$V(r) = \frac{m\omega_m^2 r^2}{2} + \Phi(0) . \tag{A.6}$$

An expression of the ion density is obtained by substituting equation A.6 into A.4,

$$n(r) = C' \exp\left(\frac{-m\omega_m^2 r^2}{2kT}\right). \tag{A.7}$$

The second order Doppler shift due to the micromotion is then obtained using equations A.7 and A.3 and integrating over the ion cloud

$$\left.\frac{\Delta f}{f}\right|_{micro} = \frac{-\int_0^\infty 2\pi r\ n(r)\left(\frac{\omega_s^2 r^2}{2c^2}\right)dr}{\int_0^\infty 2\pi r\ n(r)\ dr} = -\frac{kT}{mc^2}\left(\frac{\omega_s}{\omega_m}\right)^2 \tag{A.8}$$

The second-order Doppler shift due to the secular motion is then given by

$$\left.\frac{\Delta f}{f}\right|_{secular} = -\frac{3}{2}\frac{kT}{mc^2}, \tag{A.9}$$

and the total second-order Doppler shift is then

$$\left.\frac{\Delta f}{f}\right|_{\substack{total\\ Second-order\ Doppler}} = -\frac{3}{2}\frac{kT}{mc^2}\left(1+\frac{2}{3}\left(\frac{\omega_s}{\omega_m}\right)^2\right). \tag{A.10}$$

This result is identical to that of Meis et al. [16] for a spherically symmetric potential in a hyperbolic Paul trap. This analytical result for the second-order Doppler shift is also in agreement with the result of the numerical calculation performed by Prestage et al. for $^{199}Hg^+$ ions in a linear Paul trap [20].

References

[1] C. Tamm, D. Schnier, and A. Bauch, "Radio-frequency laser double-resonance spectroscopy of trapped ^{171}Yb ions and determination of line shifts of the ground-state hyperfine resonance," *Appl. Phys. B*, vol. 60, 19-29, 1995.

[2] R.L. Tjoelker, J.D. Prestage, G.J. Dick, and L. Maleki, "Recent stability comparisons with the JPL linear trapped ion frequency standards," in *In Proc. 1994 International Frequency Control Symposium,* 1994, pp. 739-743.

[3] R.L. Tjoelker, J.G. Prestage, G.J. Dick, and L. Maleki, "Long term stability of Hg^+ trapped ion frequency standards," *In Proc. 47th Annual IEEE International Symposium on Frequency Control,* 132, 1993.

[4] D.J. Seidel and L. Maleki, "Progress toward the development of a ytterbium ion standard," *In Proc. 1994 IEEE International Frequency Control Symposium,* 1994, pp. 761-768.

[5] P.T.H. Fisk, M.J. Sellars, M.A. Lawn, and C. Coles, "Performance of a microwave frequency standard based on laser-detected, trapped $^{171}Yb^+$ ions," *To appear in Appl. Phys. B .*

[6] P.T.H. Fisk, M.J. Sellars, M.A. Lawn, C. Coles, A.G. Mann, and D.G. Blair, "Performance of a prototype microwave frequency standard based on trapped $^{171}Yb^+$ ions.," *Proc. 45th Annual IEEE International Symposium on Frequency Control.*

[7] A. Bauch, D. Schnier, and C. Tamm, "Frequency measurement of the hyperfine splitting of $^{171}Yb^+$ ions stored in a Paul trap," in *Proc. 7th European Frequency and Time Forum,* 1993.

[8] D.J. Seidel, A. Williams, R.W. Berends, and L. Maleki, "The development of a ytterbium ion frequency standard," *In Proc. 46th Annual IEEE International Frequency Control Symposium,* 70, 1992.

[9] R. Casdorff, V. Enders, R. Blatt, W. Neuhauser, and P.E. Toschek, "A 12 GHz standard clock on trapped ytterbium ions," *Ann. Physik,* vol. 48, no. 7, 41, 1991.

[10] L.S. Cutler, R.P. Giffard, and M.D. McGuire, "Thermalization of ^{199}Hg ion macromotion by a light background gas in an RF quadrupole trap," *Appl. Phys. B,* vol. 36, 137-142, 1985.

[11] J.D. Prestage, G.J. Dick, and L. Maleki, "New ion trap for frequency standard applications," *J. Appl. Phys.*, vol. 66, no. 3, 1013, August 1989.

[12] P.T.H. Fisk, M.A. Lawn, and C. Coles, "Laser cooling of $^{171}Yb^+$ ions in a linear Paul trap," *App. Phys. B,* vol. B57, 287, 1993.

[13] P.T.H. Fisk, M.A. Lawn, and C. Coles, "Progress at CSIRO Australia towards a microwave frequency standard based on trapped, laser-cooled $^{171}Yb^+$ ions," *In Proc. 47th Annual IEEE International Frequency Control Symposium*, 139, 1993.

[14] A.J. Giles, A.G. Mann, S.K. Jones, D.G. Blair, and M.J. Buckingham, "A very high stability sapphire-loaded superconducting cavity oscillator," *Physica B,* vol. 165-166, no. 1, 145, August 1990.

[15] A. Clairon, P. Laurent, G. Santarelli, S.N. Lea, and M. Bahoura, "A cesium fountain frequency standard : Preliminary results," *To appear in IEEE Trans. Instr. Meas.,* .

[16] C. Meis, M. Desaintfuscien, and M. Jardino, "Analytical calculation of the space charge potential and the temperature of stored ions in an RF quadrupole trap," *Appl. Phys. B,* vol. 45, 59-64, 1988.

[17] G.R. Janik, J.D. Prestage, and L. Maleki, "Simple analytic potentials for linear ion traps," *J. Appl. Phys.*, vol. 67, no. 10, 6050-605, 1990.

[18] J. Vanier and C. Audoin, *The Quantum Physics of Atomic Frequency Standards.* Adam Hilger, 1989.

[19] L.S. Cutler, C.A. Flory, R.P. Giffard, and M.D. McGuire, "Doppler effects due to thermal macromotion of ions in an RF quadrupole trap," *Appl. Phys. B,* vol. 39, 251-259, 1986.

[20] J.D. Prestage, R.L. Tjoelker, G.J. Dick, and L. Maleki, "Doppler sideband spectra for ions in a linear Paul trap," *In Proc. 1993 International Frequency Control Symposium,* 148-153, 1993.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Development of a trapped Yb^+ frequency standard: Efficient quenching of population trapping

D.J. Seidel and L. Maleki
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena, California 91109

We report on a study to determine the efficiency of various buffer gases in quenching the population trapping states of laser excited ytterbium ions. The ions were confined in an rf/dc hybrid trap, and their $^2S_{1/2} \leftrightarrow {}^2P_{3/2}$ transition was excited with 329 nm laser light. Buffer gases used include H_2, N_2, CO_2, Ar, and He at several different pressure regimes. We have identified N_2 to be highly efficient in quenching the population of the trapping states, and thus be the most suitable buffer gas to use with ytterbium for the development of a trapped ion microwave frequency standard.

The 171 isotope of ytterbium is a particularly attractive choice for trapped ion frequency standards applications. This ion has a nuclear spin I=1/2 and thus a simple hyperfine level structure with an energy splitting of 12.64 GHz. The lower lying excited electronic states of singly ionized ytterbium are accessible from the ground state with light generated by frequency doubled semiconductor lasers [1,2]. This is a particularly useful feature of the ytterbium ion in connection with the development of practical standards.

The major obstacle in the development of a frequency standard based on the ytterbium ion is the existence of population trapping states. These metastable states prevent the ion from making the desired clock transition by retaining it in an energy eigenstate for a time corresponding to the lifetime of the eigenstate. The presence of low lying metastable levels in the electronic structure of the ytterbium ion has been the subject of much study and concern [3–8]. Such long lived states can "trap" the excited population of ions involved in the optical pumping of the clock transition. The observation of a "dark" state with estimated lifetime exceeding several days has been associated with population trapping in the $4f^{13}6s^2 {}^2F_{7/2}$ state [7] and has been examined in several investigations [1,4,5,9]. Initial studies of the F state have resulted in some unanswered questions concerning the mechanism for its population; these studies have not identified an efficient scheme to de-excite the ions back to the ground state. Until now the only effective approach to avoid population trapping in the "dark" state involved using additional lasers to depopulate the metastable D-states [3–5,9–11], which are believed to be the intermediary path to the excitation of the F state.

In this paper we examine the efficacy of several different buffer gases (including H_2, N_2, CO_2, He, and Ar) in quenching the population of the "dark" state in the ytterbium ion. Nitrogen is found to be extremely efficient at quenching the dark state with an estimated quenching rate of 3.78×10^4/sec/Pa for pressures less than 2.66×10^{-4}Pa. While the influence of nitrogen on the depletion of the population of the "dark" state will be the primary focus of this paper, we will also present results obtained with other gases which point to a possible path for population trapping due to molecular formation with laser excited ions.

Buffer gas studies were performed with naturally abundant ytterbium, as well as with isotopically enriched ytterbium 171 ions confined in a hybrid linear rf/dc trap [12]. The isotopically enriched ytterbium 171 study was performed to examine the effect of collisions with the buffer gas on the coherence of the microwave transition between the hyperfine levels. The trap is housed in a stainless steel vacuum chamber evacuated to a background pressure below 7×10^{-7} Pa. H_2, N_2, CO_2, He, and Ar were introduced from a manifold connected with a high precision sapphire variable leak to the vacuum chamber.

Ytterbium ions were generated by collision of counter-propagating beams of ytterbium atoms from a tantalum tube, and electrons from a hot filament. The beam directions were perpendicular to the trap axis. During ion loading a background buffer gas pressure of 6.7×10^{-4} Pa, as read on the ion gauge, was maintained. Once a sufficiently large cloud of ions was loaded ($\sim 10^6$) both the electron beam and the atom beam were turned off. After a short relaxation time the pressure was reduced to 6.7×10^{-5} Pa and the steady state fluorescence was measured as a function of buffer gas pressure.

The fluorescence measurements were made by exciting the $4f^{14}(^1S)6s^2S_{1/2} \leftrightarrow 4f^{14}(^1S)6p^2P_{3/2}$ transition using 328.9 nm laser radiation directed along the trap axis. The fluorescence rate from the decay of the excited $4f^{14}(1S)6p^2P_{3/2}$ level was monitored perpendicular to the trap axis. The fluorescence was monitored for 30 seconds, followed by blocking the laser for another 30 seconds interval; during this time the buffer gas pressure was manually adjusted from 6.7×10^{-5} to 1.3×10^{-4} Pa. After this period the laser was unblocked again and

0-7803-2500-1/95 $4.00 © 1995 IEEE

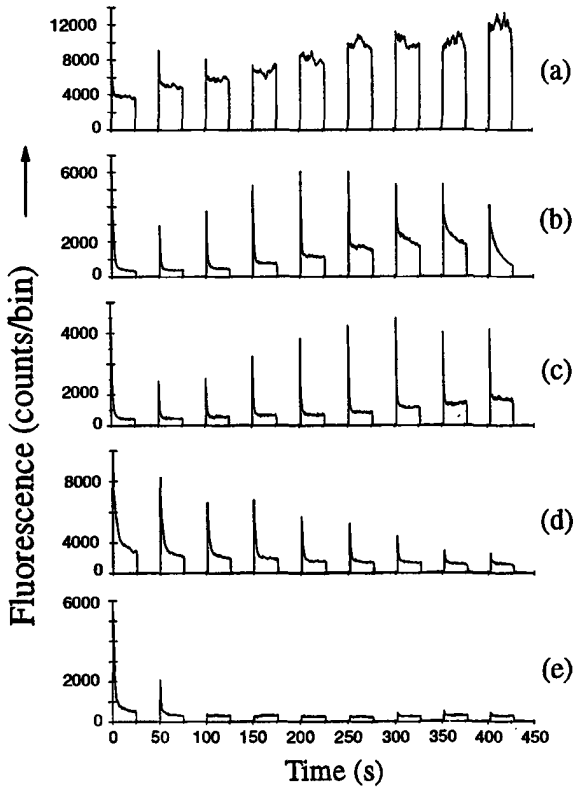


FIG. 1. Steady state fluorescence spectrum for (a) nitrogen, (b) helium, (c) argon (d) carbon dioxide and (e) hydrogen. Fluorescence rate was recorded on a multichannel scaler with a 110 msec bin width.

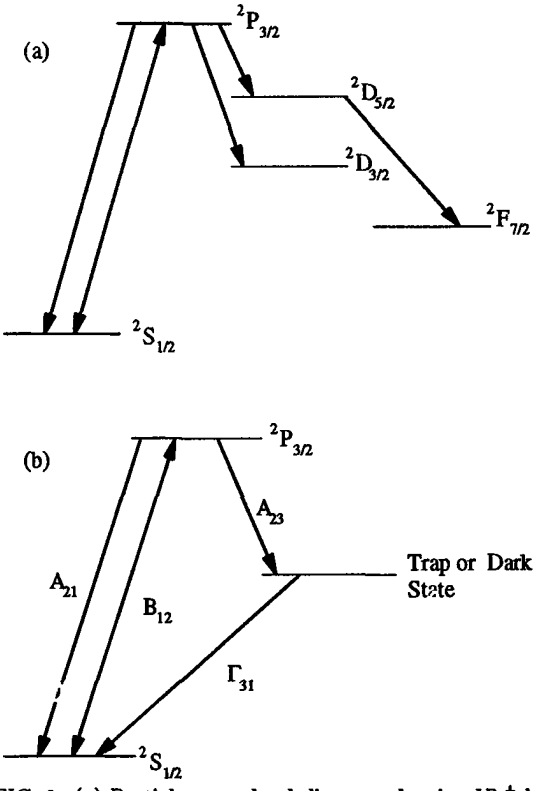


FIG. 2. (a) Partial energy level diagram showing Yb^+ levels involved in present study. (b) Further simplification where the D and F states have been combined into one state which we call the "dark" or "trap" state.

fluorescence vs. time was monitored for another 30 second interval. This procedure was continued until the fluorescence rate was measured for each of the following nine buffer gas pressures(6.7×10^{-5}, 1.3×10^{-4}, 2.7×10^{-4}, 6.7×10^{-4}, 1.3×10^{-3}, 2.7×10^{-3}, 6.7×10^{-3}, 1.3×10^{-2}, 2.7×10^{-2}) Pa. Once the steady state fluorescence measurement was completed the ions were re-loaded, to replace the ions lost at very high buffer gas pressures, and a measurement of the ion temperature was performed by scanning the laser 40 GHz at a rate of 7.63GHz/s to obtain the Doppler profile of the transition. The temperature measurements were performed at the same nine buffer gas pressures, each measurement comprising of 50 scans summed to improve the signal to noise ratio. The resonance spectrum obtained in this manner was then fitted to a sum of five Voight profiles, corresponding to the five dominant peaks present in naturally abundant ytterbium, using a commercial curve fitting program. Using the ratio of the peak fluorescence together with the temperature measurements a quenching rate for the "dark" state can be determined.

Steady state fluorescence and resonance spectra were obtained for each of the five buffer gases examined in this study. Fig. 1 show the fluorescence measurements for all five buffer gases. It is clear from the figure that N_2 is the most efficient of the buffer gases studied in its ability to quench the "dark" state. In order to obtain a more quantitative measurement of the influence of N_2 on the ytterbium ion, fluorescence vs. time measurements were repeated with higher resolution and at lower pressures. Fifteen values of N_2 pressures between 6.7×10^{-5} and 1.3×10^{-3} Pa were used to study its quenching efficacy.

Fig. 2(a) shows a partial energy level diagram for Yb^+ and 2(b) shows a further simplification where we have combined the $^2D_{3/2}$, $^2D_{5/2}$, and $^2F_{7/2}$ metastable states into a single "dark" state. This procedure is justified since the "dark" state is presumed to be resulting from population trapping via a mechanism that involves the D states, or the F state which is populated through the D states, or the combination of all three states.

Using the oscillator strengths and branching ratios quoted in Fawcett and Wilson [13] we set up a three level rate equation.

$$\dot{\mathcal{N}}_1 = -\rho(f_{12})g(f_{12})B^f_{12}\mathcal{N}_1 + \rho(f_{12})g(f_{12})B^f_{12}\mathcal{N}_2 + A_{21}\mathcal{N}_2 + \Gamma_{31}\mathcal{N}_3. \quad (1)$$

$$\dot{\mathcal{N}}_2 = \rho(f_{12})g(f_{12})B^f_{12}\mathcal{N}_1 - \rho(f_{12})g(f_{12})B^f_{12}\mathcal{N}_2 - A_{21}\mathcal{N}_2 - A_{23}\mathcal{N}_2. \quad (2)$$

$$\dot{\mathcal{N}}_3 = A_{23}\mathcal{N}_2 - \Gamma_{31}\mathcal{N}_3. \quad (3)$$

$$\mathcal{N}_o = \mathcal{N}_1 + \mathcal{N}_2 + \mathcal{N}_3 \quad (4)$$

where A an B are the Einstein coefficients, Γ_{31} is the quenching rate of the "dark" state, and $g(f_{12})$ is the lineshape factor given as the convolution of the Doppler broadening, lifetime broadening and the laser linewidth. Fawcett and Wilson quote the radiative decay rates gA_{ik}, where $g = 2J + 1$ is the statistical weight of the upper level. $A_{21} = 1.189 \times 10^{8}$ and we set $A_{23} = A(4f^{14}6p^2P_{3/2} \rightarrow 4f^{14}5d^2D_{3/2}) + A(4f^{14}6p^2P_{3/2} \rightarrow 4f^{14}5d^2D_{5/2}) = 3.383 \times 10^5 + 1.458 \times 10^6 = 1.796 \times 10^6$. B_{ik} is related to the A coefficient through the relation

$$B_{ik} = \frac{\lambda^3}{8\pi h}A_{ik}. \quad (5)$$

Since the laser linewidth ($\sim$ 1 MHz) is much smaller than either the Doppler broadening ($\sim$ 1 GHz) or the lifetime broadening ($\sim$ 19 MHz) of the S to P transition we assumed the laser to be perfectly monochromatic, reducing $g(f_{12})$ to a convolution of only the lifetime and Doppler broadening. The lifetime lineshape is given by the lorentzian

$$G_{LT}(f, f_o, \Delta f_{LT}) = \frac{1}{2\pi}\frac{\Delta f_{LT}}{(f_o - f)^2 + (\frac{\Delta f_{LT}}{2})^2} \quad (6)$$

while the Doppler broadening is give by the gaussian

$$G_d(f', f_o, \Delta f_d) = \sqrt{4\frac{\ln 2}{\pi}}\frac{1}{\Delta f_d}e^{-4\ln(2)\left(\frac{f'-f_o}{\Delta f_d}\right)^2}. \quad (7)$$

The resulting Voight profile is given by

$$\mathcal{V}(f, f', \Delta f_{lt}, \Delta f_d) = \frac{1}{2\pi}\sqrt{\frac{4\ln 2}{\pi}}\frac{\Delta f_{lt}}{\Delta f_d} \times \int_{-\infty}^{+\infty}\frac{e^{-4\ln 2(\frac{f'-f_o}{\Delta f_d})^2}}{(f_o - f)^2 + (\frac{\Delta f_{lt}}{2})^2}df_o. \quad (8)$$

Here Δf_d is the Doppler full width at half maximum which we are able to obtain through the resonance fluorescence data fitted with the sum of five Voight profiles as previously discussed (see fig. 3). Δf_{lt} is the lifetime width and is given by$A_{12}/2\pi$. f' is the center of the Doppler broadened transition where we have shifted the origin to set $f' = 0$ and also set the laser frequency f to overlap the center of this transition. Throughout the experiment the laser frequency is monitored with the aid of the ytterbium hollow cathode lamp. With $f' = 0$ and $f = 0$ the lineshape factor $g(f) = g = \mathcal{V}(f, f', \Delta f_{lt}, \Delta f_d)$

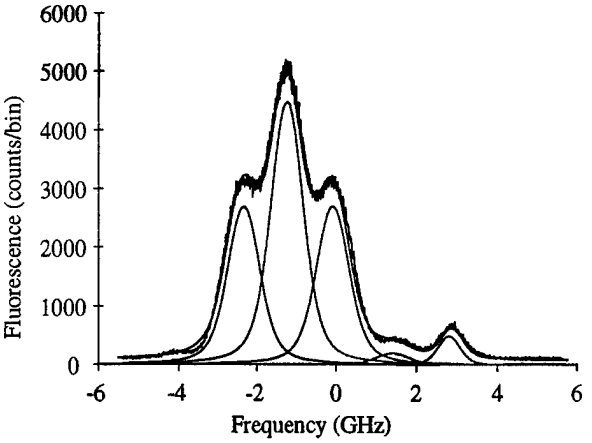


FIG. 3. Resonance spectrum of excitted ytterbium at 1.3×10^{-2} Pa of N_2, fit to a sum of five Voight profiles in order to obtain the doppler width at half maximum and thus the temperature.

in equations 1 and 2. Steady state solutions of equation 1-4 may be obtained by letting $\dot{\mathcal{N}}_1 = \dot{\mathcal{N}}_2 = \dot{\mathcal{N}}_3 = 0$ and by rearranging the equations to get

$$\mathcal{N}_2 = \frac{pgB^f_{12}\Gamma_{31}\mathcal{N}_o}{pgB^f_{12}(2\Gamma_{31} + A_{23}) + \Gamma_{31}(A_{21} + A_{23})}. \quad (9)$$

The $\mathcal{N}_2$ level population is the only one of importance since the observed fluorescence level is directly proportional to this population. Several time constants are necessary to fit the fluorescence rise and decay curves. The initial peak fluorescence is dominated by a very fast time constant in which an equilibrium is reached between the ground state population depleted by the laser and the decay of the P state via the various available decay channels. This peak fluorescence is proportional to the maximum value of $\mathcal{N}_2$ ($\mathcal{N}_2^{max}$) given by setting $A_{23}/\Gamma_{31} = 0$ in eqn. 6. This is justified since the ratio between A_{21} and A_{23} is large (66:1), which in turn implies that the peak population in level 2 is approximately equal to the steady state population in that level if Γ_{31} is very large. This is followed by a much slower time constant in which fluorescence is lost due to ions getting shelved into a metastable state. The steady state level beyond several seconds is determined by the actual value of Γ_{31}. The measured ratio $\mathcal{R}$ of the steady state fluorescence vs. the peak height is therefore equal to $\mathcal{N}_2/\mathcal{N}_2^{max}$, from which we can determine Γ_{31} as follows:

$$\Gamma_{31} = \frac{pgB^f_{12}A_{23}\mathcal{R}\mathcal{N}_2^{max}}{pgB^f_{12} - (2pgB^f_{12} + A_{21} + A_{23})\mathcal{R}\mathcal{N}_2^{max}}. \quad (10)$$

A plot of Γ_{31} vs. pressure is shown in Fig. 4. The data fit well to a polynomial of order 2 reflecting thermalizing effects [14,15] and multiple collisional effects that may be present. We have fit the first seven data points to a

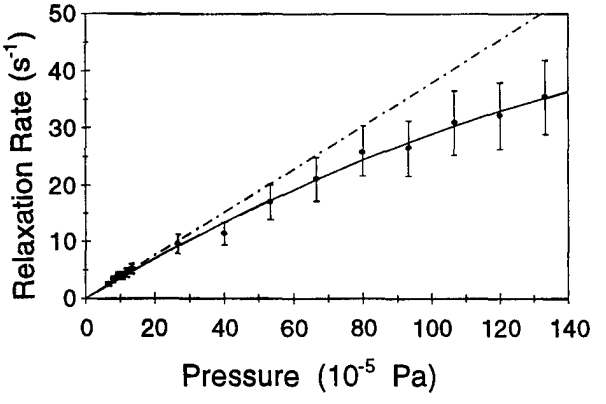


FIG. 4. Quenching rate of ytterbium's trap state vs. N_2 buffer gas pressure. The curved line represents a least squares fit of the data to the polynomial $\Gamma_{31} = aP + bP^2$; $a = 3.63 \pm 0.12 \times 10^4/s/Pa$ and $b = -7.45 \pm 1.15 \times 10^6/s/Pa^2$, while the straight line represents a linear least squares fit to pressures $\leq 26.6 \times 10^{-5}$Pa.

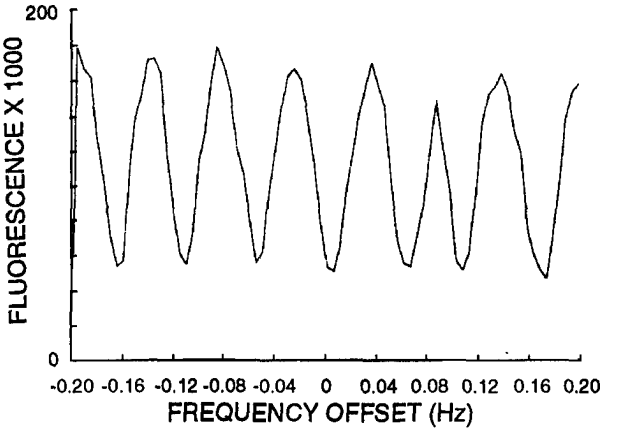


FIG. 5. 25 mHz ramsey fringes corresponding to a 20 second separation between $\pi/2$ pulse.

straight line to obtain a quenching rate $\Gamma_{31} = (3.78 \pm 0.99 \times 10^4)/s/Pa$ at low pressures ($P \leq 2.66 \times 10^{-4}$Pa). The uncertainty corresponds primarily to fluctuations in the laser power.

Highly effective quenching of metastable states in many of the alkali-earths with N_2 as a buffer gas has been known for many years and has been attributed to large cross sections for energy transport between the alkali-earths' electronic energy into vibrational and rotational states of the nitrogen molecule [14–16]. We believe the same mechanism is responsible for quenching of the alkali-like singly ionized ytterbium ion by the N_2 buffer gas.

As the data displayed in Fig. 1 shows He, Ar and CO_2 exhibit a considerably smaller efficiency for depleting the population of the dark state as compared to N_2. The absence of quenching efficiency with argon and helium is not surprising since the noble gases do not have the vibrational and rotational structure of diatomic or polyatomic molecules and therefore may only collisionally transfer the electronic energy into kinetic energy, a process with a much smaller cross section. A previous study of ytterbium ion's "trap" state lifetime as a function of buffer gas pressure using He, Ne, and Ar indicated relaxation times on the order of minutes even at pressures approaching 1 Pa [7].

It was however surprising to see that neither CO_2 or H_2 were efficient at quenching the "dark" state population. In fact, with hydrogen as the buffer gas the fluorescence quickly decayed to the level of the background light and could not be re-established even after extremely long dark periods ($\sim$ several hours). This would imply that either the trap time with hydrogen is short or hydrogen is particularly ineffectual in reducing the duration of the dark period. We verified that the trap time with hydrogen is quite long by waiting several hours, after initially loading the ions, before exposing the ions to the uv radiation. Since hydrogen is quite effective in quenching metastable states in other excited ions and atoms, [6,14,17] and has been shown to be nearly an order of magnitude more efficient than nitrogen in quenching the $^2D_{3/2}$ state in Yb^+ [6], we conjecture that molecular formation between hydrogen and excited ytterbium ions may be occurring. CO_2 proved to be a very ineffective buffer gas yielding very short trap times, thus being ruled out as a candidate buffer for use in the ytterbium ion frequency standard.

A dual decay time characteristic was exhibited when helium was used as the buffer gas at pressures of 6.7 x 10^{-3} Pa and greater. A nonlinear behavior of the decay rate of the $^2D_{3/2}$ in Yb^+ at high He pressures has been previously observed [6,18] and discussed in detail in ref [18]. We also found that at very high He pressures the trap time was greatly reduced and a competing effect between quenching the "dark" state and loss of ions from the trap appeared to occur.

Since our main objective is to exploit the 12.64 GHz transition between the hyperfine levels of the ground state of $^{171}Yb^+$ we performed microwave optical double resonance spectroscopy using N_2 as a buffer gas and found no loss of coherence in the ground state hyperfine levels due to the N_2 buffer gas. In fact we were able to transfer the population of the ions between the lower and upper hyperfine states over a 25 second period using the Ramsey separated pulse technique which yielded a linewidth of 20 mHz (see fig. 5), corresponding to a line Q of 6.3 x 10^{11}. Furthermore, nitrogen was so efficient at quenching the "dark state" that we were able to obtain single shot photon counts of 130,000 above a 50,000 count background, corresponding to a signal-to-noise ra-

tio exceeding 350:1 in the shot noise limit. A potential stability of 1.8 x $10^{-14}/\sqrt{\tau}$ is thus realizable with the present trapping configuration [19].

In summary we have determined a highly effective means for quenching the "trap" state in the ytterbium ion through introduction of N_2 into the vacuum system. A quenching rate of $(3.63 \pm 0.12 \times 10^4)/\mathrm{s/Pa} - (7.45 \pm 1.15 \times 10^6)/\mathrm{s/Pa}^2$ was determined by measuring the ratio of the peak fluorescence to the steady state fluorescence level of naturally abundant ytterbium irradiated with 329nm radiation. We found no loss of ion coherence due to the presence of nitrogen and were able to cool the ion temperature down to 600 kelvins. This approach enables the realization of the full potential of ytterbium as the active ion in a trapped ion frequency standard.

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautic and Space Administration.

---

[1] Chr. Tamm. *Appl. Phys. B*, **B56,** 295, May 1993.
[2] A. Williams, D.J. Seidel, and L. Maleki. in *OSA Proc on Advanced solid-State Lasers*, **15,** 250, 1993.
[3] D.J. Seidel, A. Williams, R.W. Berends, and L. Maleki. in *Proc. 46th Annu. IEEE Frequency Control Symposium*, page 70. IEEE Publication 92CH3083-3, 1992.
[4] A. Bauch, D. Schnier, and Chr. Tamm. *J. Mod. Opt.*, **39,** 389, 1992.
[5] A.S. Bell, P. Gill, H.A. Klein, A.P. Levick, and W.R.C. Rowley. *J. Mod. Opt.*, **39,** 381 1992.
[6] Ch. Gerz, J. Roths, F. Vedel, and G. Werth. *Z. Phys. D*, **8,** 235, 1988.
[7] H. Lehmitz, J. Hattendorf-Ledwoch, R. Blatt, and H. Harde. *Phys. rev. Lett.*, **62,** 2108, 1989.
[8] J. Yoda and K. Sugiyama. *J. Mod. Opt.*, **39,** 403, 1992.
[9] H. A. Klein, A. S. Bell, G. P. Barwood, and P. Gill. *Appl. Phys. B*, **50,**13, 1990.
[10] P.T.H. Fisk, M.A. Lawn, and C. Coles. *App. Phys. B*, **B57,** 287, 1993.
[11] A.S. Bell, P. Gill, H.A. Klein, A.P. Levick, Chr. Tamm, and D. Schnier. *Phys. Rev. A*, **44,** R20, 1991.
[12] J.D. Prestage, G.J. Dick, and L. Maleki. *J. Appl. Phys.*, **66,** 1013, 1989.
[13] B.C. Fawcett and M. Wilson. *At. Data and Nucl. Data Tables*, **47,** 241, 1991.
[14] E.S. Hrycyshyn, and L. Krause. *Can. J. Phys.*, **48,** 2761, 1970.
[15] E.A. Gislason, A.W. Kleyn, and J. Los. *Chemical Physics*, **59,** 91, 1981.
[16] E.R. Fisher and G.K. Smith. *App. Opt.*, **10,** 1803, 1971.
[17] A.A. Madej and J.D. Sankey. *Phys. Rev. A*, **41,** 2621, 1990.
[18] K. Sugiyama and J. Yoda. *IEEE Trans. on Instr. and Meas.*, **51,** 414, 1990.
[19] D.J. Seidel and L. Maleki. in *Proc. 48th Annu. IEEE Int. Frequency Control Symposium*, 1994 (to be published).

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A MERCURY LINEAR ION TRAP FREQUENCY STANDARD FOR THE USNO*

R. L. Tjoelker, J. D. Prestage, L. Maleki

California Institute of Technology, Jet Propulsion Laboratory
4800 Oak Grove Drive, Bldg 298
Pasadena, California 91109

Abstract

A frequency standard for the USNO based on mercury ions in a Linear Ion Trap is operational. Initial measurements of the new standard LITS-3 against an active hydrogen maser gives a fractional frequency stability of $\sigma_y(\tau)=7.5 \times 10^{-14}/\tau^{1/2}$ for measuring intervals up to 10,000 seconds. Stability measurements for longer averaging intervals against another ion trap standard have commenced.

Introduction

One of the most important roles of atomic frequency standards is to provide for long term definition and maintenance of time scales. The USNO maintains the United States master clock, providing the national time reference and center of all Department of Defense timed systems. It consists of a H-maser steered to a mathematical time scale based on a large number of different clocks, primarily hydrogen masers. For an individual clock the major source of frequency error for long averaging times are systematic in nature. The achievable long term stability is directly related to the size of the frequency offsets and the degree they are measured and controlled.

Mercury trapped ion standards are particularly attractive for long term stability needs because of the large atomic mass and ground state hyperfine splitting (≈40.5 GHz). Sensitivity to thermal effects and magnetic field fluctuations are reduced from standards using lighter atoms. Confinement in an ion trap eliminates hard to control wall shifts and allows long interrogation times. Continuous operation is practical using a ^{202}Hg lamp for atomic state selection and helium buffer gas for ion cooling. These advantages were recognized some years ago and trapped ion standards developed by Hewlett Packard for the USNO based on the traditional hyperbolic ion trap [1,2]. More recently, the development of the linear ion trap [3] at JPL provided for significantly improved short term performance of $\sigma_y(\tau)= 7\times10^{-14}/\tau^{1/2}$ [4] and an improved capability for measuring sensitivity to long term perturbations.

Long Term Stability Measurements of LITS-2

Typical operating parameters and environmental sensitivity of our first two research standards LITS-1 and LITS-2 have been previously published [4,5]. The original research standards have no direct regulation of ion number, and have three magnetic shields providing a shielding factor of 800. Measurements were performed in a thermal environment controlled to 0.1 °C.

Figure 1 shows the first measurement between LITS-2 and LITS-1 over a continuous 24 day interval. The measurement was performed in an environment where occasional movement of a steel electronic rack perturbed the frequency of LITS-1. With a shielded magnetic sensitivity of 5×10^{-16}/mG, the stability around 100,000 seconds, typically about 5×10^{-16} [5], is degraded in part due to magnetic perturbations.

Also shown in Figure 1 is the stability of LITS-2 measured against two hydrogen masers, SAO-26 [6] and a cavity tuned maser STSC-1 [7]. The average differential drift between LITS-2 and the STSC-1 maser over the 146 day measurement interval was measured to be $(2.1\pm0.8)\times10^{-16}$ /day. The measured long term stability of the hydrogen masers shown in Figure 1 is characteristic of the technology [8,9]. Though little effort was spent regulating and isolating LITS-1 and LITS-2 from environmental perturbations, the ion trap stability exceeds H-maser stability for averaging times between 10,000 and 10^6 seconds.

Initial Measurements with LITS-3

In the current mode of operation, the largest sensitivity that could lead to long term variation of the 40.5 GHz frequency is to variations in the external magnetic field and/or the trapped ion number (through the second order Doppler shift). A number of improvements have been incorporated into the standard for the USNO (LITS-3) including improved stability of the trapping potentials and electron current to load ions. The trap resides in five layers of magnetic shielding, increasing the longitudinal shielding factor from 800 to 17,000. The bias magnetic field has been reduced from 80 to 40 mG, reducing the sensitivity to ambient magnetic fluctuations to 1×10^{-17}/mG. The enclosure containing the trap

0-7803-2500-1/95 $4.00 © 1995 IEEE

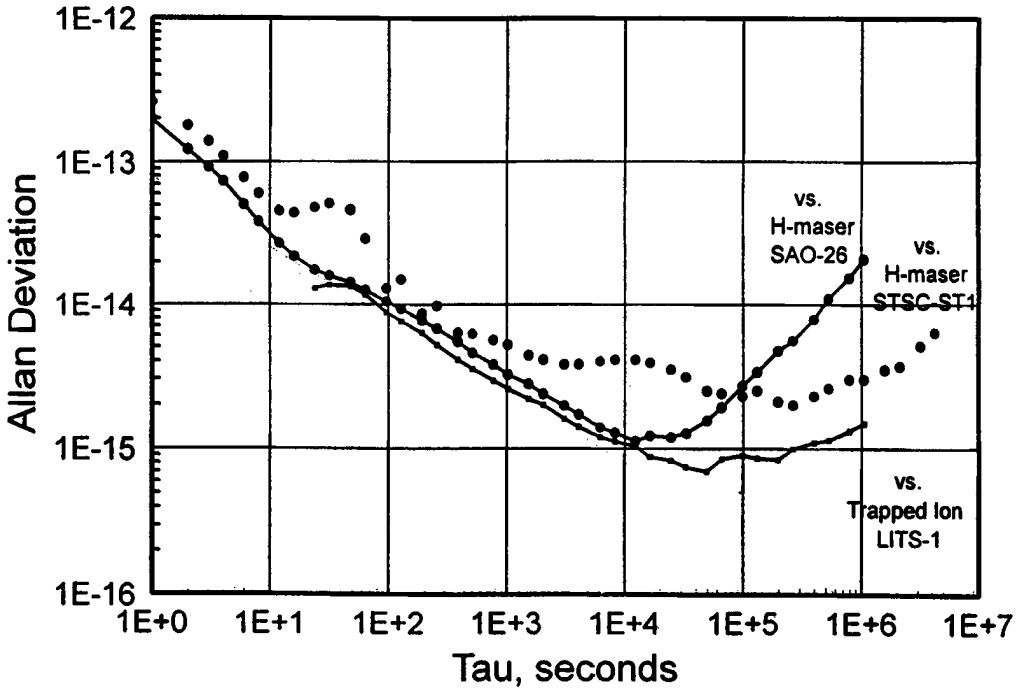


Figure 1: Fractional frequency stability of the research Hg+ standard LITS-2 compared to (a) the ion trap standard LITS-1 for 24 days, (b) a cavity tuned Hydrogen maser for 146 days, and (c) an active Hydrogen Maser.

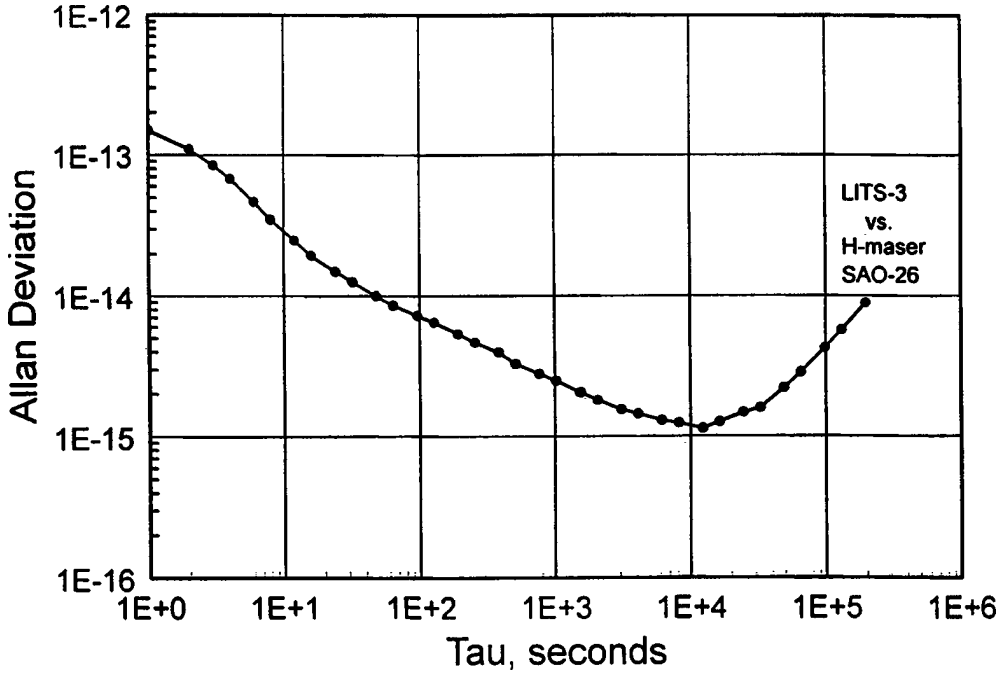


Figure 2: Short term stability of LITS-3 with a H- maser. A short term stability of $\sigma_y(\tau) = 7.5 \times 10^{-14}/\tau^{1/2}$ is shown. The stability measurement past 10,000 seconds is limited by the H-maser reference standard.

and vacuum assembly is thermally regulated to 0.1 °C allowing the frequency standard to be portable.

In Fig. 2, a five day measurement of the new standard against an active H-maser is shown. The short term Allan deviation shown of $\sigma_y(\tau)= 7.5 \times 10^{-14}/\tau^{1/2}$ is accomplished with an 8 second microwave interrogation time, a factor of two shorter than used in LITS-1 and LITS-2 for a comparable stability [4]. Beyond 10,000 seconds the measurement is limited by the stability of the hydrogen reference maser. Long term stability measurements of LITS-3 compared with other ion trap standards have commenced.

In the future, improving the magnetic field homogeneity should allow operation at still lower fields, further reducing magnetic sensitivity. Using the new extended trap (LITE) configuration [10], the magnetic shields will be much smaller. The LITE also allows performing the microwave interrogation at reduced linear ion densities. This reduces sensitivity to ion number fluctuations without compromising signal to noise and should provide even better long term stability.

*This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration.

References

[1]L. S. Cutler, R. P. Giffard, and M. D. McGuire, "Thermalization of ^{199}Hg Ion Macromotion by a Light Background Gas in an RF Quadrupole Trap", *Appl. Phys. B* **36**, 137-142, 1985.

[2] D.N. Matsakis, A.J. Kubik, J.A. DeYoung, R.P. Giffard, and L.S. Cutler, "Eight Years of Experience with Mercury Stored Ion Devices", these proceedings, (1995).

[3] J. D. Prestage, G. J. Dick, L. Maleki, "New Ion Trap for Frequency Standard Applications", *J. Appl. Phys.* **66**, No. 3, 1013-1017, August 1989.

[4] R.L. Tjoelker, J.D.. Prestage, G.J. Dick, L. Maleki, "Long Term Stability of Hg^+ Trapped Ion Frequency Standards", Proc. 47th Ann. Symp. Freq. Control, 132-138, June 1993.

[5] R.L. Tjoelker, J.D. Prestage, G.J. Dick, L. Maleki, "Recent Stability Comparisons With The JPL Linear Trapped Ion Frequency Standards", Proc. 48th Ann. Symp. Freq. Control, 739-743, June 1994.

[6] Smithsonian Astrophysical Observatory; Cambridge, Massachusetts 02138 USA.

[7] Sigma Tau Standards Corporation; Tuscaloosa, Alabama 35403 USA.

[8] Lee A. Breakiron, "A Comparative Study of Clock Rate and Drift Estimation", Proceedings of the 25th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Dec. 1993.

[9] Marc A. Weiss, Fred Walls, "Preliminary Evaluation of Time Scales Based on Hydrogen Masers", Proc. 8th European Frequency and Time Forum, 1994.

[10] J.D. Prestage, R..L. Tjoelker, G.J. Dick, L. Maleki, "Improved Linear Ion Trap Physics Package", Proc. 47th Ann. Symp. Freq. Control, 144-147, June 1993;
also see "Progress on the Linear Ion Trap Extended(LITE)II", these proceedings, 1995.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Progress Report on the Improved Linear Ion Trap Physics Package[1]

John D. Prestage, Robert L. Tjoelker, G. John Dick, and Lute Maleki

California Institute of Technology
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109

Abstract

This article describes the first operational results from the extended linear ion trap frequency standard now being developed at JPL. This new design separates the state selection/interrogation region from the more critical microwave resonance region where the multiplied local oscillator (LO) signal is compared to the stable atomic transition. Hg^+ ions have been trapped, shuttled back and forth between the resonance and state selection traps. In addition, microwave transitions between the Hg^+ clock levels have been driven in the resonance trap and detected in the state selection trap.

Introduction

Ions can be readily moved along the axis of a long linear trap [1] by applying a simple dc bias. This is a feature of ion trap based standards which has not been exploited. There are clear advantages for an Hg^+ trap layout where the ions are electrically moved from a state selection region into a compact microwave resonance region and then back for state interrogation.

In a frequency standard based on this concept [2,3], short and long term stability can be improved for several reasons. Magnetic perturbations and gradients caused by the optical state selection components, most notably the magnetic photomultipliers, can be eliminated since the optical and resonance regions are apart. Narrow resonances in the Hg^+ field sensitive transitions will enable measurement of the magnetic field in the resonance region and thus allow long term stabilization of this field. The volume of the region enclosed by the magnetic shields are reduced by a factor of 100 from that of the previous LIT design. For the present design the magnetic shields are 5 cm inside diameter by 37 cm so that multilayering to achieve high shielding factors will not add much bulk or cost.

The 2nd order Doppler shift from the finite size cloud, proportional to the linear ion density N/L, can be reduced greatly by increasing the length, L, of the resonance trap. This frequency shift can be the source of clock instability if the number of trapped ions varies over time. The present trap, described in this paper, is about 4 times longer than the previous LIT design and should reduce number sensitivity by the same factor.

One of the limitations to short term stability in the present LIT standards [4] is the relaxation of the population difference between clock states during the microwave interrogation with the multiplied LO. This population difference is created by the optical state selection process which must be turned off during microwave interrogation because the UV light will shift and broaden the clock resonance and, therefore, degrade clock performance. The neutral Hg vapor which is necessary for ion creation will, over a period of ten seconds or so, replace (via charge transfer) a state selected ion with an ion in an arbitrary state of polarization. This degrades the resonance signal size, especially for the high Q measurements. One remedy for this is to remove the neutral Hg vapor from the resonance region by reducing the temperature of the vacuum walls in this region. Neutral Hg will be cryo-pumped (LN_2 temperature or higher) onto the walls and should lead to much improved signal size and clock performance. Because the optical and resonance regions are separated, the optical system and trap can be at room temperature.

[1] This work represents one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration.

0-7803-2500-1/95 $4.00 © 1995 IEEE

Improved LIT Design

Figure 1 is a schematic view of the extended linear ion trap inside its vacuum container. A dc break forms a junction between the loading/state-selection trap (upper) and the resonance trap (lower) and allows ions to be held in either trap region and to pass from one region to another. The rf trapping voltage for transverse containment of the ions is continuous across this dc break. When the dc voltage level of all four trap rods are the same in both the upper and lower regions thermal motion of the ions will carry them through the junction with no change in axial velocity. When the four trap rods in the upper region are at positive dc voltage with respect to the lower region trap rods, ions within a trap radius or so of the junction will be transported across the junction into the lower region. Only the ions near the junction will experience the electric field forcing them across the gap. Since each ion is in thermal motion along the axis of the trap it will reach the junction within a trap length transit time (typically about 1 millisecond) and then be pulled into the upper region emptying the lower region of ions. Similarly, when the lower region is dc biased positive with respect to the state selection region all ions will be transported from the resonance region to the state selection region. The helium buffer gas will provide the required energy damping with a 1/3 second time constant[5]. A ceramic ring at the junction of the two traps, supports on its inside diameter, the ends of each of the 8 rods from the two traps. The outside diameter of this same ring centers the trap inside the vacuum jacket. Similar rings are used at each end of the extended trap. These rings have a metallized coating to prevent charging and can be electrically biased from outside the vacuum. The ring at the trap junction, when biased to a few volts positive, acts as a partition between the two traps preventing ions from moving accross the junction.

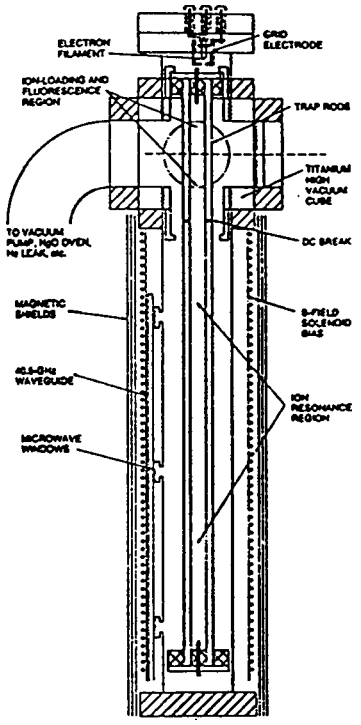


Figure 1: Improved physics unit for the trapped ion frequency standard.[199]Hg ions are created and state prepared in the upper region of the trap, then transported into the lower region, where the Hg^+ atomic resonance frequency is compared with the local oscillator frequency. The ions are then moved back to the state selection region to check for frequency detuning from the LO during the resonance comparison. Overall dimensions are 10 cm by 50 cm.

Figure 2 shows one of the early measurements of ion fluorescence vs time as ions are shuttled back and forth between the two trap regions as just described. Ions only fluoresce when in the upper trap thus the fluorescence switches on and off corresponding to ion location. When the ions are in either trap, the partition electrode is biased to 10 volts or more with respect to the rf trap rods. The sequence of applied voltages used to move the ions from the state selection trap to the resonance trap follows: (1) the partition electrode is grounded; (2) a 1.5 volt dc bias on the 4 resonance trap rods is ramped to 0 volts in about 0.1 second; (3) the dc bias on the 4 state selection trap rods is ramped from 0 to 1.5 volts, and; (4) the partition electrode is again biased to 10 volts.

Following step (2) but before step (3) the ions move freely between the two traps as though it were one long trap. Following step (3) the ions are confined to the resonance trap. The partition serves to isolate the two traps so that the electron pulse can load more ions into the loading trap while the bulk of ions are confined to the resonance trap. This is done during the "fluorescence off" period of Fig. 2. The resonant 40.5 Ghz microwave pulse is continuously on during this measurement to prevent optical pumping and the consequent reduction of atomic fluorescence.

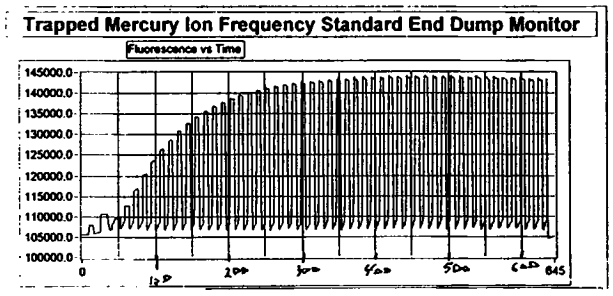


Figure 2: Ion population build-up while shuttling ions from resonance trap to state-selection trap and firing electron gun once per cycle to generate ions.

The ions spend ≈9.5 seconds in each trap before being moved. The ion population builds from an empty trap to the steady state population where loss rate equals load rate. The trap lifetime for an Hg ion is about 500 seconds for the measurements reported here. This is shorter than previous traps and probably stems from a small leak which has since been repaired. There is no significant loss of ions during the transfer from one trap to another.

Figure 3a and 3b show the signature of the F=0 to F=1 clock transition driven in the resonance region and detected after transferring the ions into the state-selection region. Consider first Fig. 3a again showing fluorescence vs. time. Following ten steps of no ions in the trap, the electron pulse fires for 100 steps loading ions into the state-selection trap (each step lasts for 1.6 seconds). During this 100 step interval there is a residual fluorescence from the ions even though they are largely pumped into the non-fluorescing F=0 state. This fluorescence is proportional to ion number and shows the trap filling while the electron pulse is on. Ten steps after the electron pulse stops, the microwave pulse switches on causing a 23,000 per step fluorescence increase from the ions as they are held in the loading trap. Ten steps after the microwaves are switched off, the ions are moved to the resonance trap where they are held for one step and then moved back. After 10 steps in the state selection trap they are again moved to the resonance trap and this time held for 2 steps. This shuttling of ions continues, each time the ions remain one step longer in the resonance trap, away from the state selecting light which pumps the ions into the F=0 non-fluorescing state. While the ions are in the resonance region the level population relaxes toward all four hyperfine levels (three F=1 states and one F=0 state) being equally populated. In the measurement of Fig. 3a this relaxation shows up as an "overshoot" in the ionic fluorescence for the first step following the return of the ions to the state selection trap. This overshoot grows with the time spent in the resonance trap and is probably due to charge transfer between the neutral Hg vapor and the state prepared ions. Also notice that since there is no overshoot for short times in the resonance trap, the transfer process from the state-selection trap to the resonance trap and back does not induce any relaxation across the 40.5 Ghz clock transition.

Fig. 3b shows the same sequence as 3a with one important difference. While the ions are in the resonance trap the 40.5 Ghz microwave pulse is applied driving the transition F=0 to F=1. The ions then return to the state-selection trap with up to 50% of the ions in the upper clock state which readily fluoresces in the state selection UV light beam showing the much larger "overshoot" in light scatter. As the ions are optically pumped back into the non-fluorescing F=0 state, this overshoot fluorescence decays in a single step.

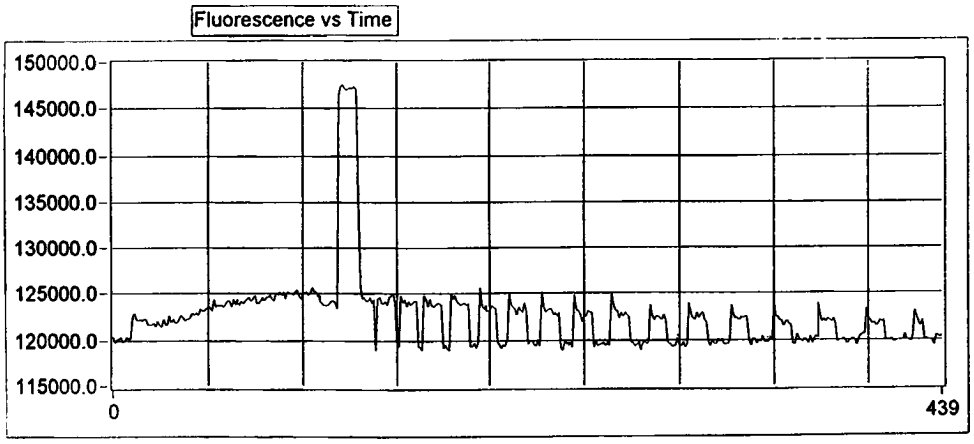


Figure 3a: Ion fluorescence shows relaxation of the population difference while ions are in the resonance trap and out of the state-selecting UV light.

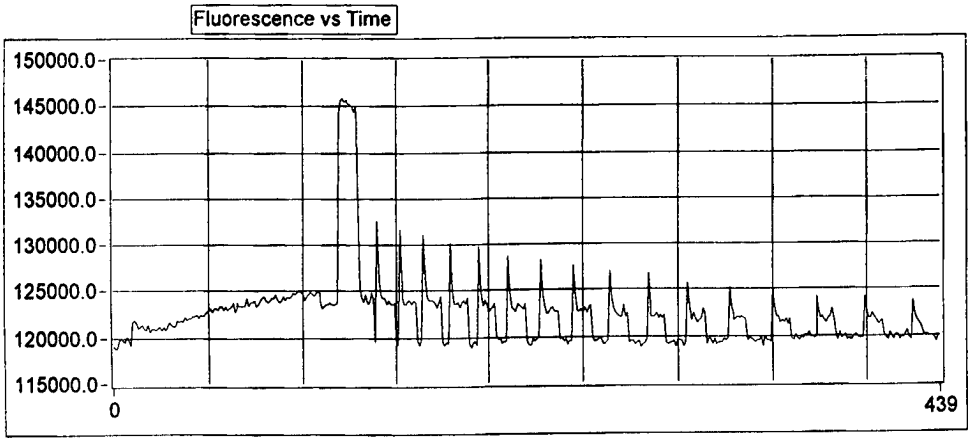


Figure 3b: Ion fluorescence "overshoot" signature for transition driven in the remote resonance trap and detected in the state-selection trap.

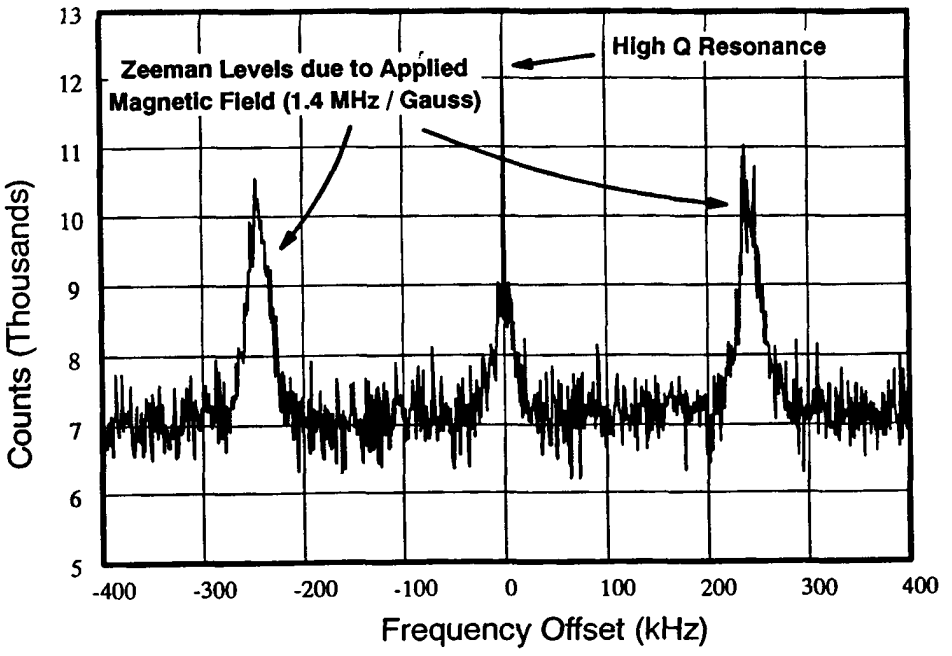


Figure 4 shows the atomic resonance driven in the resonance trap and detected after the ions had been transported back into the loading/state-selection trap. To obtain this resonance curve, the "overshoot" fluorescence as seen in Fig. 3b is collected in a 2 second time interval which starts with the arrival of the ions back into the state-selection region. This fluorescence is measured as the 40.5 Ghz source is swept in frequency by ±400 khz around the F=0, m_F=0 to F=1, m_F=0 clock transition. The two field sensitive transitions F=0 to F=1, m_F=±1 are spaced about 240 khz from the clock transition and correspond to a 170 mG magnetic field in the resonance trap. No attempt has been made to shim the field in the resonance region. These resonances are driven with high power microwaves derived from a Gunn oscillator locked to an H-maser reference source.

Conclusions

An improved architecture for a linear ion trap based frequency standard was built and ions have been trapped and moved back and forth between the state-selection and resonance trap with no loss of ions and with no relaxation of the atomic state preparation. Microwave resonances among the hyperfine levels have been driven in the remote interrogation region and detected in the optical state-selection region. Many of the design constraints of the present configuration are eliminated and a smaller, cheaper more stable frequency standard should result.

References

[1] J. D. Prestage, G. J. Dick, and L. Maleki, *New Ion Trap for Frequency Standard Applications*, J. Appl. Phys., vol. 66, pp. 1013-1017, 1989.

[2] J. D. Prestage, R. L. Tjoelker, G. J. Dick and L. Maleki, *Progress Report on the Improved Linear Ion Trap Physics Package*, Proc. 1994 IEEE Frequency Control Symposium, pp. 755-760, 1994.

[3] J. D. Prestage, R. L. Tjoelker, G. J. Dick and L. Maleki, *Improved Linear Ion Trap Physics Package*, Proc. 1993 IEEE Frequency Control Symposium, pp. 144-147, 1994.

[4] R. L. Tjoelker, J. D. Prestage, G. J. Dick, and L. Maleki, *Recent Stability Comparisons with the JPL Linear Trapped Ion Frequency Standards*, Proc. 1994 IEEE Inter. Freq. Control Symposium, pp. 739-743, 1994.

[5] L. S. Cutler, R. P. Giffard, and M. D. McGuire, *Thermalization of ^{199}Hg Ion Macromotion by a Light Background Gas in an rf Quadrupole Trap*, Appl. Phys. B36, pp. 137-142,1985.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

EIGHT YEARS OF EXPERIENCE WITH MERCURY STORED ION DEVICES

D.N. Matsakis, A.J. Kubik, and J.A. DeYoung
United States Naval Observatory
Washington, DC 20392-5420

and

R.P. Giffard and L.S. Cutler
Hewlett-Packard Laboratories
Hewlett-Packard Physics Laboratory
Palo Alto, CA 94304

1. Abstract

The performance of three similar trapped-ion frequency standards, which use mercury in a hyperboloidal Paul trap, is presented. Initially, frequency deviations well below 10^{-14} rms were observed on time scales of 1-10 days, but the standards later underwent frequency variations of up to 10^{-12} s/s, which are probably due to impurity effects.

2. Introduction

The U.S. Naval Observatory (USNO) has maintained three frequency standards, whose fundamental measurements are based on microwave sampling of mercury ions confined within a hyperboloidal Paul trap. Since their installation in 1986 and 1987,[1] they have been operated under a policy of minimal disturbance. None of the vacuum systems have ever been exposed to air, and humans would enter their environmentally controlled rooms only to calibrate or repair the standards or other equipment in the room. The frequencies of the three units (Figure 1) have displayed similar long-term patterns of staying roughly constant for a few years, with a daily rms below 10^{-14} s/s, then falling in frequency by up to 10^{-12} s/s over the next several years. During 1993, the getter-based vacuum pump in unit 1 was periodically reactivated, resulting in a temporary restoration of the frequency to its former level.

3. Instrumental Description

The frequency standards are based upon optical pumping of the 40.5 GHz hyperfine transition of Hg^{199} ions (Figure 2) confined within a sphere of radius about 2.5 mm by the time-varying electric fields of a hyperboloidal potential[2,3] (Figure 3). Within the four stages of the 2.5-second measurement cycle, ions are sequentially:

1. Created by bombardment with an ionizing electron current.

2. Pumped with ultraviolet (uv) light out of the F=1 S state to the P states, from which they quickly cascade to both S states, eventually depleting the population of the F=1 S state and maximizing the population of the F=0 S state.

3. Interrogated with a 40.5 GHz microwave field, which has the effect of removing ions from the F=0 S state and repopulating the F=1 S state. The uv lamp is set to a dim state so as to minimize the "light shift."[4]

4. Re-exposed to uv light, which is absorbed by any ions in the F=1 S state and

0-7803-2500-1/95 $4.00 © 1995 IEEE

reemitted. Some of the reemitted photons are detected with a photomultiplier tube (pmt), as are some photons which were reflected into the pmt without being absorbed by the mercury.

A frequency measurement is achieved by counting the number of uv photons reemitted by the mercury. A high uv count indicates that the 40.5 GHz microwave frequency was close enough to the resonance line center to stimulate transitions from the F=0 state back to the F=1. The 40.5 GHz microwave field is generated from and phase-locked to a 5 MHz signal from a maser, which itself can be directly referenced to the USNO Master Clock. The center of the resonance line is found by determining what frequency offset equalizes the uv counts when the microwave field is offset in frequency by a half-linewidth above and below the 0.85-Hz-wide resonance. In order to eliminate the effects of linear drifts, frequency offsets are alternated on each side of the line, and each measurement is compared to the average of the two measurements on the opposite side of the resonance which immediately precede and follow it.

The number of ions and their macromotion frequency (defined below) are kept constant by periodically shocking the cloud with a 50-kHz high-voltage impulse and observing the resulting image currents in the trap circuit. The strength of electron current used to create ions is varied in each cycle so as to keep the amplitude of the image currents constant, and the strength of the rf confining voltages are set to keep the frequency of the response current at 50 kHz.

The permanent data record consists of 21-minute averages of the frequency offsets, pressure, vacuum pump currents, several temperatures, heater status, trap confinement parameters, cloud creation parameters, relative humidity, frequency variances, and average uv count. Figures 4-12 plot some of these as a function of modified Julian day (MJD). For the last few years, these data have been supplemented by daily observations of the Zeeman components (Table 1), which give an indication of the magnetic field experienced within the trap shields.

The trap vacuum is controlled by two vacuum pumps and a helium-leak input. In order to cool the mercury ions and minimize their Doppler shifts, helium is allowed to leak into the vacuum through a pyrex tube, whose temperature is varied in order to modify the diffusion rate so as to keep the vacuum pressure (as measured by a Granville-Phillips Ionization Gage) constant. The vacuum is maintained by a zirconium-aluminum getter pump and an ion pump which ionizes the gas and then buries those ions under sputtered titanium.

The three units are not identical. Units 1 and 2 are similar, but data from unit 2 (which was disassembled in February, 1993) are of lesser quality because the endcap shields of its hyperboloidal trap were grounded. Unit 3 has several improved electronic circuits, one of which allows software to independently vary the dc potential of the two endcaps, whereas for the other units the dc potential of the two endcaps is the same.

The temperature at the trap itself is measured with a thermocouple. The temperature is controlled at the magnetic shields, the neck below the trap, the uv lamp, the HgO sublimation bottle, and the helium-leak bottle.

4. Trap Dynamics

If it is in the presence of an oscillating quadrupole electric field, an ion will undergo an oscillatory micromotion as it is sinusoidally accelerated. The instantaneous electric potential experienced by the ion is

given in cylindrical coordinates by

$$\phi(t) = \frac{[U - Vcos(\Omega t)](\rho^2 - 2z^2)}{\xi^2} \quad (1)$$

where U and V are the dc and rf voltages applied, Ω is the rf frequency (555.555 kHz), and ξ is a length characterizing the trap electrodes. It is well known that the net effect of the oscillating part of the field is to create a centrally directed average restoring force (the pondermotive force) on an ion, provided the displacement during a trap cycle is small compared to the distance from the trap center.[5,6] The oscillatory part of the potential can be replaced by a harmonic pseudopotential

$$\phi_{macro} = \frac{q^2V^2(\rho^2 + 4z^2)}{m\Omega^2\xi^4} \quad (2)$$

where q and m are the charge and mass of the trapped ion.

For timescales larger than $2\pi/\Omega$, one can therefore rewrite the effective potential as:

$$\phi_{macro} = \frac{1}{2}m(\omega_\rho^2\rho^2 + \omega_z^2 z^2) \quad (3)$$

where

$$\omega_\rho^2 = \frac{2q^2V^2}{m^2\Omega^2\xi^4} + \frac{2qU}{m\xi^2} \quad (4)$$

$$\omega_z^2 = \frac{8q^2V^2}{m^2\Omega^2\xi^4} - \frac{4qU}{m\xi^2} \quad (5)$$

It is standard practice to set the dc voltage so that

$$U = \frac{qV^2}{m\Omega^2\xi^2} \quad (6)$$

and then $\omega_z = \omega_\rho = \omega = \frac{2qV}{m\Omega\xi^2}$
and in spherical coordinates:

$$\phi_{macro}(r) = \frac{1}{2}m\omega^2r^2 \quad (7)$$

The complete potential Φ is given by the sum of the ϕ_{macro} and $q\phi_{sc}$, where ϕ_{sc} is the potential due to the space charges of the ion cloud.

If the ion density is assumed to follow a Boltzmann distribution, then the density n(r) is given by

$$n(r) = n(0)e^{-E/kT} \quad (8)$$

where

$$E = \frac{1}{2}mv^2 + \Phi \quad (9)$$

Integrating out the statistically independent v-dependence of n(r), and using Poisson's equation ($\nabla^2\phi_{sc} = -qn/\epsilon_0$), it is possible to derive the following nonlinear second-order equation for n(r):

$$n'' - \frac{n'^2}{n} + \frac{2n'}{r} - \frac{n^2q^2}{\epsilon_0 kT} + \frac{3m\omega^2 n}{kT} = 0 \quad (10)$$

Numerical solutions show that the ions tend to accumulate in the trap center so as to cancel the pondermotive force.[7-9] Due to collisions with the gas and reflections off the cloud boundaries, each ion can be expected to sample all possible velocities and positions in the course of absorbing a microwave photon. The first-order Doppler effect causes a modulation of the microwave frequency in the frame of the ion, and experiments have shown that the resulting sidebands are displaced from the resonance by about 20 kHz.[10]

The observed line frequency is decreased by the average of all possible second-order Doppler shifts throughout the cloud, whose total magnitude is of order 5×10^{-12}. The second-order Doppler shift depends upon the average instantaneous velocity squared, which depends upon the density and temperature of the cooling helium buffer gas (maintained at a measured pressure of 1.5×10^{-6}

torr), the strength of the trap confining fields, and the ion cloud's size and shape. At any given point, however, the contribution of the micromotion to the second-order Doppler shift is given by

$$\frac{-<v^2>}{2c^2} = \frac{-\omega^2(\rho^2+4z^2)}{4c^2} \qquad (11)$$

It follows that larger clouds will have smaller frequencies.[7]

The existence of contaminants can affect frequency measurements in several ways. The frequency of the mercury ions could be changed by collisions with any background gas.[11] Another effect of any non-helium vacuum component would be to change the proportionality between the true vacuum pressure and the current measured with the vacuum ionization gage, thus leading to a variation in the true density of the helium cooling gas and the final temperature of the mercury cloud.

An ionized non-mercury component could be trapped with the mercury, if its mass is larger than about 50 amu. Any such species with a charge-to-mass ratio larger then Hg^+ would experience a larger pondermotive force than the Hg^+ and would preferentially occupy the center of the cloud, forcing the Hg^+ to the outer regions, where the micromotion velocities are larger.[12] This would have the effect of increasing the average second-order Doppler shift, and decreasing the frequency. Further distortions would be caused by changes in the response to the high-voltage impulse; the exact amount depends among other things upon how close the macromotion frequency of the contaminant is to a resonance of the 50 kHz setpoint frequency.

Since the ratio of the dc to rf voltages which causes the spherical symmetry of the charge distribution depends upon the charge-to-mass ratio, contamination will result in non-spherical distributions and can also lead to collective oscillations.[13]

Numerical estimates can be obtained by noting that according to equations (4) and (5) the resonant macromotion frequencies of a contaminant of mass m_c, $\omega^2_{\rho,c}$ and $\omega^2_{z,c}$, will be equal neither to each other nor to ω. If the contaminant is singly and positively ionized, the condition of Boltzmann equilibrium allows one to compute the ratio of contaminant density n_c to mercury density:

$$\frac{n_c(r,z)}{n(r,z)} = \frac{n_c(0,0)}{n(0,0)} \times e^{\frac{-((m_c\omega^2_{\rho,c}-m\omega^2_\rho)\rho^2+(m_c\omega^2_{z,c}-m\omega^2_z)z^2)}{2kT}} \qquad (12)$$

The existence of vacuum impurities can be indirectly inferred from some of the recorded instrumental parameters. The most straightforward is from the current of ion pump, which, for each species, is proportional to the number of atoms ionized by the pump in the process of removal. Another measure of vacuum purity is the strength of the trapping fields required to maintain the magnitude and frequency of the ion image currents. All of these are subject to various long-term aging effects.

The large frequency variations in unit 1 after MJD 48986 were initiated after the getter pump was heated in an attempt to refurbish it by outgassing hydrogen and driving all other absorbed impurities inward. On MJD 48946 the getter pump heater, which had been heating the getter to a temperature 320°C, was turned off. On MJD 49076, 49167, and 49258 the heater was turned up so as to reach 350°C for several days before being again turned off. It can be seen that this had the effect of restoring the frequency towards its original value, but that the improvement was rapidly lost. As shown in Figure 5, as the frequency falls so does the

temperature of the helium leak, which indicates that less helium must be introduced to maintain the measured pressure. This could be due to an increase in the concentration of just one impurity, although more complex possibilities involving the differential sensitivity to different contaminants cannot be ruled out. In the middle plot, the decreasing magnitude of the current of the ion pump indicates that either fewer atoms or a different mixture of atoms are being removed.

The correlation between frequency variation and ion pump current is also evident in plots of the entire edited dataset (Figures 5, 8, and 11) and also in some of the data which were removed by the editing process. These edited data were often associated with the failure of the temperature stabilization circuits, which aside from affecting the ion cloud temperature directly, could cause extra contaminants to be emitted by the walls or alter the properties of the pump system.

Further indirect evidence for contamination can be found in the dependence of frequency upon the helium pressure. Cutler et al.[7] have found that at pressures below 10^{-6} torr the frequency rapidly decreases with decreasing pressure, due to the reduced cooling. At higher pressures, the mercury thermalization becomes complete, but the frequency continues to increase at a slower rate expected from a collisional shift.[14] Figure 13 shows data taken near MJD 48100, in which the pressure was varied superimposed upon the curve of Cutler et al. The rise of frequency with pressure below 10^{-6} torr is shown to be less steep than expected. This could be due to the presence of a more effective cooling agent in the cloud than helium or, alternately, due to the release of contaminants by temperature variations associated with the helium leak pressure control system. The daily average indicated vacuum pressure is usually kept constant to within 10^{-10} torr; which implies, in the absence of changes in the composition mixing ratios, insignificant frequency variations due to the pressure shift. The actual temperature at the helium leak rises slowly with time so as to keep a constant flow rate while the reservoir empties, then suddenly falls after the reservoir is refilled (every three or four years).

Although we have no way of directly inferring what kind of contaminants could be involved, barium (137 amu) is a prime candidate since it is known to be emitted in the electron gun, and others may be present in the mercuric oxide which is sublimated to produce neutral mercury for ionization.

5. The Magnetic Field

Each clock has three layers of mu-metal magnetic shielding, so that the magnetic field is principally determined entirely by the current through an enclosed coil, which we typically set to produce fields of order 1 milligauss. If the field varies uniformly, then it will shift the frequency at a rate[15] of

$$2.4B^2 fs/s/mG^2 \qquad (13)$$

It will also cause the Zeeman components to shift at a rate of 1.4 kHz/milligauss. If the magnetic field were exactly zero, or such as to cause a frequency overlap between a Zeeman component and a frequency sideband, collisions would shift ions between magnetic substates, resulting in coherence time degradation and a decrease the signal to noise ratio.

Since 1991, semi-weekly measurements of the Zeeman components have been made for units 1 and 2. Unit 2 showed broader Zeeman components, as well as variations in the lineshapes. Table 1 presents the measured Zeeman splitting, derived magnetic fields, range of observed magnetic variation, and

theoretical effect on the frequency measurement. The variations are due to any irregularities in the coil current, as well as any magnetic field that may have leaked through the magnetic shields.

Figure 14 shows how the lineshape of unit 3 was affected after MJD 48852, when the dc fields of the trap (whose potential is U above) were altered. Although the exterior controls were set so as to restore U to its nominal value, it is evident that the coherence properties were affected, perhaps through residual magnetization.

6. Effect of Lamp Brightness

Another source of measurement error is the purely random noise caused by uv photons which arrived at the detector by way of reflections instead of absorption through the mercury resonance. After averaging for 21 minutes, the true signal of absorbed photons is typically less than 3% of the noise of reflected photons, and this noise source would be expected to decrease with the familiar inverse square root of the averaging time. In Figures 15-17, the lamp brightness for each unit is given in the top plot and compared with the rms scatter of the frequency measurements (averaged each 10 days, each day, and each 21 minutes), as well as the average frequency compared to the maser clock. The discontinuities in the lamp brightness plot correspond to times when the lamps were replaced or adjusted. It is to be noted that while increasing lamp brightness at times decreased the scatter on scales of 21 minutes, other noise sources are dominant on longer time scales. It is interesting that for many periods there seems to be an anti-correlation between the average frequency and the short-term rms; both of these are probably caused by vacuum contamination.

7. Absolute Accuracy in the Absence of Contamination

Inspection of Figure 1 shows that the very best data were obtained before MJD 47400 and for units 1 and 3. The last figure shows the relevant data, whose statistical parameters are summarized in Table 2. If the two standards and the TAI were truly independent, it would be possible to derive the variances of all three time series using the standard "3-cornered-hat" method. This is not the case, due to continental time-transfer errors. The data do suggest, however, that when the mercury clocks were giving their optimal performance, their absolute accuracies on intervals of 1-10 days were a few parts in 10^{-15}. Efforts to improve the vacuum in units 1 and 3 are currently underway.

8. Acknowledgments

We thank L.A. Breakiron, C.A. Flory, R.T. Clarke, W.J. Klepczynski, L. Maleki, J.D. Prestage, M. Stone, R.L. Tjoelker, P.J. Wheeler, and G.M.R. Winkler.

References

1. L.S. Cutler, R.P. Giffard, P.J. Wheeler, and G.M.R. Winkler, 1987 *Proc. 41ST Ann. Freq. Control Symp,* p. 12.
2. F.G. Major and G. Werth, 1973 *Phys. Rev. Lett.* 30, 1155.
3. L.S. Cutler, R. P. Giffard, and M.D. McGuire, 1983 *Proc. 37TH Ann. Freq. Control Symp.*, p. 32.
4. M.D. McGuire, 1977 *Proc. 31ST Ann. Freq. Control Symp.*, p. 612.
5. H.G. Dehmelt, 1967 *Advan. Atom. Mol. Phys.* 3, 53.
6. H.G. Dehmelt, 1969 *Advan. Atom. Mol. Phys.* 5, 109.
7. L.S. Cutler, R. P. Giffard, and M.D. McGuire, 1985 *Appl. Phys. B* 36, 137.

8. C. Meis, M. Desaintfuscien, and M. Jardino, 1988 *Appl. Phys. B* 45, 59.
9. J.D. Prestage, G.J. Dick, and L. Maleki, 1989 *J. Appl. Phys* 66 (3), 1013.
10. L.S. Cutler, C.A. Flory, R.P. Giffard, and M.D. McGuire, 1986 *Appl. Phys. B* 39, 251.
11. S. Dushman and J.F. Lafferty, 1962 Scientific Foundations of Vacuum Technique, 2nd edn. (Wiley, NY), p. 324.
12. C. Flory, 1995, private communication
13. K. Jungmann, J. Hoffnagle, R.G. DeVoe, and R.G. Brewer, 1987 *Phys. Rev. A* 36, No. 7, 3451.
14. J. Vetter, M. Stuke, and E.W. Weber, 1975 *Z. Physik A* 273, 129.
15. M.D. McGuire, R. Petsch, and G. Werth, 1978 *Phys. Rev. A* 17, No. 6, 1999.

Table 1: Magnetic Field Properties.

The peak line strength of the outer Zeeman components were typically 0.7 the strength of the main component for unit 1 and 0.5 the main component in unit 2. Unit 3 displays no Zeeman components within 10 kHz of line center, perhaps due to the alignment of the magnetic component of its rf field. Columns (3) and (4) give the frequency difference between the main and satellite hyperfine components. The last two columns indicate the associated frequency shift of the central component. The nominal magnetic field, determined from the current in a Helmholtz coil was 1 mG for units 1 and 2. The nominal field for unit 3 was initially 4 mG, but changed to 1 mG on mjd 48286, and again varied between 4 mG and 40 mG after the event of MJD 48852, when the dc fields in the trap were temporarily changed.

		Separation Hz	Separation milligauss	Δf_{shift} Hz	Δf_{shift} fs/s
Unit 1	Minimum	1340	.96	8.9×10^{-5}	2.2
	Maximum	1400	1.0	9.7×10^{-5}	2.4
	Max-Min	60	.04	$8. \times 10^{-6}$	0.2
Unit 2	Minimum	3299	2.36	5.40×10^{-4}	13.3
	Maximum	3276	2.34	5.31×10^{-4}	13.1
	Max-Min	23	.02	$9. \times 10^{-6}$	0.2

Table 2: Statistical Properties of Data From MJD 47132 to 47400.

Daily TAI frequencies were derived by differencing and interpolating the 10-day Circular T values published by the BIPM. All units are in fs/s ($10^{-15} s/s$)

	RMS 1-day clump	Allan variance 1-day clump	RMS 10-day clump	Allan variance 10-day clump
Unit1-TAI	22.2	11.5	18.2	15.5
Unit3-TAI	17.4	7.7	15.1	16.8
Unit1-Unit3	14.4	12.6	6.7	4.7

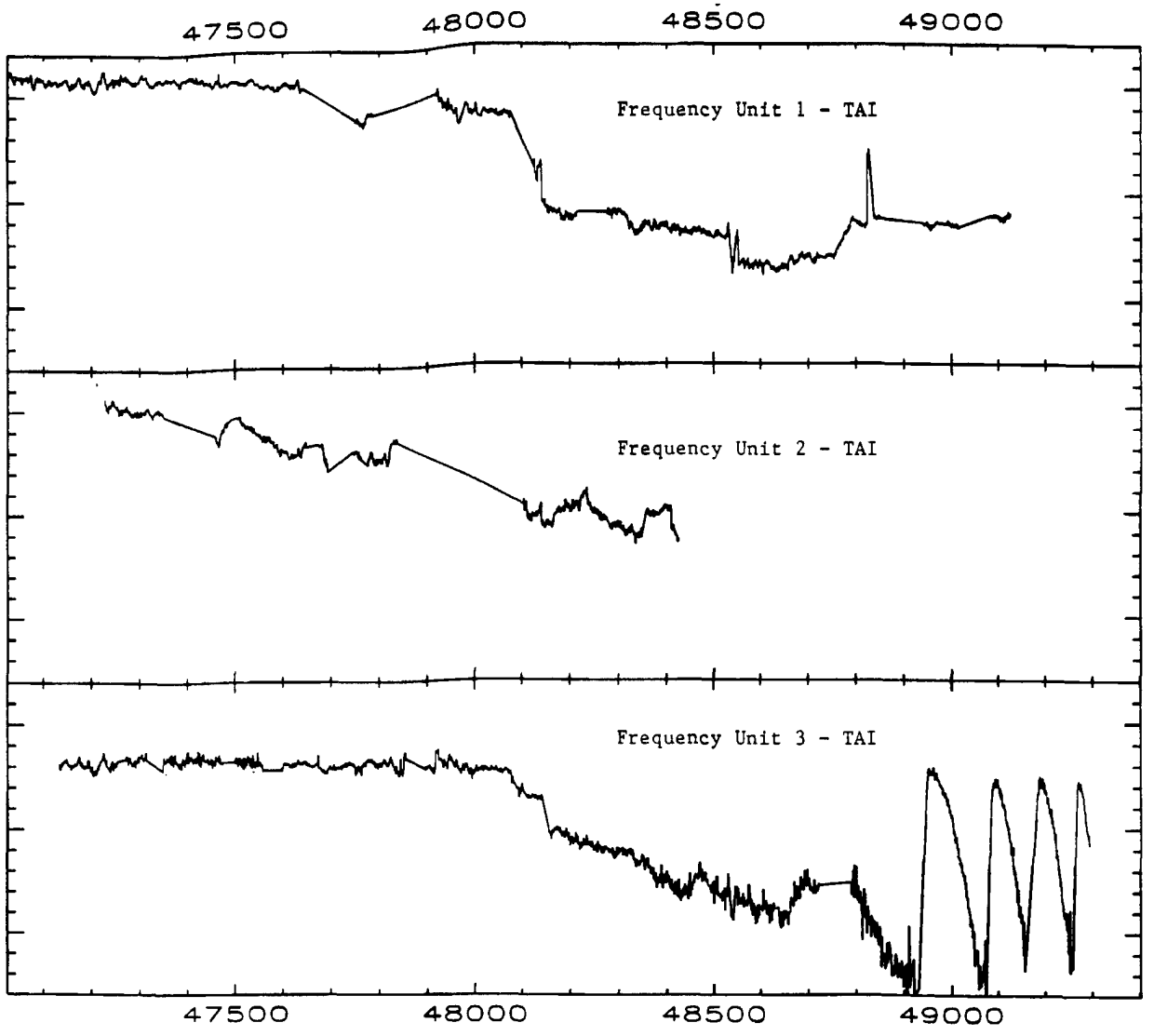


Figure 1: The daily frequency differences of mercury frequency standards and the TAI. To generate these plots, daily comparisons of the mercury clocks with their reference, which was the USNO Master Clock (MC2) until it was switched to the maser P19 on MJD 48732, were corrected using interpolated difference values obtained from the BIPM Circular T. Upper plot is unit 1, middle is unit 2, and lower is unit 3. All plots are 1500 fs/s full scale.

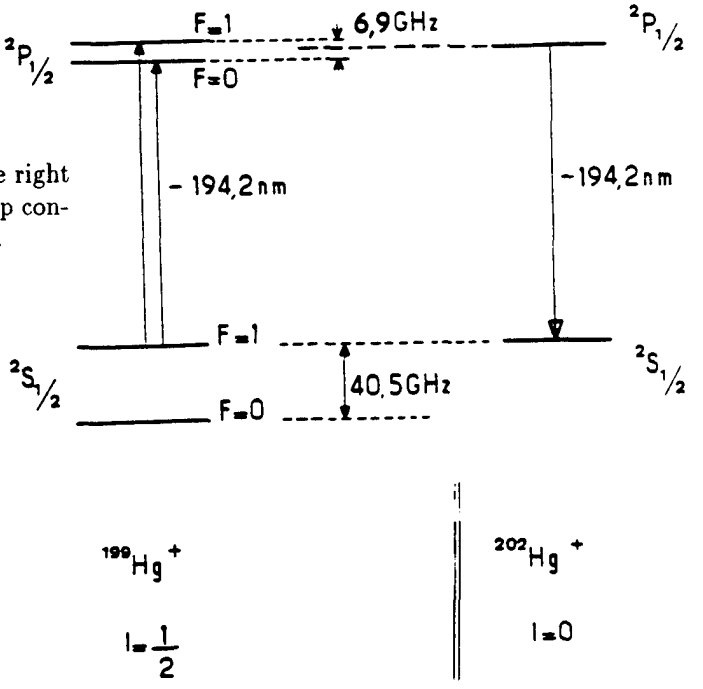


Figure 2: Hg^{199} energy level diagram. The right half shows how a uv emission line of a lamp containing Hg^{202} is used for optical pumping.

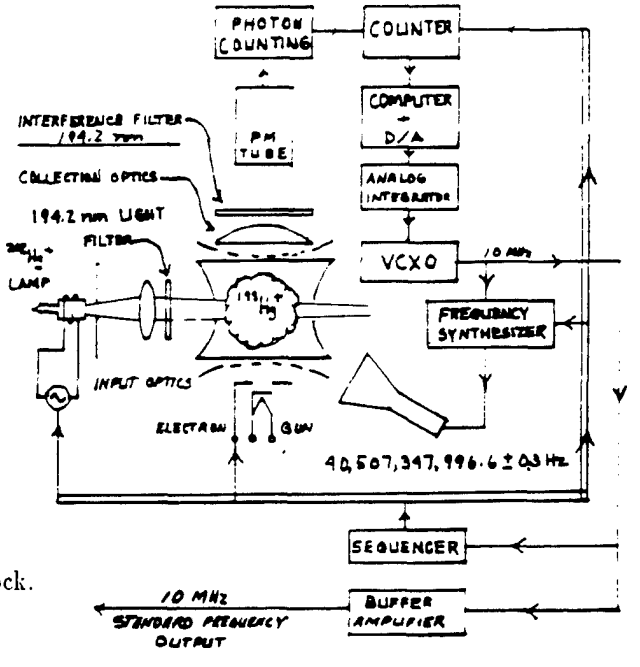


Figure 3: Hardware design of mercury clock.

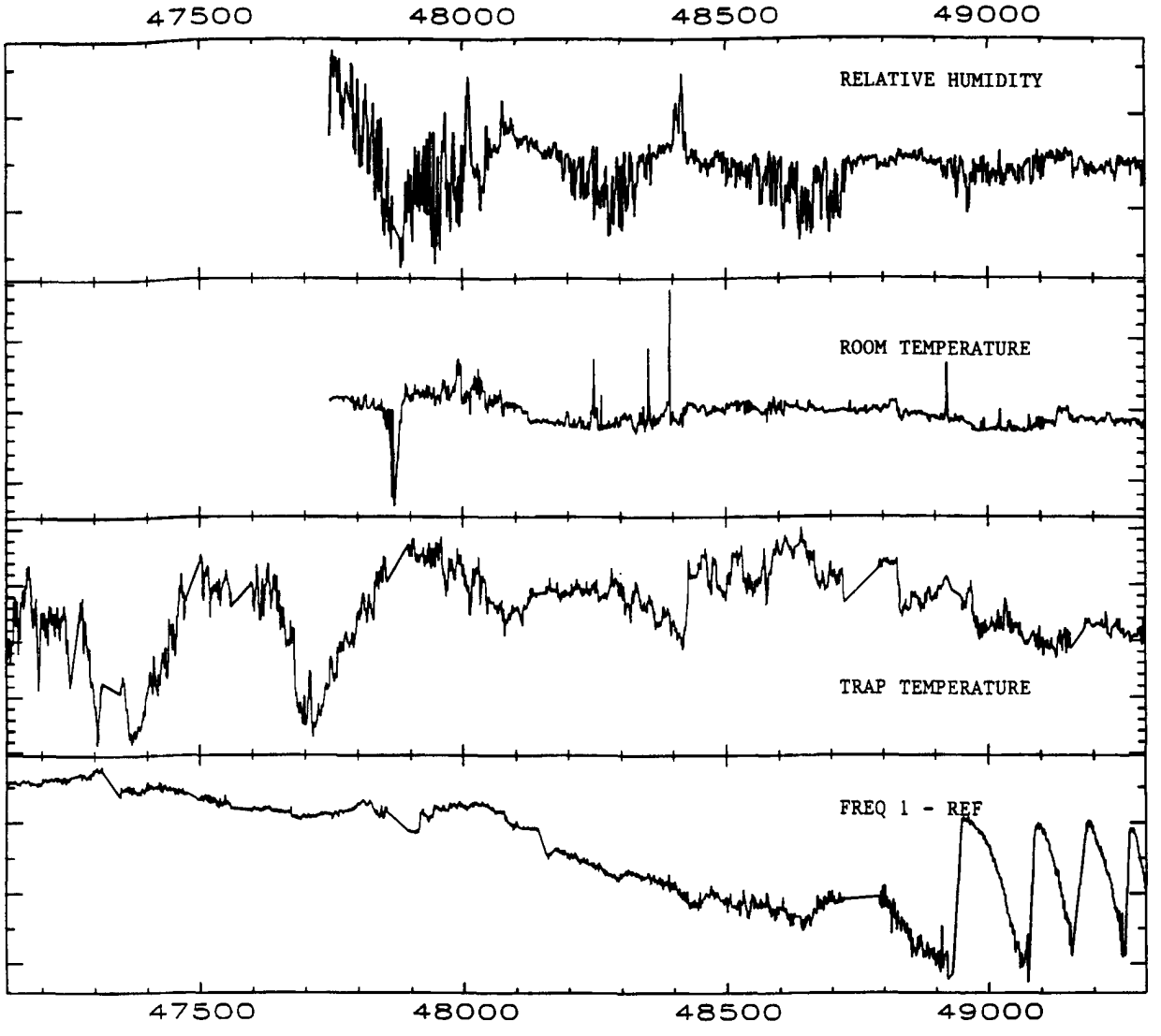


Figure 4: Autoscaled plots of unit 1 daily average clock parameters. From top to bottom, with values of minimum and maximum major tick marks in parentheses, plots are of room relative humidity (20%, 40%), room temperature (65 °F, 75 °F), trap temperature (46.6 °C, 47.0 °C), and output frequency minus reference (range 987 fs/s).

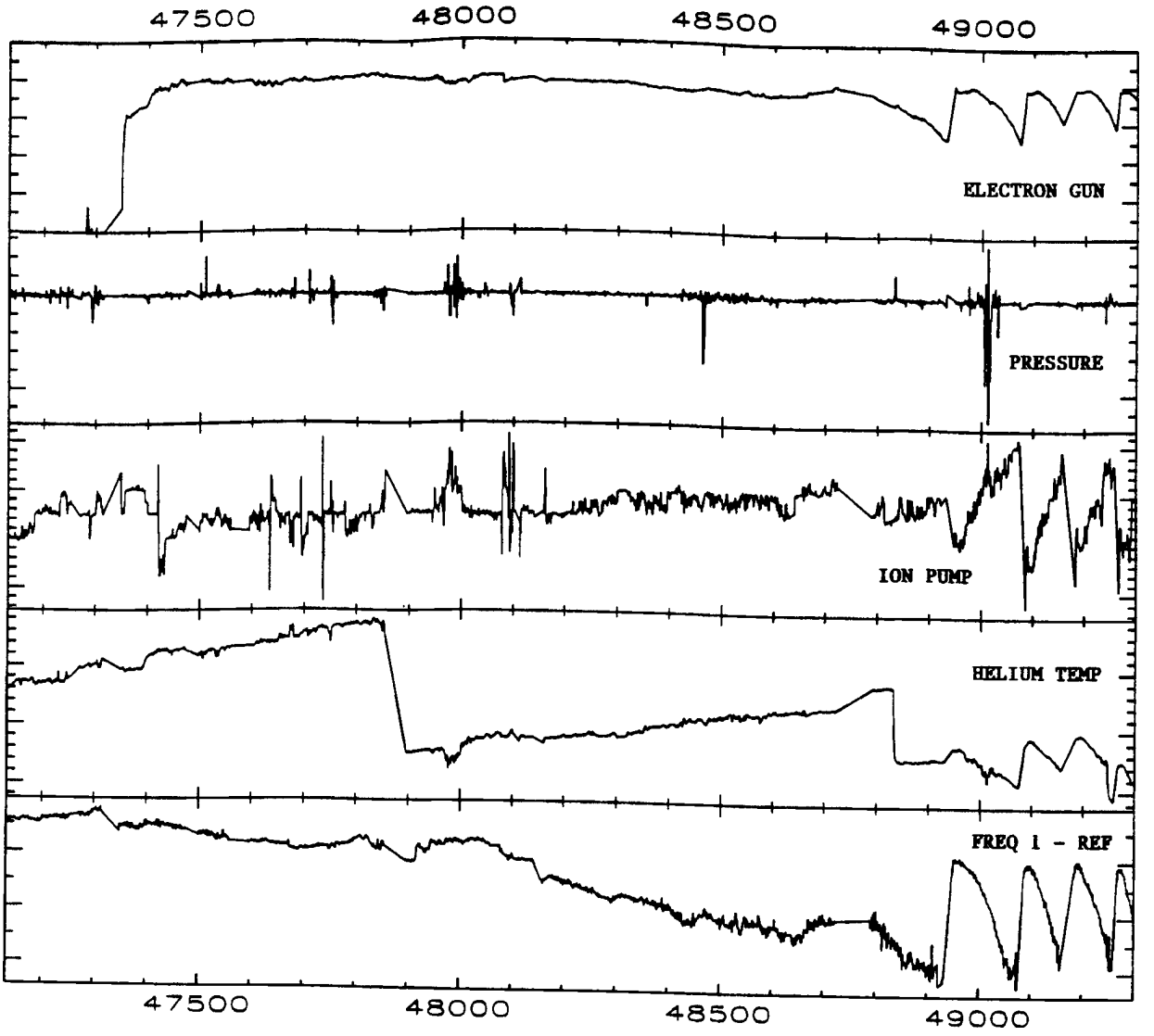


Figure 5: Autoscaled plots of unit 1 daily average clock parameters. From top to bottom, with values of minimum and maximum major tick marks in parentheses, plots are of electron gun voltage used to stabilize the measured ion number (-.8 v, .2 v), the vacuum pressure (1.495×10^{-6} torr, 1.500×10^{-6} torr; data previous to MJD 47314 decreased by 1×10^{-6} torr), ion pump current (-.01, -8×10^{-3}), temperature of helium leak (40°C, 60°C), and output frequency minus reference (range 987 fs/s).

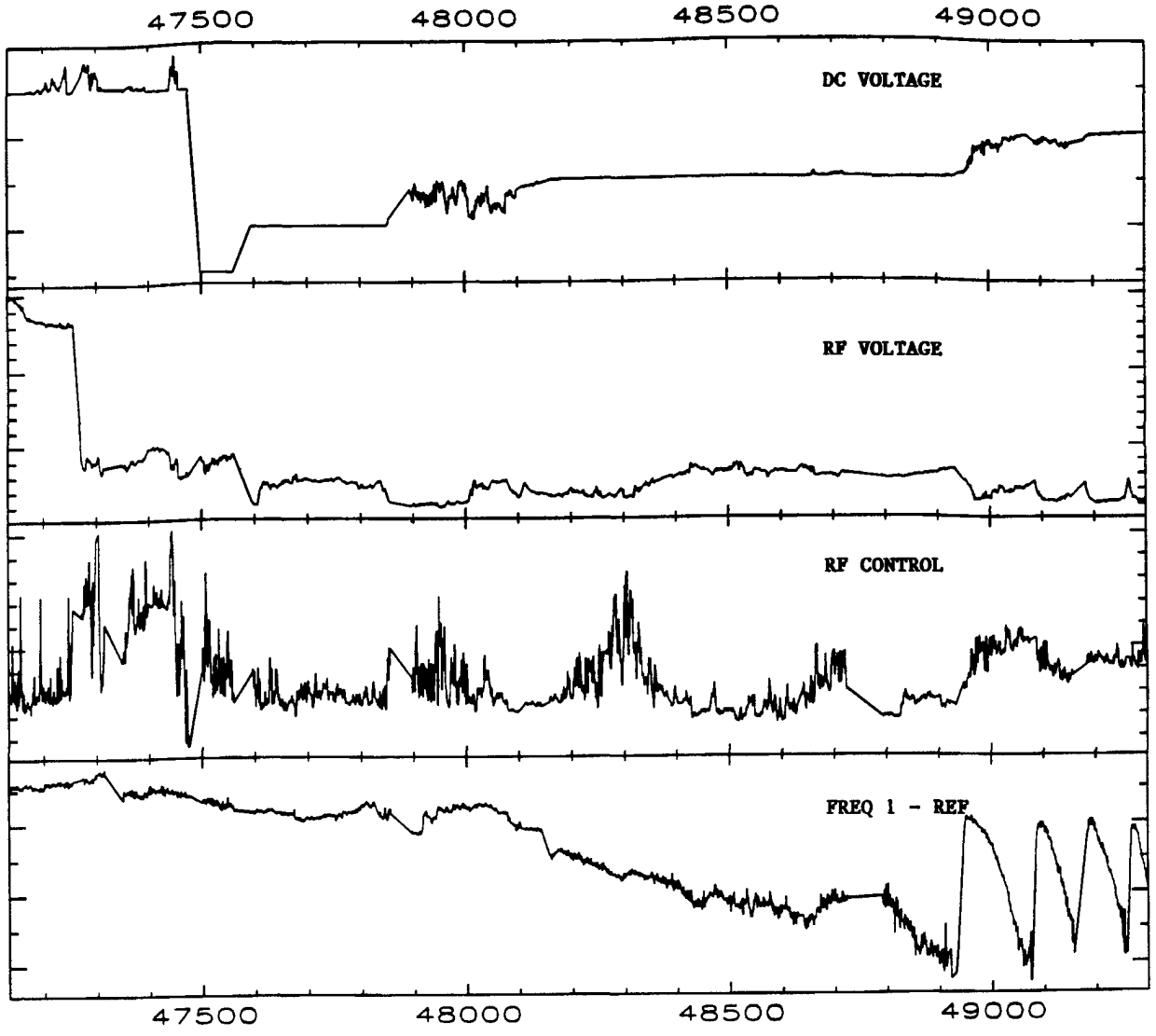


Figure 6: Autoscaled plots of unit 1 daily average clock parameters. From top to bottom: the dc voltage difference between trap endcaps (-15.65 v, 15.6 v), actual amplitude of trap rf (348 v, 350 v), control voltage used to set the rf voltage so as to stabilize macromotion frequency (5 v, 10 v), output frequency minus reference (range 987 fs/s).

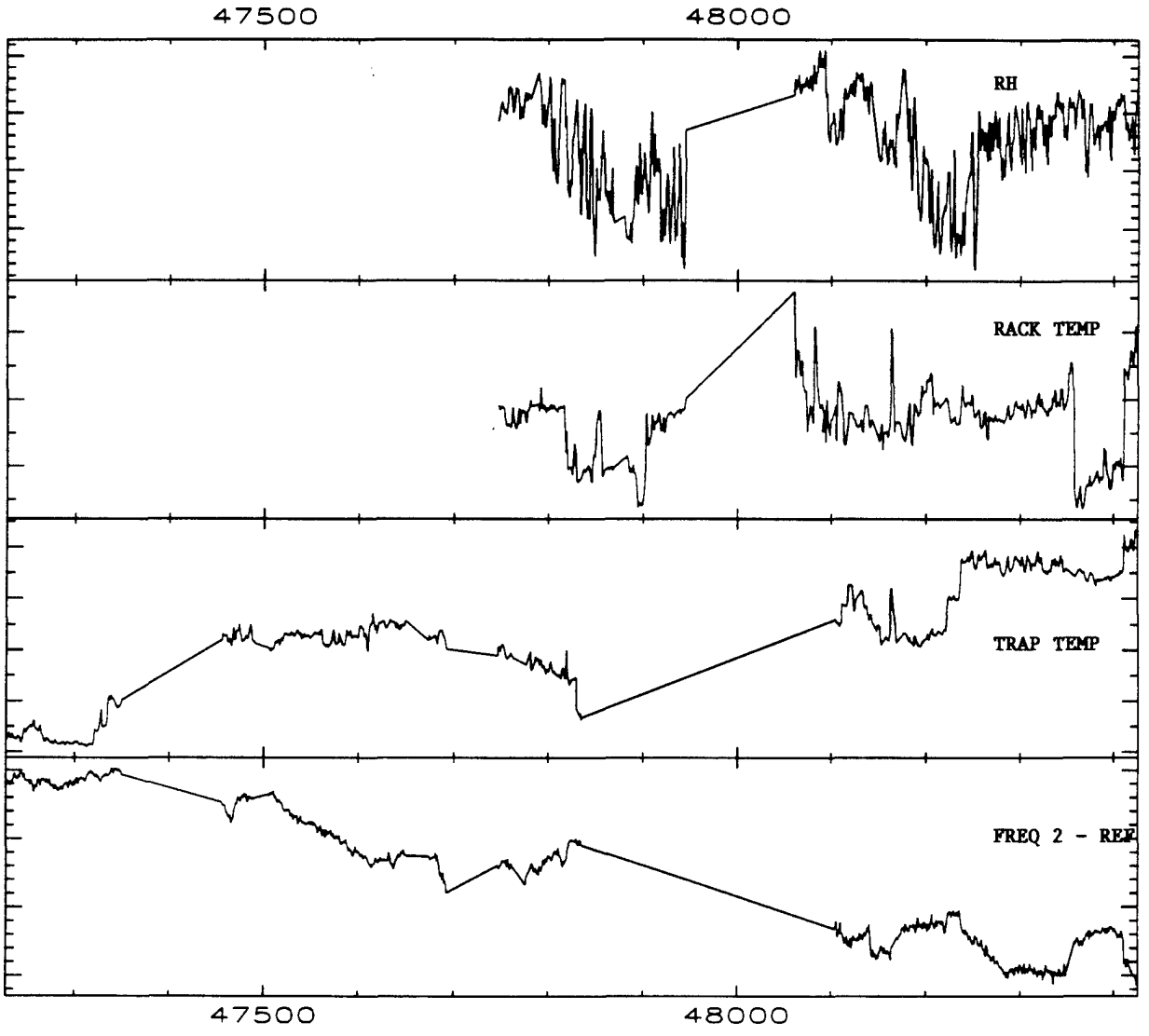


Figure 7: Autoscaled plots of unit 2 daily average clock parameters. From top to bottom, with values of minimum and maximum major tick marks in parentheses, plots are of room relative humidity (20%, 50%), rack temperature (82°F, 86°F), trap temperature (41.6°C, 42.4°C), and output frequency minus reference (range 740 fs/s).

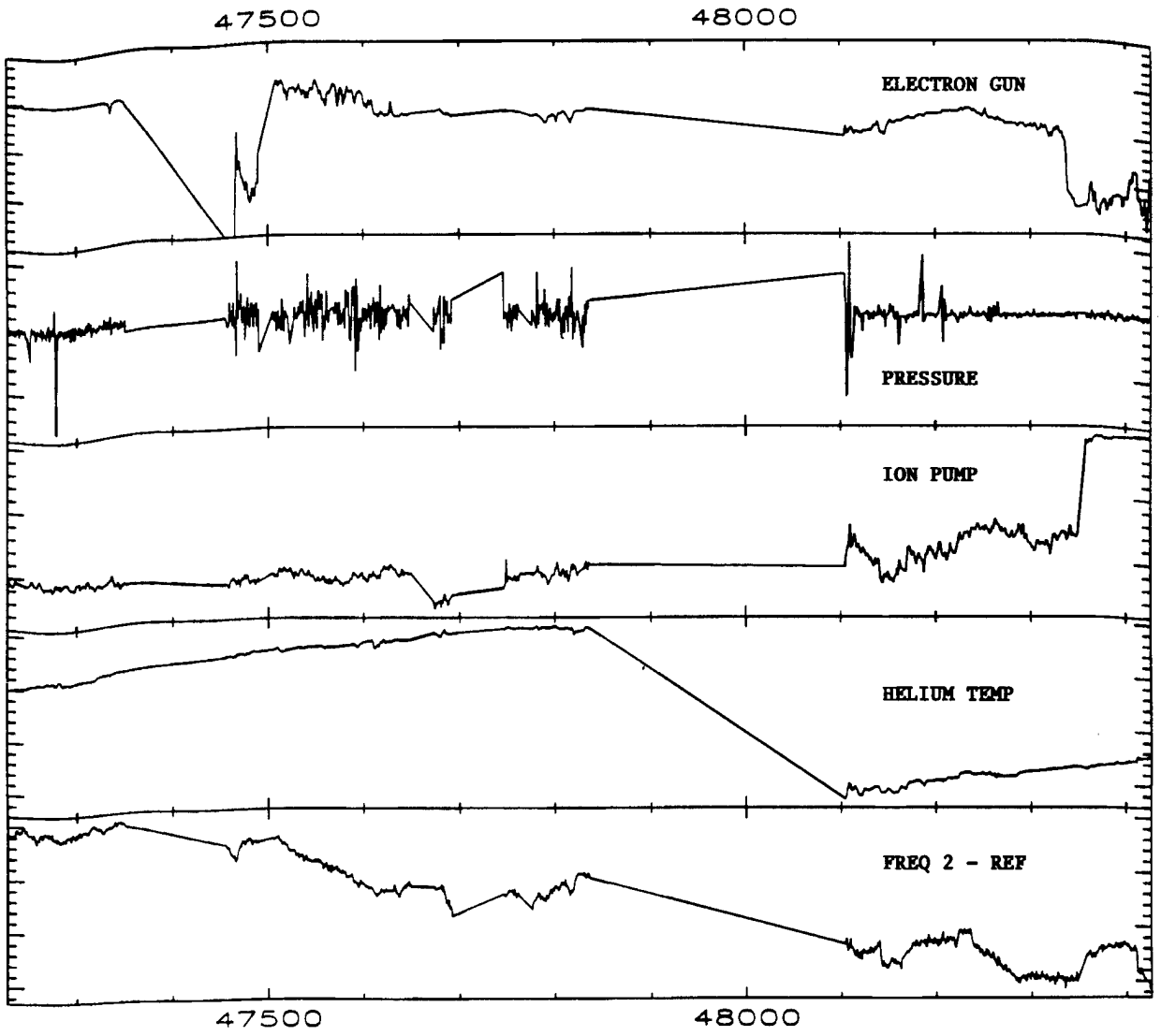


Figure 8: Autoscaled plots of unit 2 daily average clock parameters. From top to bottom, with values of minimum and maximum major tick marks in parentheses, plots are of electron gun voltage used to stabilize the measured ion number (-.4 v, -.2 v), the vacuum pressure (1.4995×10^{-6} torr, 1.5005×10^{-6} torr), ion pump current (-.002, 0), temperature of helium leak (45°C, 60°C), and output frequency minus reference (range 987 fs/s).

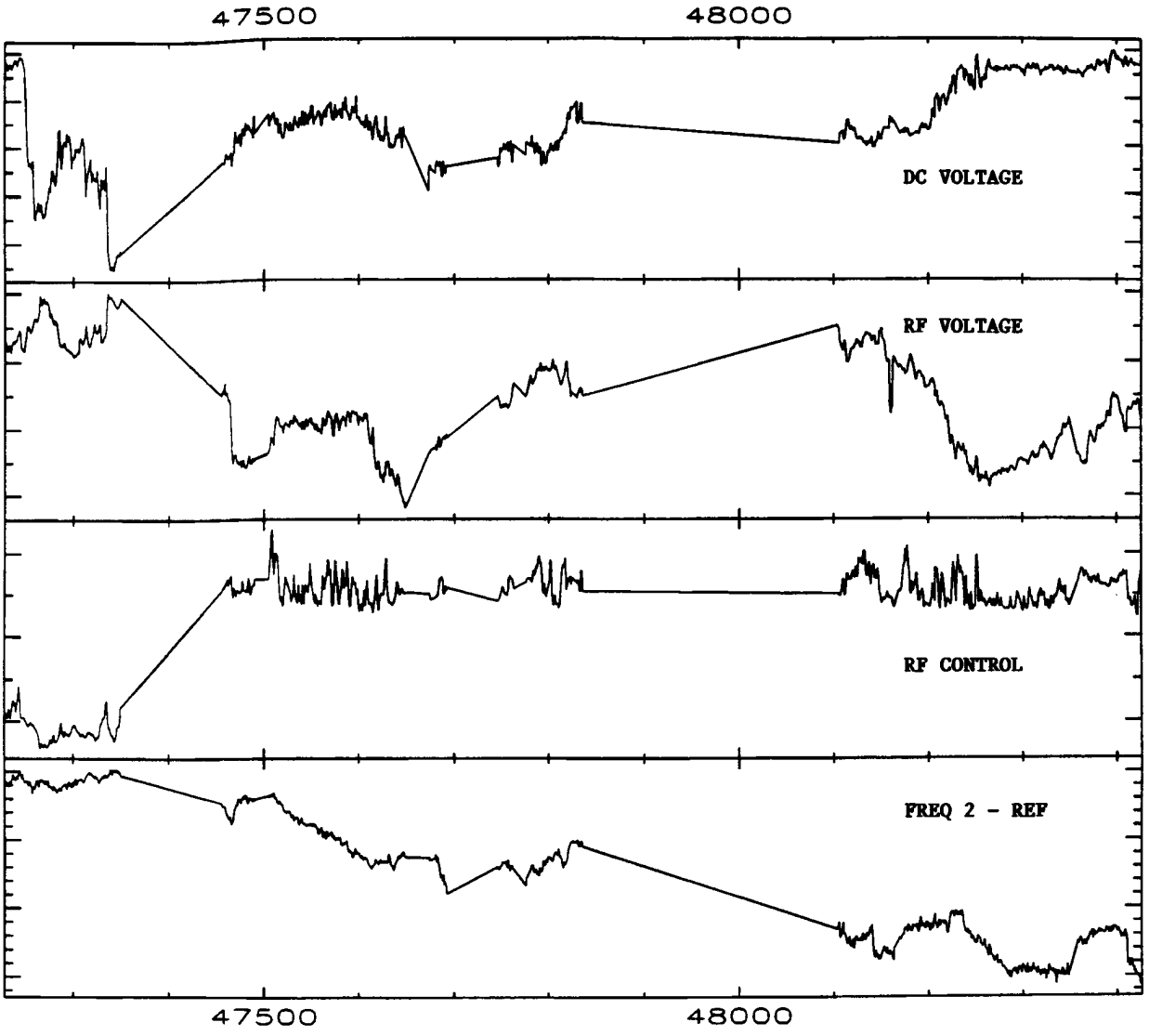


Figure 9: Autoscaled plots of unit 2 daily average clock parameters. From top to bottom, with values of minimum and maximum major tick marks in parentheses, the plots are of the dc voltage difference between trap endcaps (-14.989,-14.980), actual amplitude of trap rf (316.3 v,317.2 v), control rf voltage so as to stabilize macromotion frequency (2 v, 6 v), and output frequency minus reference in fs/s (740 fs/s range).

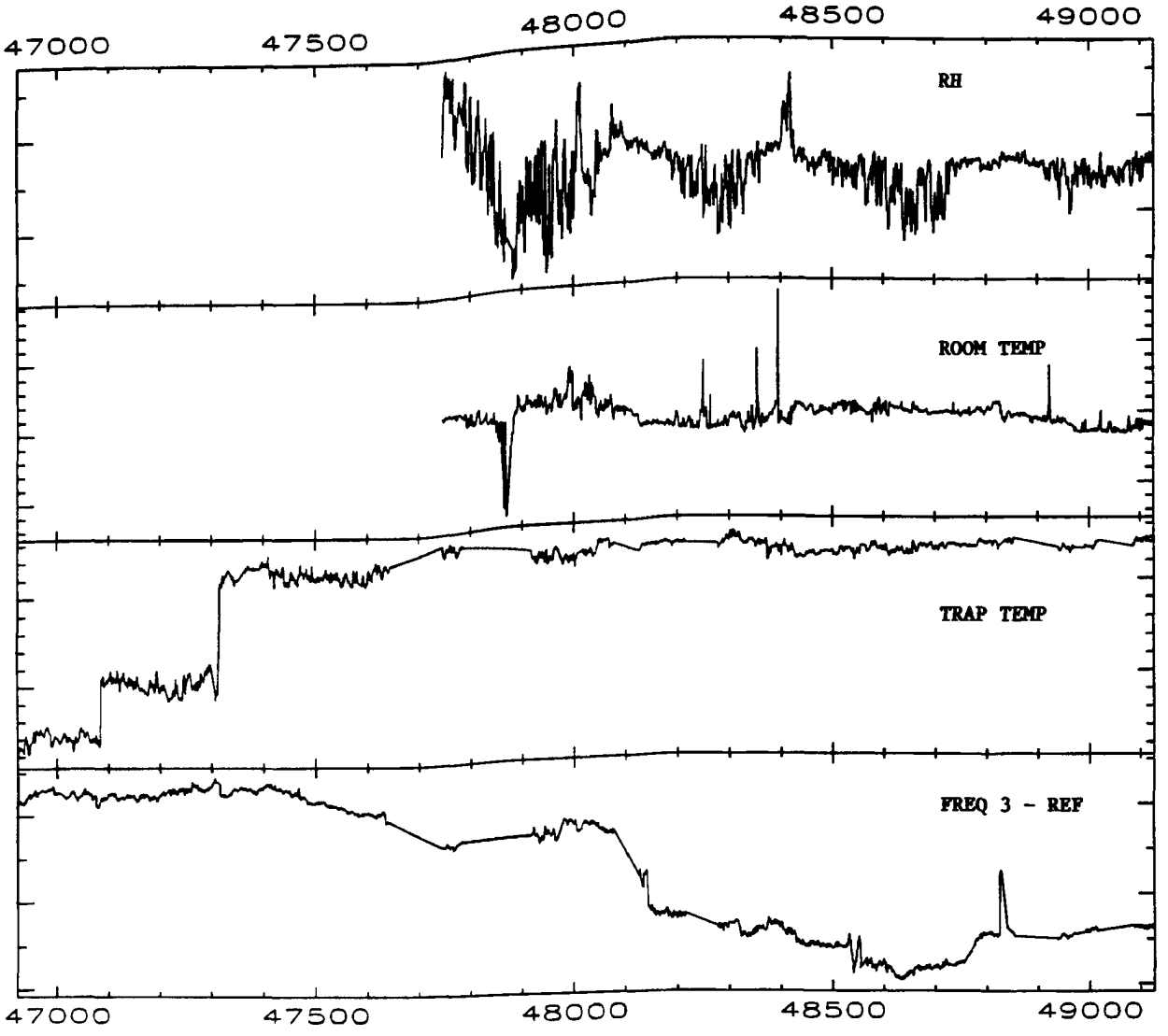


Figure 10: Autoscaled plots of unit 3 daily average clock parameters. From top to bottom, with values of minimum and maximum major tick marks in parentheses, plots are of room relative humidity (20%, 40%), room temperature (65°F, 75°F), trap temperature (43.5°C, 44°C), and output frequency minus reference in fs/s (range 740 fs/s).

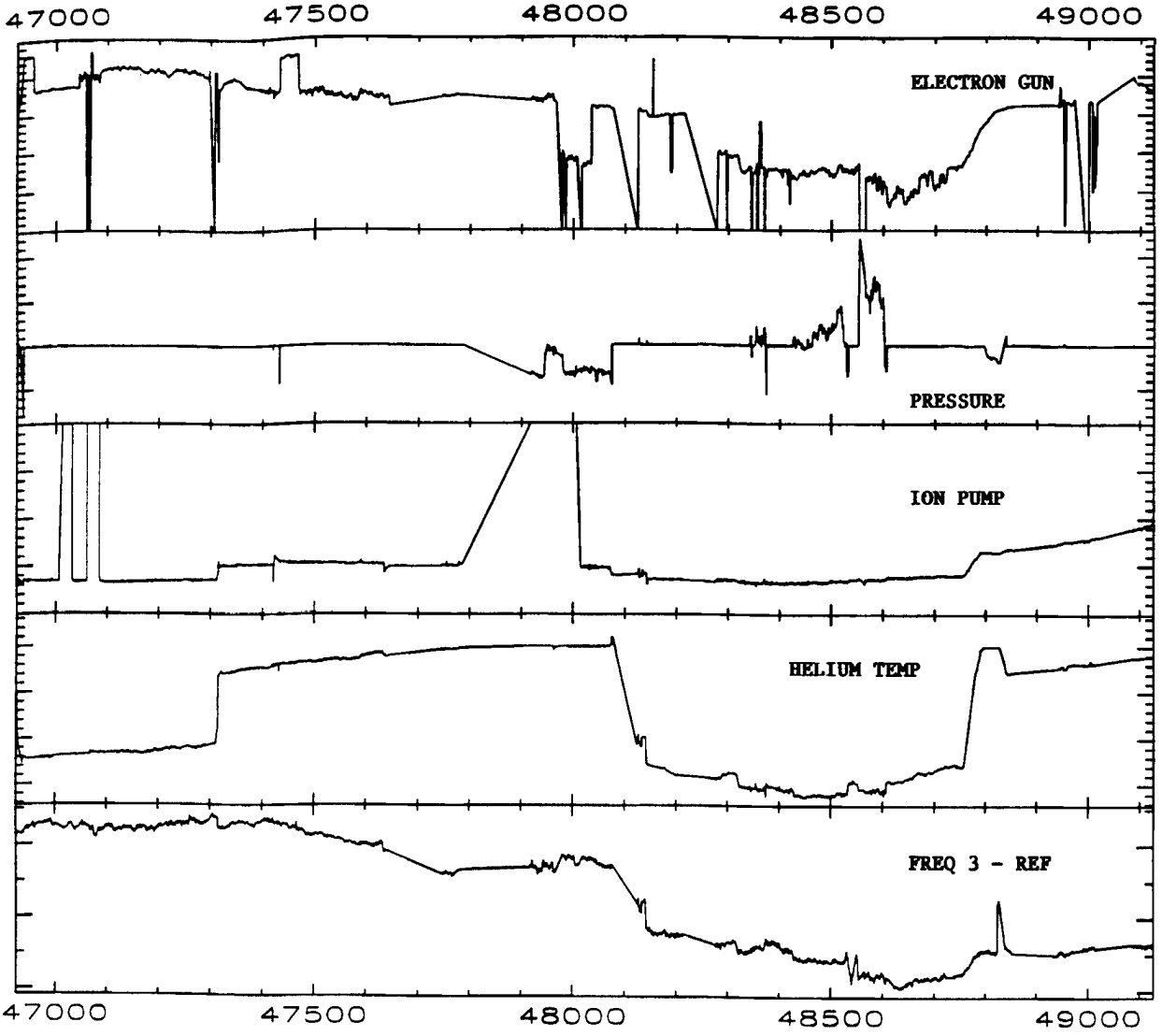


Figure 11: Autoscaled plots of unit 3 daily average clock parameters. From top to bottom, with values of minimum and maximum major tick marks in parentheses, the plots are of electron gun voltage used to stabilize the measured ion number (-0.1v, -0.4v), the vacuum pressure (1.45×10^{-6}, 1.60×10^{-6}torr; data previous to MJD 47314 decreased by 5×10^{-7}), the ion pump current (0.01, 0.04), the temperature of helium leak (65, 80 C), and the output frequency minus reference (range 740 fs/s).

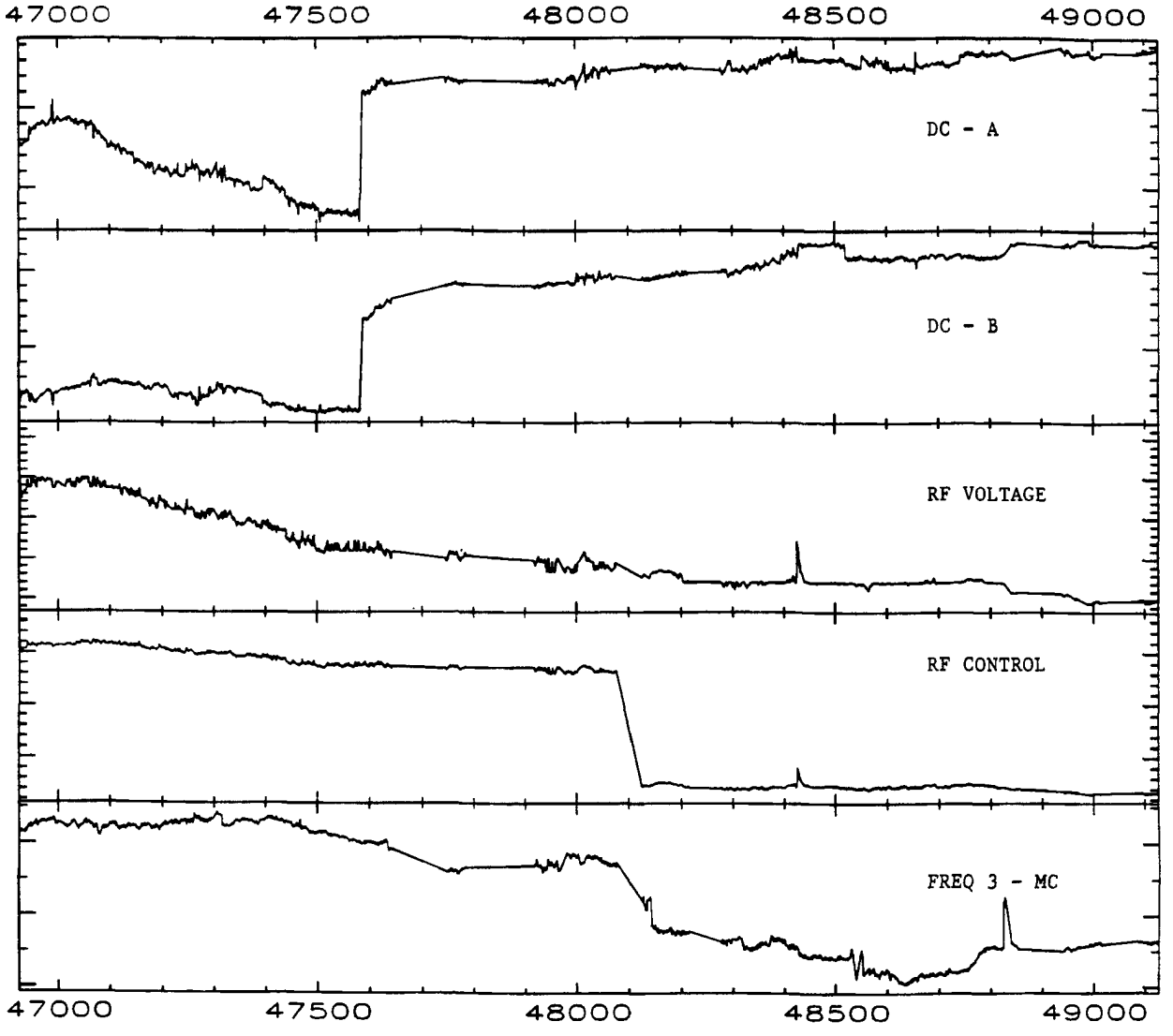


Figure 12: Autoscaled plots of unit 3 daily average clock parameters. From top to bottom: actual dc voltage of endcap a (-14.620 v, -14.615 v), actual dc voltage of endcap b (-14.570 v, -14.565 v), actual amplitude of trap rf (330 v, 334 v), control voltage used to vary the rf voltage so as to stabilize the macromotion frequency (-2.0 v, 0.0 v), and the output frequency minus reference (range 740 fs/s).

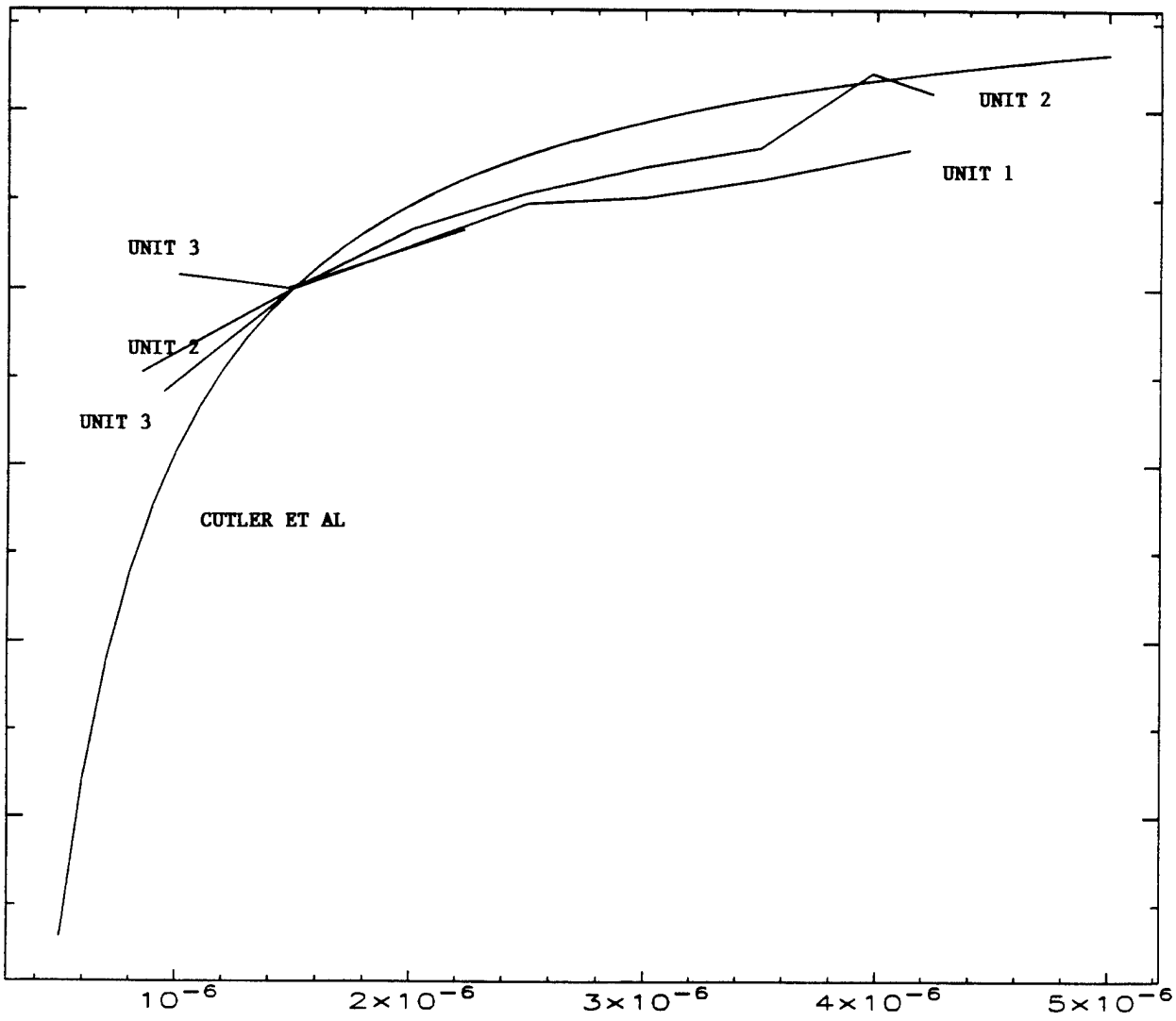


Figure 13: Frequency of stored ion devices as a function of pressure, and compared to fitted curve of Cutler et al. (1985). Vertical tick marks are .01 Hz, horizontal axis is helium pressure in torr.

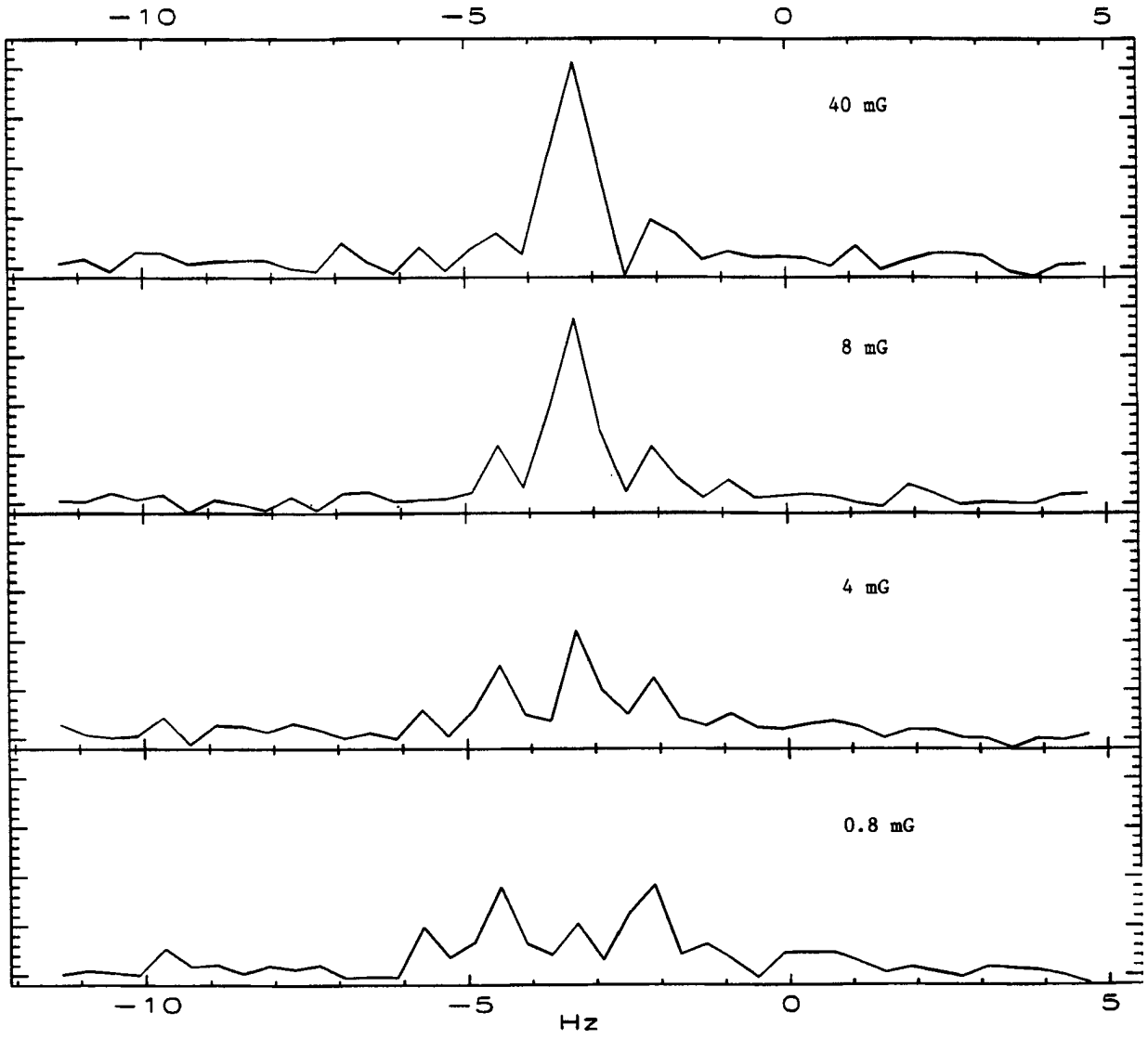


Figure 14: Resonance lineshapes of unit 3, as a function of magnetic field. Major vertical tick marks range from 0 to 2000 counts, horizontal tick marks are 1 Hz intervals. Top plot shows resonance with a 5 milliamp current generating a magnetic field of 40 mG. Lower plots are resonance in fields of 8 mG, 4 mG, and 0.8 mG (lowest plot).

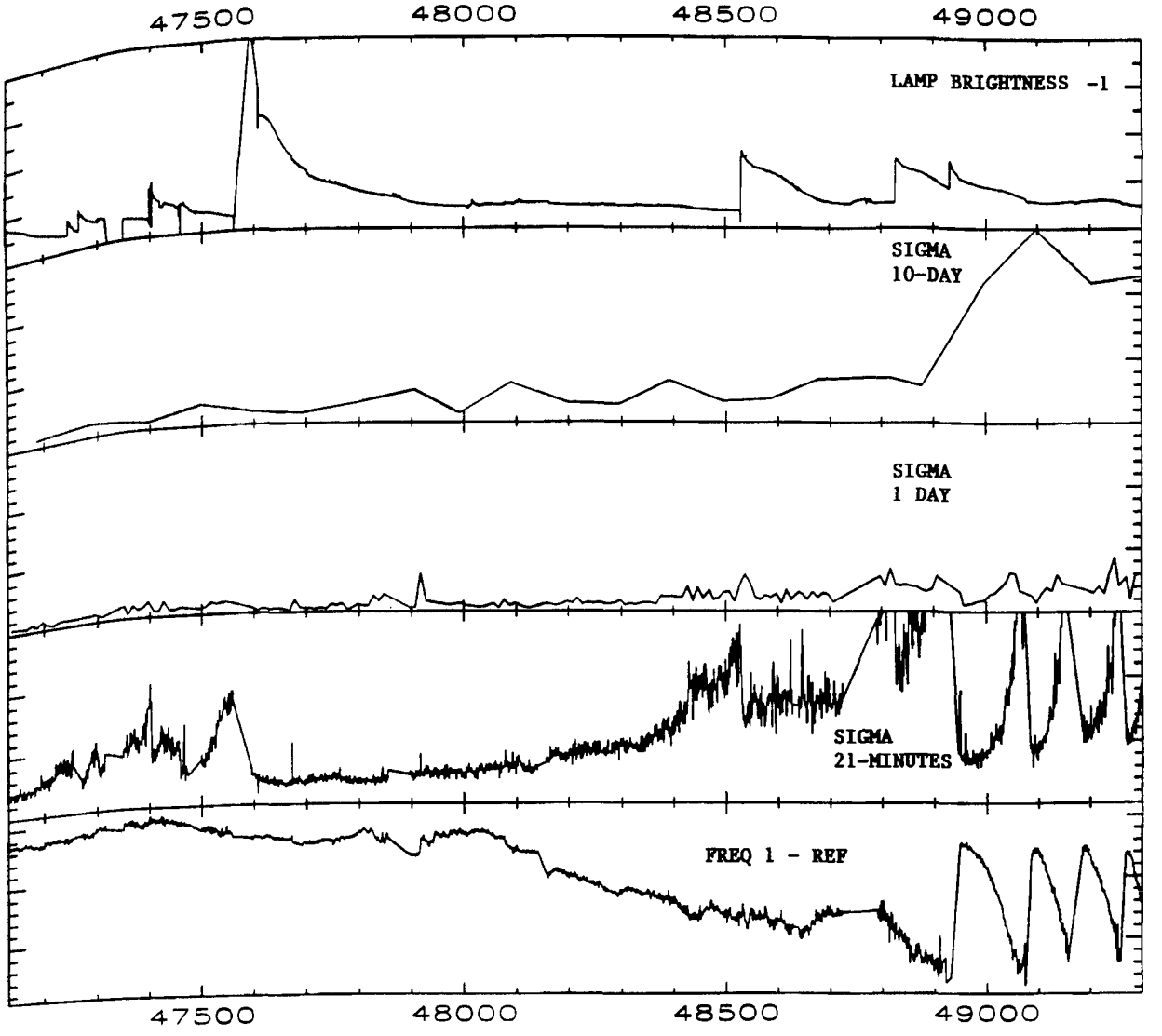


Figure 15: Lamp brightness and clock scatter of unit 1. Top plot gives off-resonance uv counts in arbitrary units, which is proportional to the measurement noise and the amplitude of the scattered and reflected uv light. The next plot is the rms of successive 10 ten-day averages, for every 100-day interval (full scale 0-300 fs/s). The third plot from the top is the rms of 1-day points (full scale 0-300 fs/s), for ever 10-day interval. The fourth plot from the top is the daily rms of the 21-minute averages (full scale 0-300 fs/s). The lowest plot is the frequency of the mercury clock minus its phase reference, plotted over a range of 1408 fs/s. In all cases the rms is computed weighting each point by the number of 21-minute runs it includes.

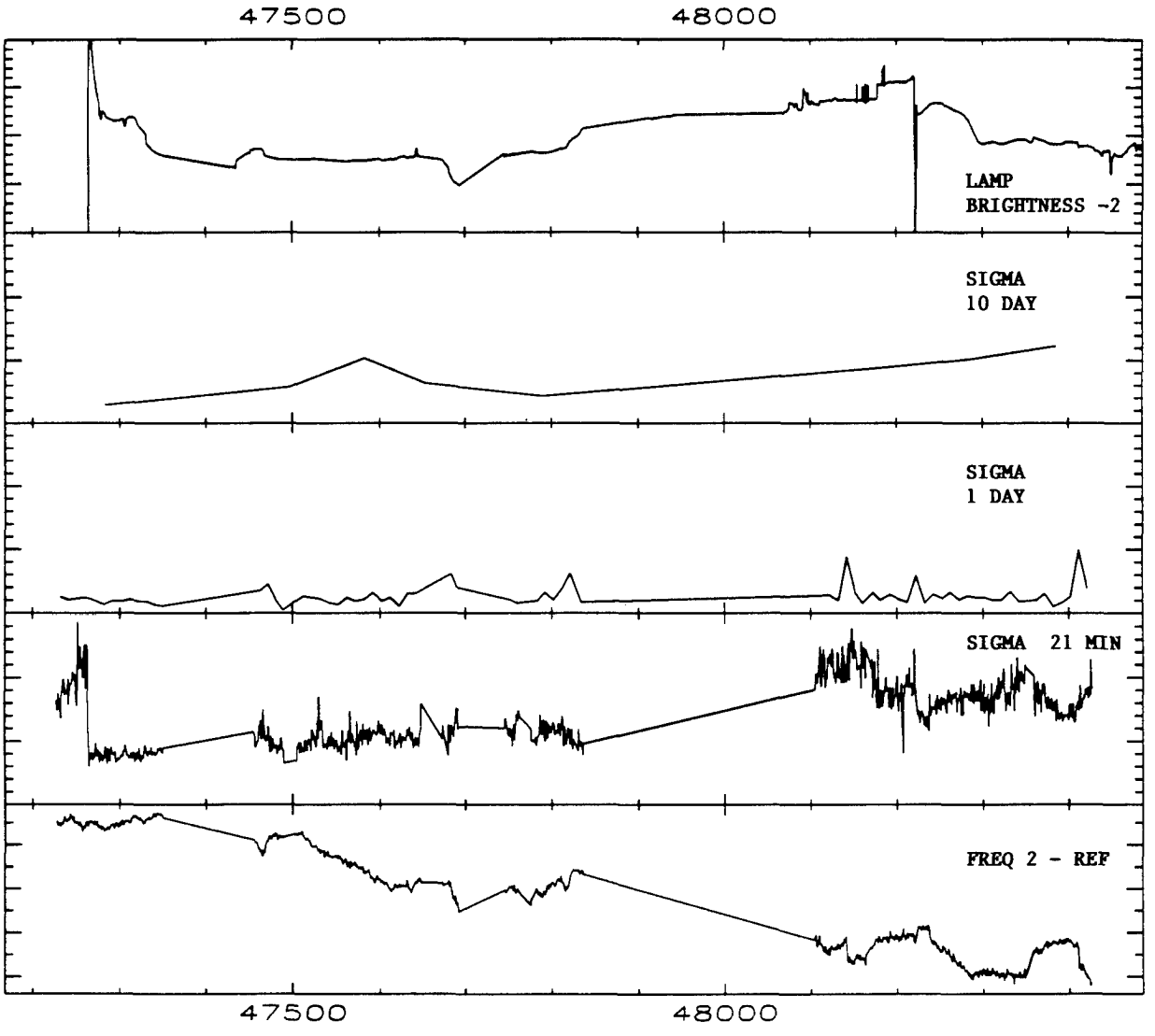


Figure 16: Lamp brightness and clock scatter of unit 2. Plots are as in previous figure, except that they apply to unit 2, the full scale on the rms plots is 160 fs/s, and on the lowest plot the range of the frequency plot is 860 fs/s.

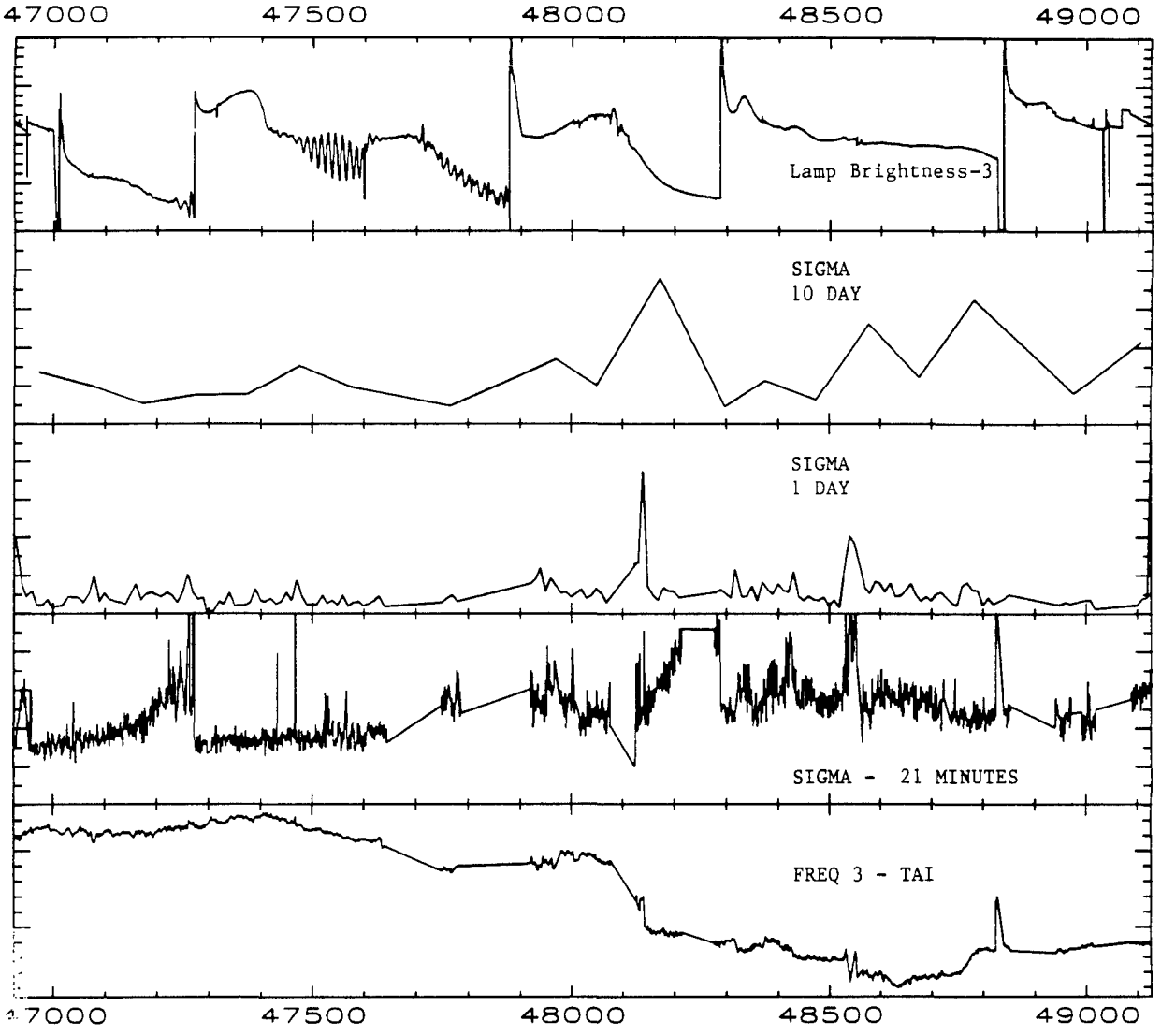


Figure 17: Lamp brightness and clock scatter of unit 3. Plots are as in previous figure, except that they apply to unit 3, the full scale on the rms plots is 100 fs/s, and on the lowest plot the range of the frequency plot is 1150 fs/s.

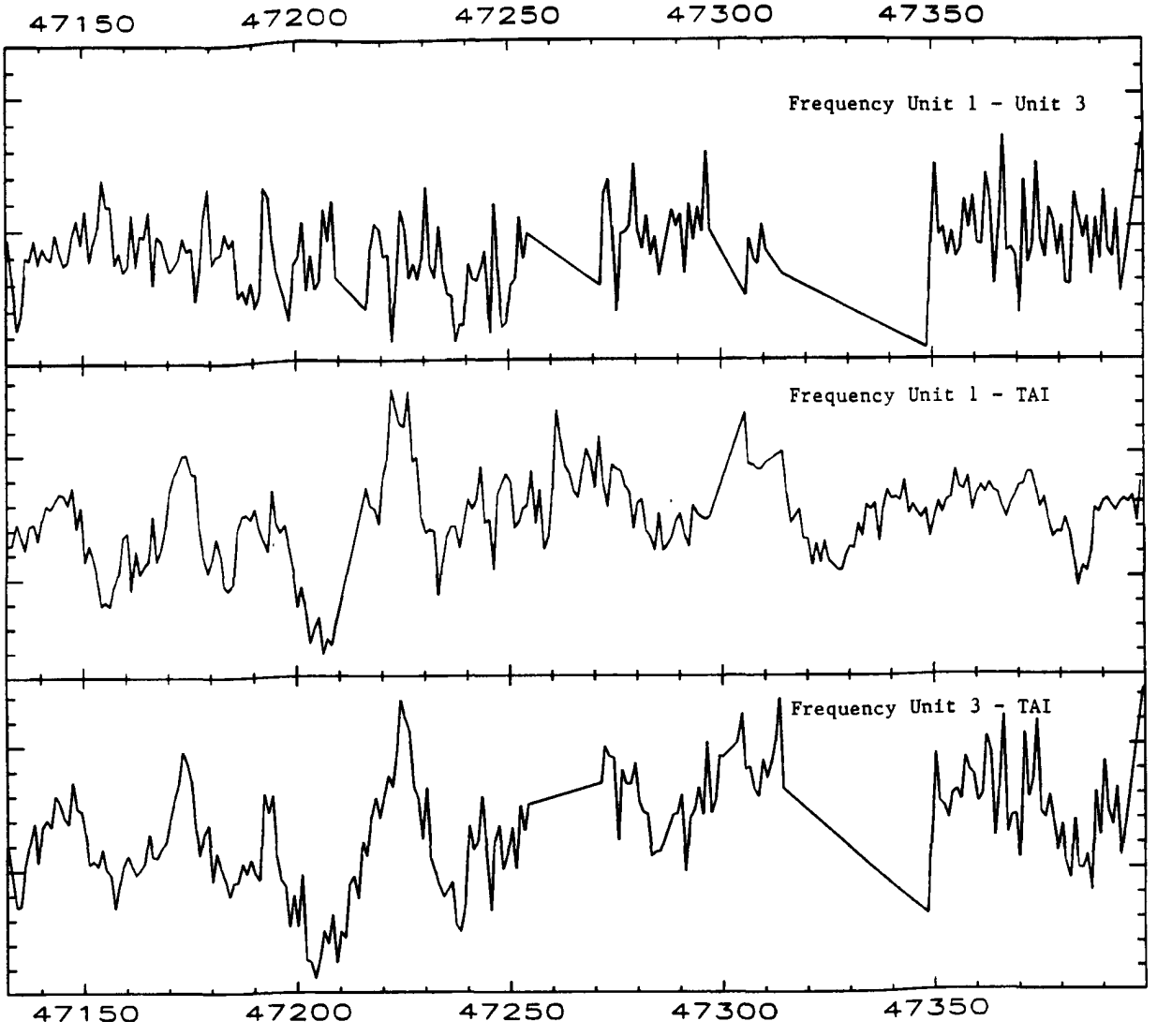


Figure 18: Frequency comparison of clocks with TAI for times of best data. Top plot is the daily average frequency difference unit 1 - unit 3. Second plot is unit 1 - TAI. Lower plot is unit 3 - TAI. Full vertical scales are $1.2 \times 10^{-13} s/s$

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A CRYOGENIC LINEAR ION TRAP FOR $^{199}Hg^+$ FREQUENCY STANDARDS*

J.D. Miller, M.E. Poitzsch,† F.C. Cruz, D.J. Berkeland J.C. Bergquist, W.M. Itano, and D.J. Wineland
Time and Frequency Division, NIST
Boulder, Colorado 80303, USA

Abstract

A cryogenic linear rf trap has been constructed to study the accuracy of a clock based on laser cooled $^{199}Hg^+$. Crystals of tens of ions have been observed. The residual micromotion of the ions has been minimized and Ramsey interrogation periods of 25 s have been demonstrated for the 40.5 GHz transition. This system should eventually provide a frequency standard of high precision and accuracy. We report on the development of compact, efficient, solid-state sources at 194 nm and 282 nm for laser cooling of $^{199}Hg^+$ and investigation of the $^2S_{1/2} \rightarrow {}^2D_{5/2}$ quadrupole transition of interest for an optical frequency standard.

$^{199}Hg^+$ Atomic Clock System

Trapped, laser cooled ions are well suited to be the basis of future frequency standards. The use of trapped ions enables long interrogation times and high resolution. Reducing the kinetic energy of the trapped ions through laser cooling decreases systematic shifts to extremely small levels. To realize the advantages of trapped, laser cooled ions for frequency standards we have constructed a linear rf trap for the confinement of $^{199}Hg^+$ ions. A frequency standard based on the 40.5 GHz ground-state transition using 50 ions in this trap has a projected stability of $\sigma_y(\tau) = 5.5 \times 10^{-14}\tau^{-1/2}$.

The 40.5 GHz ground-state hyperfine transition of $^{199}Hg^+$ ions provides the basis for a high-performance frequency standard [1,2,3,4,5]. An rf-trapped $^{199}Hg^+$ ion frequency standard (using buffer gas cooling) has been demonstrated to have high frequency stability [3]. It contained $N \sim 2 \times 10^6$ ions and had a fractional second-order Doppler shift of $\sim -2 \times 10^{-12}$. Short-term fractional frequency stability of $< 7 \times 10^{-14}\tau^{-1/2}$ has been demonstrated in a linear trap geometry (also using buffer gas cooling) [4]. Operating with $N \sim 2.5 \times 10^6$ ions, they estimated a fractional second-order Doppler shift of $\sim -4 \times 10^{-13}$. In comparison, the fractional second-order Doppler shift of a single $^{199}Hg^+$ ion laser-cooled to the Doppler limit in an rf trap is -2.3×10^{-18} [5]. To improve the signal-to-noise ratio (and hence the fractional frequency stability), it would, however, be desirable to have multiple $^{199}Hg^+$ ions, all with equally low Doppler shifts.

Cryogenic Linear RF Ion Trap

The linear rf quadrupole trap, which uses four rf rods to achieve radial confinement and a static axial potential for longitudinal confinement, was developed as a way of confining multiple ions, all with the same low Doppler shift [6,7]. In this scheme, the four rods are configured as in an rf mass analyzer, with a zero-field node all along the centerline instead of at a single central point as in a conventional quadrupole Paul rf trap [8]. Axial confinement is achieved by applying static potentials at the ends of the trap, using positively biased rings, pins, or split sections in the trap rods. Recently, we [5] have demonstrated laser cooling in a linear rf trap in the small-N regime. In that apparatus, operating at room temperature in a vacuum of about 10^{-8} Pa, we were able to "crystallize" as many as several tens of $^{199}Hg^+$ ions at fixed positions in a single row along the trap's nodal centerline. Such a geometry is optimal for the present frequency standard application, since the ions can be imaged independently for improved signal-to-noise ratio, yet all have the same low second-order Doppler shift as a single ion in a quadrupole trap. The major limitation of this apparatus was the background gas pressure in the UHV chamber, which was still high enough that ions would be lost due to chemical reactions after times on the order of a few tens of minutes. At this pressure, pressure shifts could also limit the accuracy [9].

Our solution to the background gas pressure problem is to maintain the trap and vacuum vessel at liquid helium temperature ($\sim$ 4 K). At this low temperature, most gases cryopump to the walls of

*This work was supported by the ONR. Work of the U.S. Government, not subject to U.S. copyright.
†Present address: Schlumberger-Anadrill, Sugar Land, TX 77478.

U.S. Government work not protected by U.S. copyright

the chamber, giving a very low background pressure. In a similar sealed vacuum can, lowered to 4 K, Gabrielse *et al.* [10] report background pressures below 10^{-14} Pa. By thus lowering the pressure by several orders of magnitude, we should be able to store trapped ions for at least several days, interrogate them with Ramsey free-precession times as long as tens or hundreds of seconds, and be relatively insensitive to possible pressure shifts of the 40.5 GHz clock frequency. Operation of the trap at 4 K also reduces fractional shifts of the clock transition due to blackbody radiation to $< 1 \times 10^{-20}$ [11].

We have constructed and are testing a prototype apparatus based on these concepts. The trap is a small linear rf quadrupole, with four 0.40 mm diameter rods centered on a radius of 0.64 mm from the trap axis (about half the size of our previous trap [5]). Axial confinement is achieved with positively biased rings at either end of the four-rod quadrupole. The separation of the two rings is 4 mm. The trap and related apparatus are mounted in an indium-sealed OFHC copper vacuum can, inside a nested LHe/LN_2 dewar, heat-sunk to the outside bottom of the LHe reservoir. In addition to the trap, the vacuum vessel contains magnetic field coils, a miniature 40.5 GHz microwave antenna, a 5-element $f/1$ lens for 194 nm that can survive temperature cycling from 373 K to 4 K, and an HgO oven and a tungsten filament for loading ions into the trap. The trap is driven at 13 MHz with a few mW of rf using a superconducting helical resonator (immersed in the liquid helium) to step up the drive voltage to $\sim$ 100 V. Optical access to the trap region is through baffled windows around the base of the dewar. Magnetic shielding is achieved with conventional shields external to the liquid helium dewar.

Preliminary Results and Prospects

The trap and related apparatus are currently being tested. We can load and optically resolve individual cold ions, coalesced into linear crystals with inter-ion spacings of 10–30 μm. We have seen crystals ranging in number from one to several tens of ions. The rf micromotion of the ions has been minimized in three dimensions and crystals of tens of ions remain crystalized when the cooling beam is blocked for approximately 100 s. These crystals are very stable over periods of several hours when the cooling beam is on. We have seen no loss of ions from the trap while the ions are laser cooled, indicating that the background gas pressure is low. We have observed 20 mHz linewidths of the 40.5 GHz clock transition. Narrower linewidths should be possible with the use of a new 40.5 GHz source with higher resolution. With sufficient magnetic shielding, a fractional inaccuracy of $< 1 \times 10^{-16}$ appears attainable.

In addition, this apparatus should allow us to investigate new effects based on motional Zeeman coherences. These include a novel cooling scheme (proposed by Harde [12]) using optical pumping in conjunction with a motional magnetic coupling between the spin orientation and the harmonic oscillator state of the ions in the trap potential, as well as a scheme for "squeezing" the total ensemble spin, which could improve the signal-to-noise ratio in frequency standards where the dominant noise contribution is quantum fluctuations [13].

Solid-State 194 nm Source

Progress is being made toward an all solid-state source of 194 nm radiation which is needed to laser cool and optically detect mercury ions. Currently, the 194 nm light is generated by frequency doubling an argon-ion laser locked to a hyperfine component in iodine at 514.67 nm to obtain radiation at 257.34 nm. This is then sum frequency mixed with a stabilized diode laser operating at 790 nm. In the future, a frequency-doubled, compact, efficient, diode-pumped Yb:YAG laser that operates at 1.03 μm will replace the argon-ion laser [14]. More than 1 W of single frequency output has been obtained at 1.03 μm from the Yb:YAG laser pumped by three diode lasers. Each diode laser delivers about 1 W of power at 807 nm. We expect to double the Yb:YAG laser to 515 nm with greater than 50% efficiency [15].

Optical Frequency Standard

For the optical frequency standard based on the 282 nm $^2S_{1/2} \rightarrow ^2 D_{5/2}$ quadrupole transition in $^{199}Hg^+$, a quadrupled solid-state Nd:FAP laser at 1.126 μm is being developed. Single frequency efficiency at 1.126 μm of approximately 30%, when pumping with 807 nm, has been obtained. This laser will be frequency doubled to 563 nm in a monolithic cavity and then doubled again to 282 nm in a single-pass configuration. The better inherent frequency stability of a diode-pumped, all solid-state laser system should give a compact, reliable source for the optical local oscillator.

Acknowledgements

We thank the U.S. offices of Naval Research for support.

References

[1] F.G. Major and G. Werth, Phys. Rev. Lett., Vol. 30, p. 1155, 1973.

[2] M.D. McGuire, R. Petsch, and G. Werth, Phys. Rev. A, Vol. 17, p. 1999, 1978.

[3] L.S. Cutler, R.P. Giffard, P.J. Wheeler, and G.M.R. Winkler, in Proc. 41st Ann. Symp. Freq. Control, p. 12, 1987.

[4] R.L. Tjoelker, J.D. Prestage, G.J. Dick, and L. Maleki, in Proc. 47th Ann. Symp. Freq. Control, p. 132, 1993.

[5] M.G. Raizen, J.M. Gilligan, J.C. Bergquist, W.M. Itano, and D.J. Wineland, Phys. Rev. A, Vol. 45, p. 6493, 1992.

[6] H.G. Dehmelt, in Frequency Standards and Metrology, ed. A. DeMarchi. Berlin: Springer, p. 286, 1989.

[7] J.D. Prestage, G.J. Dick, and L. Maleki, J. Appl. Phys., Vol. 66, p. 1013, 1989.

[8] W. Paul, O. Osberghaus, and E. Fischer, Forsch. Ber. des Wirtschaftsministeriums Nordrhein-Westfalen, Nr. 415, 1958.

[9] D.J. Wineland *et al.*, in Laser Manipulation of Atoms and Ions, ed. E. Arimondo, W.D. Phillips, and F. Strumia. Amsterdam: North-Holland, p. 553, 1992.

[10] G. Gabrielse *et al.*, in Atomic Physics 12, ed. J.C. Zorn and R.R. Lewis. New York: AIP, p. 549, 1991.

[11] W.M. Itano, L.L. Lewis and D.J. Wineland, Phys. Rev. A, Vol. 25, p. R1233, 1982.

[12] H. Harde, in Int. Conf. on Quantum Electronics Tech. Digest, Series 1990, Vol. 8. Washington, DC: OSA, p. 178, 1990.

[13] D.J. Wineland, J.J. Bollinger, W.M. Itano, F.L. Moore, and D.J. Heinzen, Phys. Rev. A, Vol. 46, p. R6797, 1992.

[14] T.Y Fan and J. Ochoa, submitted to Appl. Phys. Lett. 1995.

[15] Z.Y. Ou, S.F. Pereira, E.S. Polzik and H.J Kimble, Opt. Lett., Vol. 17, p. 640, 1992.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Systematic Errors in Cesium Beam Frequency Standards Introduced by Digital Control of the Microwave Excitation

W. D. Lee, J. H. Shirley, F. L. Walls, and R. E. Drullinger
National Institute of Standards and Technology
Time and Frequency Division
Boulder, Colorado, USA
Tel: 303-497-7674

Abstract— **As part of the accuracy evaluation of the NIST-7 primary frequency standard, we have investigated potential sources of frequency offsets due to the electronic servo system. We present results from investigations of RF spectral purity and synthesizer switching transients.**

I. Introduction

The Ramsey technique of separated oscillating fields is widely used in the field of atomic frequency standards. We have investigated potential sources of frequency shifts associated with the microwave radiation interrogating the atoms when using the Ramsey technique. We first measured the frequency bias that results from an interrogation signal whose spectrum contains imbalanced sidebands. We also investigated the frequency offsets that arise in determining the center of the Ramsey fringe using slow square-wave digital servos. These servos include a blanking interval to allow the atomic beam tube to reach steady state during the frequency modulation cycle. In addition to atomic beam transit time, there is an additional time delay that should be included in the servo's blanking interval. This delay is caused by anomolous switching transients in a direct digital frequency synthesizer.

II. RF Sideband Pulling

The presence of spurious, discrete sidebands in the spectrum of the microwave radiation can lead to errors in the determination of the center of the Ramsey peak. While these effects are not unique to slow square-wave digital frequency servos, they may be more pronounced in systems employing direct digital synthesis (DDS) due to phase accumulator truncation [1], and electromagnetic interference from high-speed digital electronics.

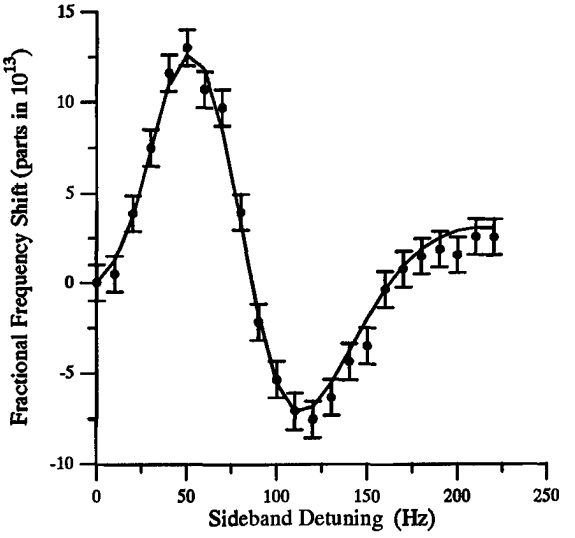


Fig. 1. Frequency Shift Due to Sideband Pulling. Solid line is the theoretical calculation. Dots are the experimental measurements. Carrier power: -2.5 dB relative to optimum power; lower sideband: -29 dBc; upper sideband: -66 dBc. Experiment duration: 300 s per point.

The frequency shift due to a discrete sideband has been calculated theoretically [2]. An independent calculation by one of us (J.H.S.) has confirmed that result, except that in their equation (9) the drift time T should be replaced by $T + \tau$, where τ is the transit time through one interaction region. For sideband detunings of the order of the Ramsey linewidth, we obtain shifts like those shown by the solid line in figure 1 when we average over the velocity distribution.

The calculations were verified experimentally using the United States primary frequency standard NIST-7 by intentionally modulating the microwave signal in both amplitude and phase to produce a well defined spectrum. The amplitude and phase of the carrier must be modulated coherently to yield imbalanced

U.S. Government work not protected by U.S. copyright

sidebands. Consider the sum of three signals: a carrier, an amplitude modulated carrier, and a phase modulated carrier.

$$\begin{aligned} E(t) &= \mathrm{Re}\{Ae^{i\theta}e^{i\omega t} \\ &+ \beta Ae^{i\theta}e^{i\omega t}[1+\epsilon\cos(\Omega t+\phi)] \\ &+ \gamma Ae^{i\theta}e^{i\omega t}e^{i\eta\sin(\Omega t+\psi)}\} \end{aligned} \quad (1)$$

where A is the carrier wave amplitude, ω is the carrier frequency, θ is the carrier phase, ϵ is the amplitude modulation index ($\epsilon \leq 1$), Ω is both the amplitude and phase modulation frequency, ϕ is the phase of the amplitude modulation, η is the phase modulation index, and ψ is the phase of the phase modulation. β and γ are the relative amplitudes of the amplitude modulated and phase modulated signals with respect to the unmodulated carrier. If $\epsilon \ll 1$ and $\eta \ll 1$ this can be rewritten in three terms representing the carrier, the upper sideband and the lower sideband, respectively:

$$\begin{aligned} E(t) &= \mathrm{Re}\{(1+\beta+\gamma)Ae^{i\theta}e^{i\omega t} \\ &+ \frac{1}{2}Ae^{i\theta}(\epsilon\beta e^{i\phi}+\eta\gamma e^{i\psi})e^{i(\omega+\Omega)t} \\ &+ \frac{1}{2}Ae^{i\theta}(\epsilon\beta e^{-i\phi}-\eta\gamma e^{-i\psi})e^{i(\omega-\Omega)t}\}. \end{aligned} \quad (2)$$

The upper and lower sideband relative amplitudes can then be written

$$S_+ = \frac{1}{2(1+\beta+\gamma)}(\epsilon\beta e^{i\phi}+\eta\gamma e^{i\psi}), \quad (3)$$

$$S_- = \frac{1}{2(1+\beta+\gamma)}(\epsilon\beta e^{-i\phi}-\eta\gamma e^{-i\psi}). \quad (4)$$

If we choose $\epsilon\beta = \eta\gamma = \delta$, then the relative sideband intensities are

$$|S_+|^2 = \frac{\delta^2}{2(1+\beta+\gamma)^2}[1+\cos(\phi-\psi)], \quad (5)$$

$$|S_-|^2 = \frac{\delta^2}{2(1+\beta+\gamma)^2}[1-\cos(\phi-\psi)]. \quad (6)$$

The upper and lower sidebands have equal intensity for

$$(\phi-\psi) = \frac{\pi}{2} \pm n\pi, \text{ for } n = 0, \pm1, \pm2, \ldots. \quad (7)$$

To measure the frequency pulling with high accuracy and within a short time, we chose to introduce the maximum sideband asymmetry (a single sideband). This is achieved by setting

$$(\phi-\psi) = 0 \pm n\pi, \text{ for } n = 0, \pm1, \pm2, \ldots. \quad (8)$$

Such a single sideband modulator is realized with the circuit of figure 2. A low phase noise, 9-GHz microwave source (OSC-MAIN) is modulated in both amplitude and phase by two independent modulators of the type described in [3]. One modulator is optimized to produce high purity AM, while the other is optimized to generate only PM. Two phase-locked direct digital synthesizers ($\mathrm{DDS_{AM}}$ and $\mathrm{DDS_{PM}}$) provide the audio frequency modulation signals ($\cos(\Omega t+\phi)$ and $\sin(\Omega t+\psi)$) to two mixers. The synthesizers are adjusted (under computer control) so that $\epsilon\beta = \eta\gamma$ and $\phi = \psi + \pi$. This is verified by monitoring the microwave spectrum with a spectrum analyzer (FFT) at an intermediate frequency of 50 kHz, using a local oscillator (OSC-LO). The amplitudes of the carrier, lower, and upper sidebands are then recorded by the computer.

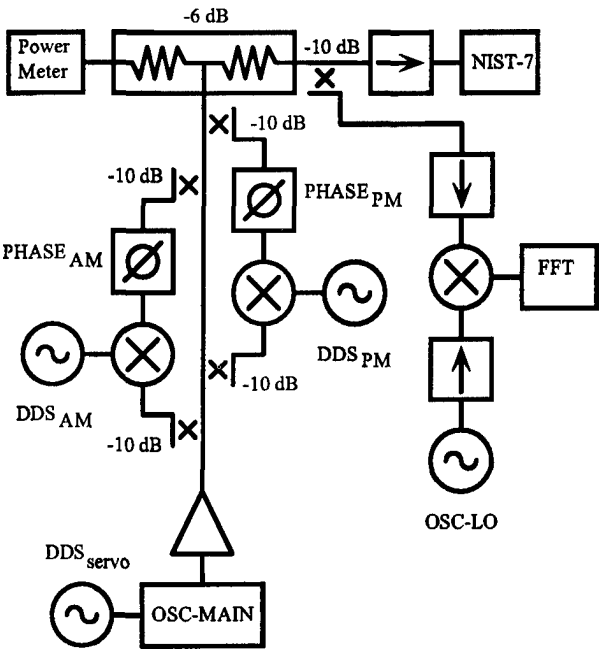


Fig. 2. Single Sideband Modulator.

The frequency bias introduced by this single sideband is determined by measuring the center of the Ramsey fringe using a slow square-wave frequency servo. The computer controlled synthesizer ($\mathrm{DDS_{SERVO}}$) modulates and steers the microwave source (OSC-MAIN). Figure 1 shows the measured frequency pulling (dots) as a function of the sideband detuning Ω. The error bars represent $\pm1\sigma$ statistical uncertainty. Since the experimental velocity distribution [4], sideband intensities, and detunings were used in calculating the theoretical curve, no free parameters were adjusted to obtain the agreement shown in figure 1. We have obtained similar results for detunings com-

parable to the width of the Rabi pedestal.

The theoretical curve in figure 1 is antisymmetric about zero detuning; hence symmetric sidebands produce shifts that cancel each other. Pure amplitude modulation generates symmetric sidebands about the carrier and does not lead to a frequency shift. Similarly, neither does pure phase modulation. This was verified experimentally by enabling only one branch of the single sideband modulator of figure 2 (either AM or PM) and measuring the center of the Ramsey fringe as a function of modulation frequency. Figure 3 shows that there is no significant frequency pulling for either amplitude modulation or phase modulation sidebands at -29 dBc.

The sideband pulling data of figure 1 can be used to calculate the frequency bias due to a pair of sidebands near the Ramsey fringe if the power dependence of the shift is known. The theory predicts that the shift is linear in sideband power. This was verified experimentally by measuring the center frequency of the Ramsey fringe as a function of sideband power. The results are shown in figure 4.

From these results, estimation of the frequency bias due to spurious RF sidebands can be made a routine part of the accuracy evaluation of a primary frequency standard.

III. Blanking Intervals in Slow Square Wave FM Servos

The use of direct digital synthesis in microwave synthesis chains has grown in recent years. This is due to improvements in the frequency resolution and agility of such synthesizers. Slow square-wave frequency servos are readily implemented when DDS instruments are computer controlled. The task of evaluating the frequency biases in such a servo is greatly simplified if the beam tube and electronics are allowed to reach steady state before the clock signal is measured. Thus, after a frequency step command is sent to the DDS, a blanking interval τ_b is established, during which the clock signal is discarded. After the blanking interval, the clock signal is recorded during the acquisition interval τ_a. It is desireable to have τ_a occupy a large fraction of the modulation period in order to enhance frequency stability. Simply reducing the modulation frequency is an unacceptable approach to increasing the ratio τ_a/τ_b, because it increases the system's sensitivity to noise that exhibits $1/f$ power spectra.

The blanking interval can be written as the sum

$$\tau_b = \tau_r + \tau_\phi + \tau_\rho, \qquad (9)$$

where τ_r (10 ms) is the response time of the DDS, that is, the time from the transmission of a frequency

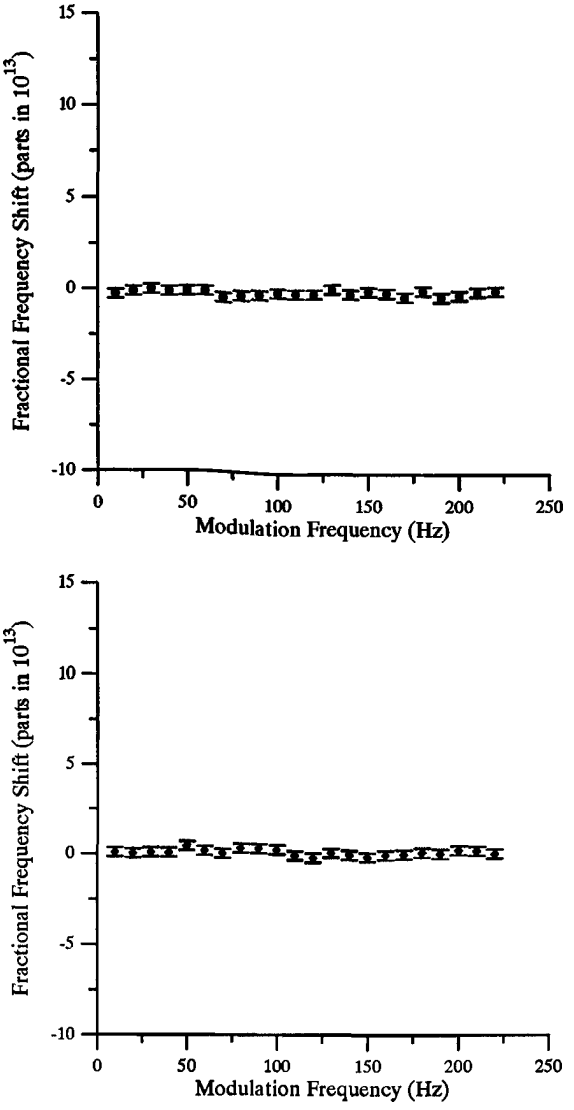


Fig. 3. Frequency Shift for Pure AM or PM
Upper Plot: frequency shift vs. amplitude modulation frequency. Lower Plot: frequency shift vs. phase modulation frequency. The modulation sidebands for both cases are -29 dBc. Experiment duration: 3000 s per point. For comparison, the horozontal and vertical scales are the same as figure 1.

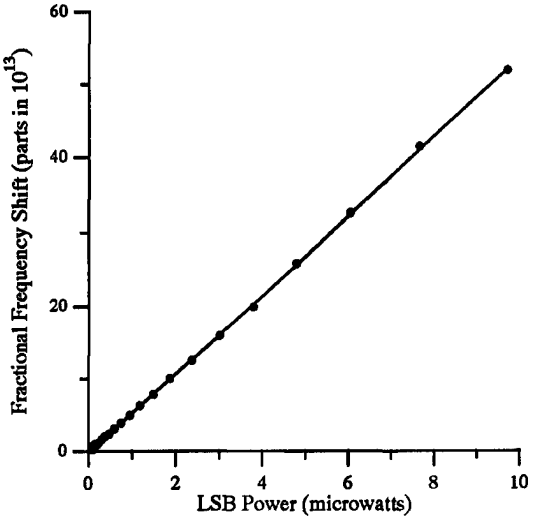


Fig. 4. Frequency Shift vs. Sideband Intensity
Dots are the experimental measurements. The 1σ uncertanties are equal to the symbol size. Carrier power: -2.5 dB relative to optimum power; sideband detuning: 50 Hz below carrier. Experiment duration: 1800 s per point. Solid line is a linear fit to the data.

step command to the change in frequency at the DDS output. τ_ϕ is phase-settling time of the DDS and τ_ρ (25 ms) is the transit time of the slowest atoms of interest across the Ramsey cavity and to the detection region. There is a lower limit placed on τ_b by τ_r and τ_ρ. We have, however, discovered techniques whereby τ_ϕ may be eliminated, improving frequency stability and eliminating a potential frequency bias.

IV. Synthesizer Switching Transients

We have found a potential frequency bias due to phase transients in a commercial direct digital synthesizer. The transients occur whenever the synthesizer is sent frequency control instructions by the computer. These transients are therefore synchronous with the frequency modulation, some occurring tens, and even hundreds of milliseconds after the frequency step is performed. Since the position of the Ramsey fringe is sensitive to the phase difference between the two ends of the cavity as seen by atoms taking several milliseconds to traverse the cavity, such phase transients can give rise to significant frequency offsets.

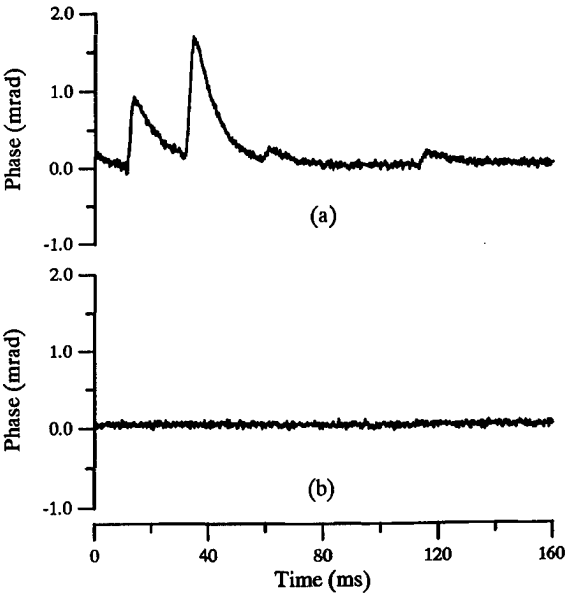


Fig. 5. Phase Transients Due to Computer Control.
Trace (a): phase transients occuring after frequency control command is sent to synthesizer at $t = 0$ ms. Trace (b): supression of phase transients when internal oscillator is bypassed.

The transients were measured by synchronously recording the relative phase of two synthesizers while one was sent frequency control commands by a computer. To simplify the measurement, the same frequency control command was sent to the synthesizer repeatedly, not changing the output frequency. Figure 5(a) shows the resulting phase transients. The peak phase deviation of 1.7 mrad occurs 35 ms after the instruction is sent to the synthesizer.

When a DDS unit is used in a microwave synthesis chain, it is common to provide the device with an external frequency reference. The DDS then phase-locks its own internal crystal oscillator to the external signal. We think that the "housekeeping" operations of the microprocessor within the DDS are inadvertently coupled to this phase-locked loop, generating phase transients that are synchronous with its operation. We were able to eliminate the phase transients by removing the instrument's internal crystal oscillator and instead injecting a suitable external reference frequency directly into the synthesizer's timing circuits. Figure 5(b) shows the dramatic improvement in phase stability.

Figure 6 shows the effect of the phase transient upon the atomic beam fluorescence as well as the normal

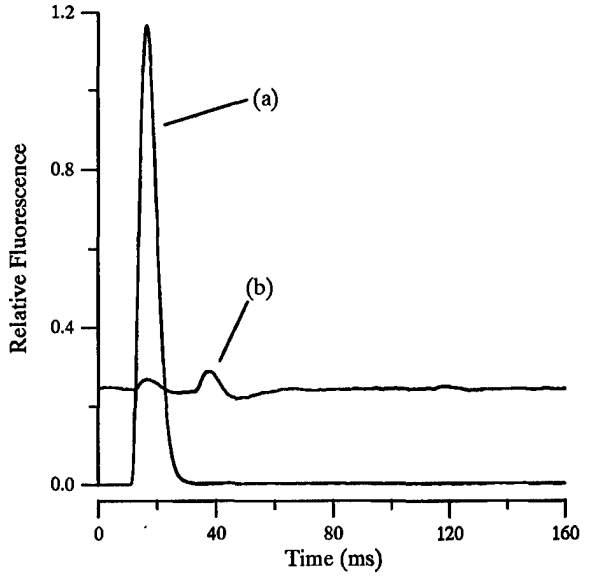


Fig. 6. Effect of Synthesizer Phase Transients. Trace (a): Normal slow square wave FM transient due to the atomic transit time distribution. This trace was taken with the frequency servo operating, thus the steady state fluorescence levels before and after the transient are equal. Trace (b): Effect of DDS phase transients (figure 5(a)) on atomic beam fluorescence with frequency fixed on the side of the Ramsey fringe. Trace (b) has been offset for clarity.

modulation transient. If we were to determine a blanking interval based solely upon the response time of the synthesizer and the atomic beam transit time distribution, the effect of the synthesizer switching transient would still introduce a significant contribution to the measured fluorescence.

The size of the frequency bias due to the switching transient was determined by measuring the center of the Ramsey fringe using two different direct digital synthesizers. Both units were identical except that one unit was modified to eliminate the phase transient. With a blanking interval of 40 ms, the fractional frequency bias produced by the transient was $(50 \pm 1) \cdot 10^{-13}$. Although the blanking interval might be extended to avoid the initial switching transients, periodic transients (visible in figure 5(a) at 120 ms) are virtually unavoidable.

We emphasize that these phase transients preserve their sign regardless of the sign of a frequency step performed by the DDS. Thus, the resulting bias cannot be eliminated through complex or pseudorandom modulation sequences. By removing these transients, we were able to transfer at least 20 ms from the blanking interval to the acquisition interval in each half of the modulation cycle.

V. Conclusion

We have investigated sources of bias in digitally controlled microwave synthesis chains: spurious RF sidebands and synthesizer switching transients. The experimental results obtained for RF sideband pulling are in excellent agreement with theory. They apply not only to digital servo systems but also to traditional analog servos. While the synthesizer switching transients we observed were unique to a specific model and might be avoided by selecting a different DDS, our modifications yielded a DDS with phase stability superior to other models we have tested.

References

[1] Nicholas, H. T., Samueli, H., "An Analysis of the Output Spectrum of Direct Digital Frequency Synthesizers in the Presence of Phase-Accumulator Truncation," 41st Annual Frequency Control Symposium, 1987, pp. 495-502.

[2] Audoin, C., Jardino, M., Cutler, L. S., and Lacey, R. F., "Frequency Offset Due to Spectral Impurities in Cesium-Beam Frequency Standards," IEEE Trans. Inst. Meas. vol. IM-27, 1978, pp. 325-329.

[3] Walls, F. L., "Extending the Range and Accuracy of Phase Noise Measurements," Proceedings of the 42nd Annual Frequency Control Symposium, 1988, pp. 432-441. U. S. Patent 4,968,908, "Method and apparatus for wide band phase modulation." (1990).

[4] Lee, W. D., Shirley, J. H., and Drullinger, R. E., "Velocity Distributions of Atomic Beams by Gated Optical Pumping," Proceedings of the 1994 IEEE International Frequency Control Symposium, pp. 658-661.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A TRANSPORTABLE HYDROGEN MASER FOR AUSTRALIA TELESCOPE

J. W. He, C. F. Lin, Z. C. Zhai
Shanghai Observatory, the Chinese Academy of Sciences,
Shanghai, 200030, P. R. of China
W. E. Wilson* and H. Fagg *
*Australia Telescope National Facility, CSIRO,
PO Box 76, Epping, NSW, 2121, Australia

Abstract

To equip the Chinese VLBI network, a transportable compact active hydrogen maser has been developed at Shanghai Observatory(SO) since 1985. So far 5 such masers have come into use for VLBI radio astronomy and some military. These hydrogen masers are all compact, relatively light, rugged and easily transportable. Including all the electronics system, power supplies and stand-by batteries, their size is 57 cm x 83 cm x 127 cm and their mass is not more than 300 Kg. In Nov., 1993, one of these masers was transported to Australia Telescope National Facility(ATNF) to determine its performance as a stable oscillator for VLBI. This paper describes the design and construction of the hydrogen maser and presents the results tested at Parkes Observatory of ATNF.

1. Introduction

Very Long Baseline Interferometry (VLBI) involves the use of a network of radio telescope to observe an astronomical source so that a high resolution map can be made of that source. During the observations the receiver local oscillator at each telescope is phase-locked to a high stability oscillator to allow recording of the signals in phase coherence with the other observations. The stable oscillators in common use are hydrogen masers. To equip the VLBI network in China, Shanghai Observatory began to develop a transportable, compact active hydrogen maser in 1985[1]. This device is still based on the traditional TE_{011} mode cavity masers made early at Shanghai Observatory[1], but there are many improvements in the new design. It emphasizes ruggedness and mechanical and thermal stability. All electronics are electrically integrated into the overall design. The entire device was to be self-contained in a rack cabinet. Its size is 127 cm X 57 cm X 83 cm and its weight is about 300 kg. The maser is operated from 220 volts AC power, the internal stand-by batteries make it possible to operate the maser during AC power failures. The overall electrical consumption is 200 watts at +24 Volts DC. Figure 1 shows the photograph of one of this kind of transportable hydrogen masers located at Gravity Wave Laboratory of the University of the Western Australia (UWA), which is being compared with the sapphire oscillator as a reference.

This paper we describe the design and construction of the hydrogen maser and present the results of tests at Parkes Observatory for a VLBI stable oscillator.



Figure 1. The photograph of the transportable hydrogen maser located at gravity research laboratory of UWA.

2. Design and Construction

0-7803-2500-1/95 $4.00 © 1995 IEEE

I. Physical Package

Figure 2 shows a overall physical package of the hydrogen maser. In this physical package design, we also incorporated many ideas from SAO(Smithsonian Astrophysical Observatory)[2].

The cavity-bulb construction is to be as inherently stable as possible. The cavity is constructed of a material with a very small thermal expansion coefficient (1 to $2x10^{-7}/^{0}C$) and long-term resistance to dimensional creep. The mass of the storage bulb is reduced to 160g to reduce the effect of thermal coefficient due to dielectric loading. Possible cavity-stress changes owing to expansion of the external cylinder holding the cavity together are taken up by the baliville Washer. Similar thermally induced radial stress in the base is relieved by the rollers. Strain on the bell jar owing to barometric pressure change is isolated from the cavity by the double-base structure. Thus a very tightly controlled thermal and mechanical environment is to be established for the cavity to permit maser operation over long period of time without auto-tuning or any other adjustment. A typical achieved temperature coefficient is 150 $Hz/^{0}C$.

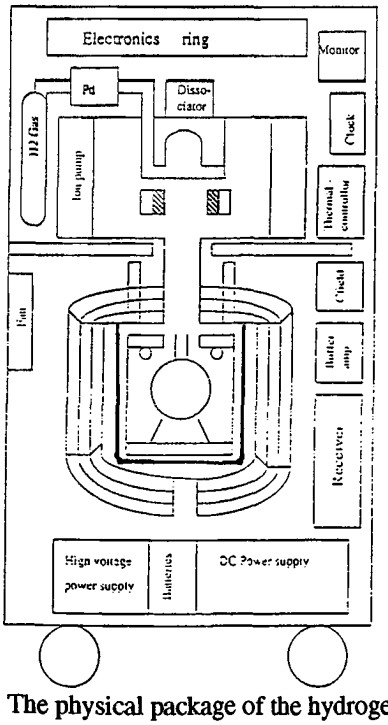


Figure 2. The physical package of the hydrogen maser.

The vacuum system is pumped by a single vac-ion pump with a capacity of 150L s^{-1}. This pump can function continuously for 4 to 5 years.

Magnetic Shielding is produced by four-layer covers within foam-thermal insulation. Their shielding factor is about 40,000. A multi-layer printed-circuit C field coil system which can effectively reduce the spurious magnetic field is located inside the innermost magnetic shield cover.

The hydrogen gas at a pressure of about 0.1 torr from a two-litre gas bottle is admitted to a glass tube(its dimension is 3.8 cm in diameter X 4.0 cm in length) surrounded by a inductance coil of RF oscillator, which produces a plasma discharge within the tube. The required DC power is 5 to 8 Watts and the resonant frequency is about 86 MHz. The palladium Leak is controlled by comparing the ion pump current with a set value in a comparator. The state selector for atoms is a classical hexapole magnet with a smaller size than before.

To minimize the temperature gradient, the vacuum tank and the oven are divided into three zones: dome, cylinder and base; and two zones: dome-cylinder and base, respectively. Each zone has its own sensing thermistor, amplifier and heater windings and can respond independently to external thermal loads. In addition, there are a independent thermal-control system which controls the temperature of a isolator-preamplifier box and a circulating air system which controls the temperature of the pump, the dissociator and the upper maser electronics.

II. Electronics

Phase-locked Receiver

The phase-locked receiver (Figure 3) is a low noise, phase-stable superheterodyne converter with 2 intermediate frequency at 20.4 MHz and 405.7 kHz. A small integrated digital synthesizer delivers a 405.7 kHz local oscillator signal. The frequency of the 100 MHz phase-locked voltage controlled oscillator (VCO) can be adjusted by means of the synthesizer with relative steps of $7X10^{-15}$.

To obtain the best short term frequency stability, we use, 1) a low-noise amplifier with less than 1.5 dB noise figure and a more than 90 dB isolator, 2) a 1400 MHz phase-locked multiplier with a small thermal phase drift, 3) a image reject filter located in front of the first-mixer.

To achieve the best long term frequency stability transfer of the maser to VCO, a optimized phase-locked loop has been designed.

Digital Clock

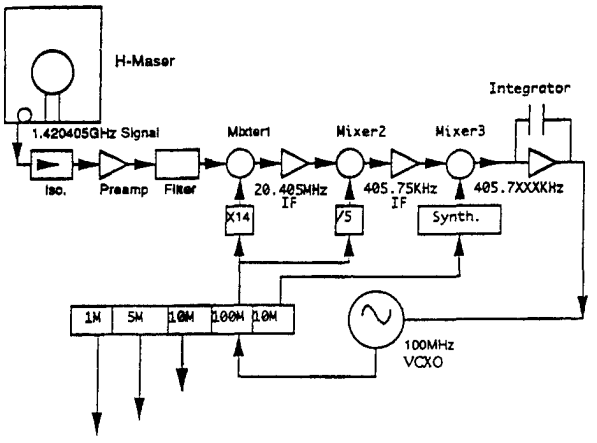


Figure 3. The block diagram of the phase-locked receiver system

The digital clock provides two 1 pps signals which are produced from a 10 MHz signal which in turn is derived from the 100 MHz phase-locked VCO. The timing of the 1 pps signals may be adjusted in steps of 0.1μs. The two 1 pps signal are also synchronised to an extra 1 pps signal by means of an extra synchronic input connector. A LED display indicates time in hours, minutes and seconds.

Monitor system

Four front-panel mounted meters with an eight-position selector switch for each are located on the monitor panel to provide quick-look monitoring of 32 functions. These include all main power-supply voltages, all heater voltages, hydrogen dissociator operation conditions, receiver/synthesizer signal level and phase-lock control voltages.

Available Signals

There are five 5 MHz signals (+13 dBm), one 1 MHz signal (+13 dBm), one 100 MHz signal (>+7 dBm) and one 1400 MHz signal (>0 dBm). There are also two 1 pps signals whose timing is adjustable in 0.1 μs steps (TTL level). Isolation between the 5 MHz signals is better than 65 dB under operation conditions.

3. Measurement Results

To determine its performance as a VLBI stable oscillator, the hydrogen maser was installed at Parkes Observatory of ATNF from Nov. 4 to Dec. 15, 1993[3]. The results are as follows.

Figure 4 shows the short to medium term frequency stability of the hydrogen maser. The results were measured by comparing two 800 MHz signals, one locked to the Parkes maser made by SAO and the other locked to the Shanghai hydrogen maser.

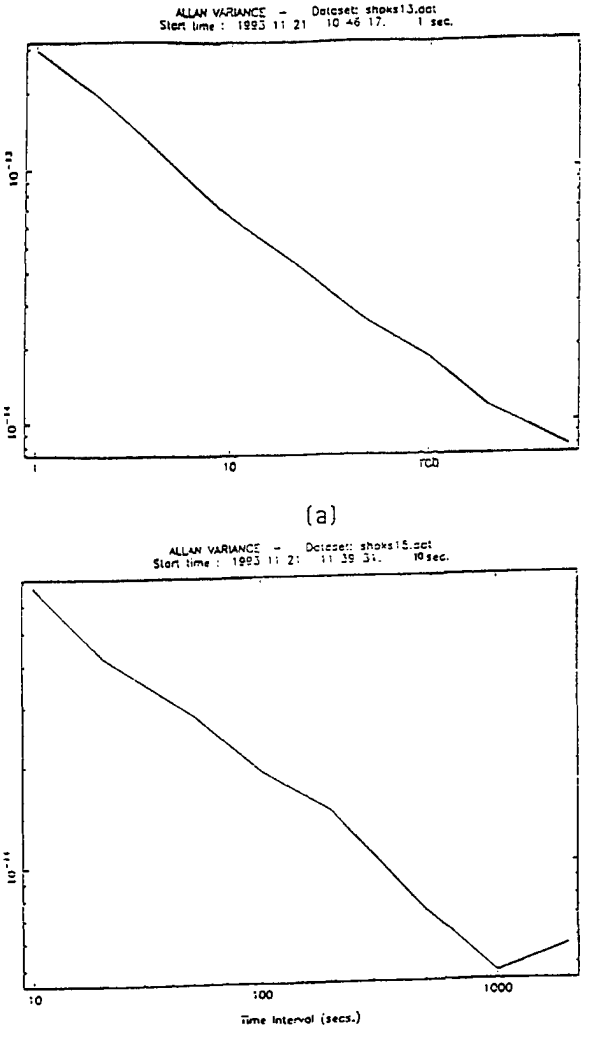


Figure 4. The Allan standard deviation(square root of the Allan variance) of the hydrogen maser compared with SAO VLG-10 maser. The results aren't divided by a square root of 2.

Figure 5 shows the long term relative frequency stability of the two hydrogen masers. The result was measured by including the 1 pps signal from the Shanghai hydrogen maser in the standard Parkes time calibration system. This allowed comparison

with 1 pps signals from the Parkes maser, GPS and the Tidbinbilla maser, whose 1 pps signal is transferred to Parkes via a microwave link.

Figure 6 shows an estimate of the amount of coherence loss due to instrumental phase noise-not including the atmosphere-that one might expect in a VLBI experiment which used the two masers as frequency standards. The observing frequency was assumed to be 40 GHz and an integration time of 50 seconds was chosen.

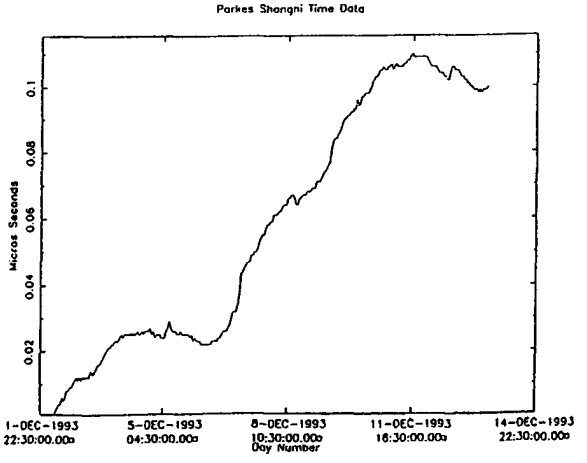


Figure 5. The long term frequency stability of the hydrogen maser compared with SAO VLG-10 Maser at 1 pps.

The loss of coherence at the beginning and end of the 16 hour dataset is probably an indication of the inadequacy of the linear clock model which was subtracted from the data.

Tests were made of the environmental systematic effects at Shanghai Observatory before the hydrogen maser was shipped to Australia. In these tests, the frequency of the maser output signal was carefully monitored while one of the environmental conditions was varied. The reference was another of the same type masers. The results are summarised in Table 1 [4].

Table 1. Environmental Sensitivities

Conditions	Sensitivities
Temperature (22^0C to 31^oC)	Slow Fast ($x10^{-14}/^oC$) T up 2.9, 1.2 T down 3, 1.3
Barometric pressure (0<Δp<50mmHg)	$<1.2x10^{-15}$/mmHg
Magnetic Field(+/-0.4G)	$< 1.1X\ 10^{-13}$/G

4. Conclusions

The short to medium term frequency stabilities tested above have shown the hydrogen maser would be a very good VLBI frequency standard. At present, three of this kind of hydrogen masers have been used in China VLBI network and two have been used in mobile tracking stations. This transportable hydrogen maser will also be used in the Australia Telescope.

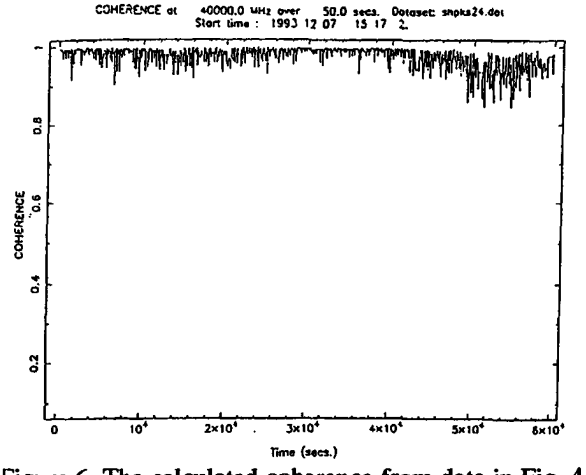


Figure 6. The calculated coherence from data in Fig. 4 for VLBI when the hydrogen masers are used at each telescope.

5. Acknowledgements

This work was supported by Shanghai Observatory, the Chinese Academy of Sciences and Australia Telescope National Facility, CSIRO. The authors would also like to thank A/Prof. D G Blair (UWA) for his suggestions during writing this paper.

6. References

[1]. Zhai Z C, Huang H X, Jiang G X,]etc., "The new generation of hydrogen maser at Shanghai Observatory", Proc. of 41st Annual Symposium of Frequency Control, 1987.

[2]. R F C Vessort, M M Levine, E M Mattison, etc. ,"Space-born Hydrogen Masers Design", Proc. of 18th PTTI Meeting, 1986.

[3]. W E Wilson, "Shanghai Maser Tests at Parkes" , ATNF, 1994

[4]. Z C Zhai, H X Huang, J W He, etc.,"The Development Progresses of Hydrogen Maser at Shanghai Observatory", Proc. of 6th European Frequency and Time Forum", 1992.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DORIS PRECISE ORBIT DETERMINATION AND LOCALIZATION SYSTEM DESCRIPTION AND USOs PERFORMANCES

M. BRUNET, A. AURIOL, P. AGNIERAY, F. NOUEL

CENTRE NATIONAL D'ETUDES SPATIALES (CNES),
18 Avenue E.Belin, 31055 TOULOUSE, FRANCE

ABSTRACT

The Centre National d'Etudes Spatiales (CNES), the French Space Agency, has designed and developed the Doppler Orbiting and Radiopositioning Integrated by Satellite (DORIS) system, able to determine within few centimeters, the location of satellites and ground beacons. DORIS payload has been integrated on three French SPOT satellites, on TOPEX NASA satellite, and will be embarked on ENVISAT ESA satellite.
Based on one way Doppler accurate measurements from a network of 50 ground beacons, localization performances depends strongly on the frequency stability of on-board and ground Ultra Stable Oscillators (USO).
The paper describes elements of the DORIS system, and gives performances of orbit computations and precise positioning. It emphasizes performances analysis of on-board and beacons USOs, measured by CNES T/F lab, and via orbit measurements.

1. THE DORIS SYSTEM

1.1 The DORIS system components (figure 1)

The DORIS system comprises a "flight segment" composed of a receiver, an USO and a dual frequency omnidirectional antenna, a "ground segment" which includes the control, command and data processing center, and a network of Orbitography Determination Beacons (ODBs) and Ground Location Beacons (GLBs). GLBs are functionally identical to ODBs, but field packaged.

1.2 The flight segment

The on-board DORIS instrument comprises:
- A fixed omnidirectional dual frequency antenna.
- A receiver performing the Doppler measurements every ten seconds. Current receivers perform measurements over one beacon at a time. The nominal functioning mode is a programmed mode in which it tracks the beacons signals according to a preplanned work table which is uploaded daily.
- An Ultra Stable Oscillator (USO) delivering a very stable frequency reference.

DORIS instrument has been integrated on SPOT 2 and SPOT 3 French satellite in orbit since February 1990 and October 1993, on TOPEX-POSEIDON NASA satellite in orbit since September 1992, and on SPOT 4 French satellite which will be launch in 1998. DORIS instrument will be also integrated on ENVISAT ESA satellite. SPOT and TOPEX satellites are on circular orbits at an altitude of 830 km and 1 300 km respectively.

1.3 The ground segment

The raw data are received from satellite ground segment by the DORIS control and processing center located in Toulouse, France and operated by C.L.S. company (a CNES subsidiary).

The DORIS control and processing center performs the following:
- Acquisition of telemetry raw data.
- Ground beacons and on-board instrument monitoring.
- Instrument programming.
- Measurements time-tagging and pre-processing.
- Operational orbit determination.
- Beacon precise positioning.
- Data monitoring and archiving.

The CNES Service d'Orbitographie DORIS (SOD) performs the precise orbit determination and evaluation. The ground segment includes also a network of 50 Orbitography Determination Beacons (ODBs) (see figure 2) installed in very accurately known places (geodesic points). They are used to calculate continuously the orbit parameters and are the reference system for the location determination. Installation and maintenance of the network is realised by the French Institut Geographique National (IGN). The location of these beacons is precisely linked to an International Reference Frame (IERS) to get absolute positioning. The frequency reference scale for the DORIS system is provided by two master beacons in Toulouse, France and Kourou, French Guyana, which are drived by an USO slaved to an atomic frequency standard. The time reference is TAI to avoid leap seconds.

0-7803-2500-1/95 $4.00 © 1995 IEEE

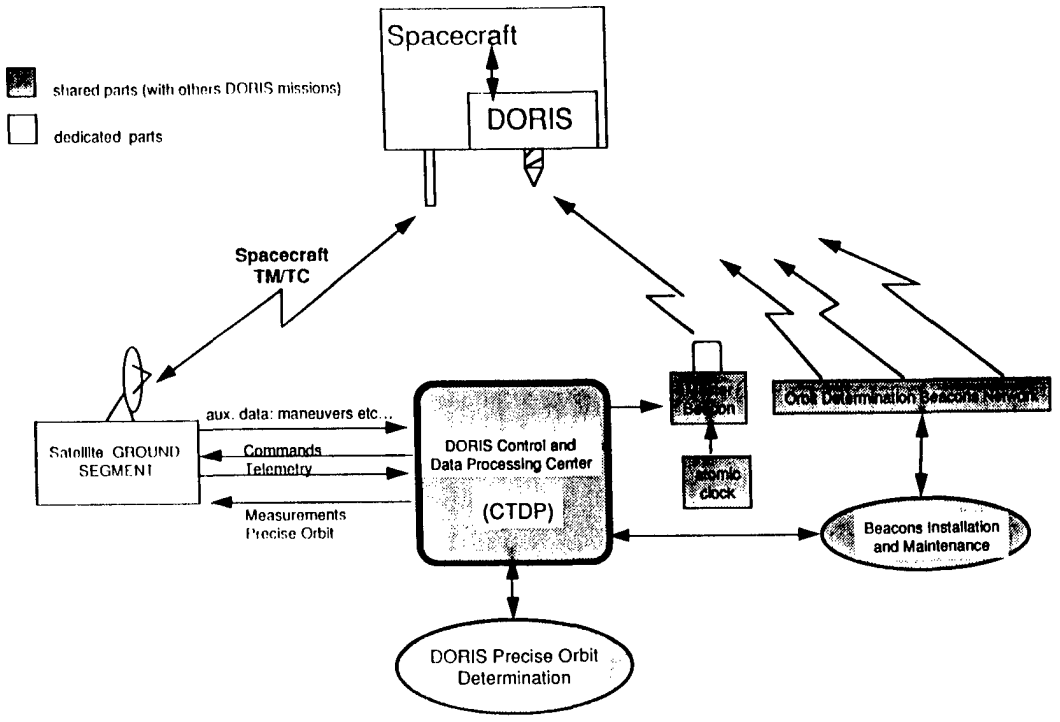


Figure 1: DORIS system components

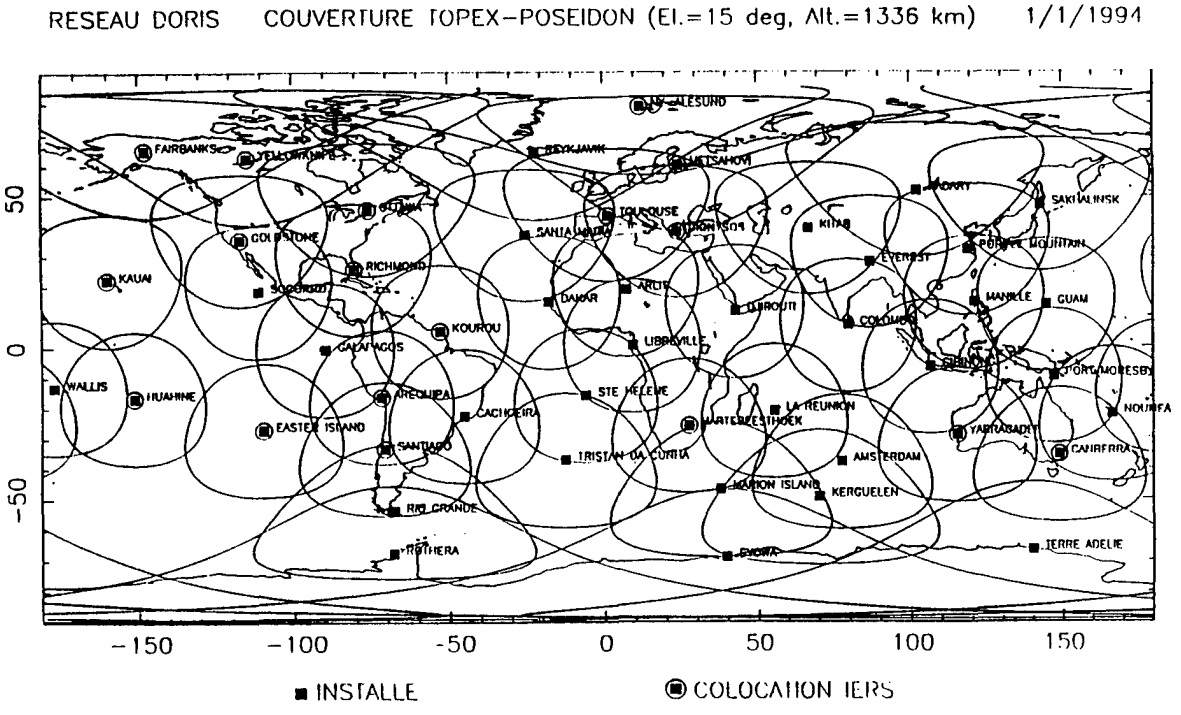


Figure 2: DORIS network (March 1993) and 15° visibility circles for T/P

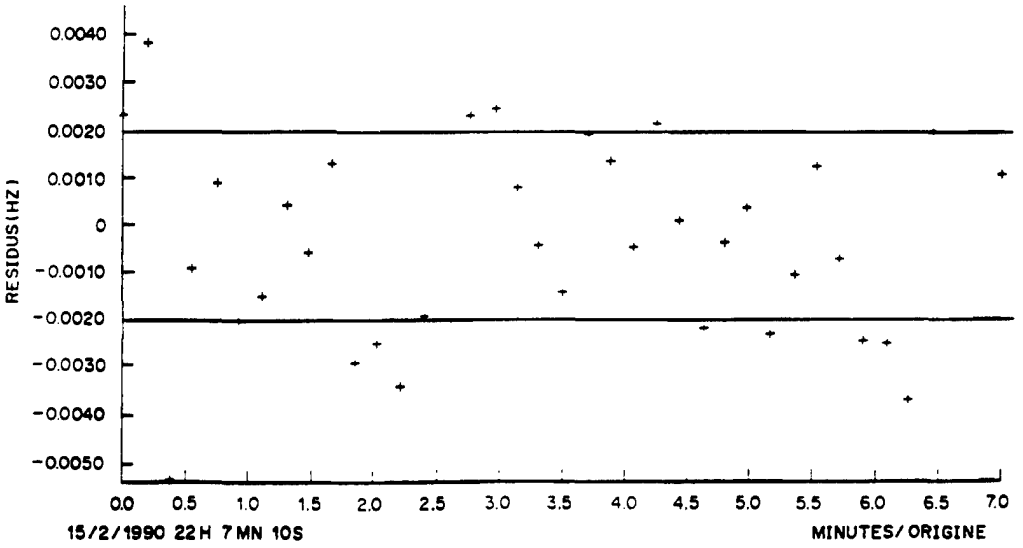


Figure 3: Results of "Guier" filtering: short term noise

ODBs have 4 units

- A main unit including an USO whose performances are similar to the on-board USO, driving 2 transmitters at 2.03625 GHz and 401.25 MHz with power of 5 Watts dual frequency.
- An omnidirectional dual-frequency antenna.
- A battery back up unit to provide autonomy while main AC supply is off.
- A meteorological package providing pressure, temperature and humidity measurements which are transmitted to the up link to allow tropospheric correction to the measurements.

Ground Location Beacons (GLBs) are also installed according the different users needs.

GLBs have same units as ODBs, but they are packaged and equiped with autonomous power supply (Solar array, ...).

2. DORIS PERFORMANCES

2.1 Instrument performances

The main characteristics and performances of the on-board instrument are summarised as follow:

- Total mass of the units composing the instrument (antenna, receiver, USO): 15 kg.
- Power consumption: 20 W.
- On-board time-tagging accuracy: 100 μs.
- Doppler frequency measurement precision: standard deviation 0.5 mm/sec or 1.6 . 10^{-12}. This figure includes effects of ground and on-board instruments (USOs, ground transmitter, on-board receiver, antennas) and atmospheric propagation effects. This standard deviation is composed of:
  1) Short term noise (rms of 10 seconds measurements) coming from on-board and ground oscillators, thermal noise, quantization and other instrumental errors. This noise is easy to evidence in the orbit residuals, since it has only high frequencies, and may be isolated by filtering using GUIER. Some very quiet pass show a 0.20 mm/sec runs (6.6 . 10^{-13}).
  2) The medium term instrument errors are due mainly to the natural noise of oscillators (Allan variance of DORIS USOs vary from 5 . 10^{-13} over 10 sec to 1 . 10^{-12} over 10 minutes), and to tropospheric propagation mismodelling during the satellite pass. Vertical tropospheric values are about 2.4 meters, estimated with an accuracy of 2-3 cm from the three meteorological ground sensors.

Figure 3 from [1] gives the short term noise on individual 10 seconds measurement over 10 minutes pass, using the state of the art "GUIER" filtering technique which removes the best "orbit-like" Doppler curve. The solid line shows the typical rms value of 0.3 mm/sec, or 2 mHz on the 2 GHz channel.

2.2 Coverage and availability

DORIS-SPOT 2 is in use since February 1990, DORIS-TOPEX since September 1992, and DORIS-SPOT 3 since October 1993.

The actual coverage (percentage of time during which the instrument is doing measurements) is 65 % for SPOT altitude (830 km) with an minimum elevation of 12° and 80 % for TOPEX altitude (1 300 km) with a minimum elevation of 15°.

The reliability of the system is about 95 % in term of Doppler measurement production. Data losses are mainly due to Single Event Upsets which necessitate re-initialization of the on-board instrument programming. Nevertheless, thanks to short data outages. TOPEX precise orbit has been produced 100 % of the time since the beginning of the mission. Furthermore, new receivers will use a SEU less sensitive processor.

2.3 Precise orbit determination performances

The Centre National d'Etudes Spatiales (CNES) prepared for the T/P precise orbit determination task in collaboration with the Space Geodesy Branch of the Goddard Space Flight Center (GFSC) and the Center for Space Research (CSR) of the University of Texas. Besides an excellent relation ship motivated by the will to succeed, and a scrupulous desire for rigorous comparison, there were also significant individual differences in the strategies to compute orbits among centers. There are worth describing and bring additional confidence in the results, especially when centimeters of accuracy is the result. While there is no absolute way to certify to the users that the 10 cm accuracy is obtained, an effective strategy is to produce several orbits with different software or procedures and compare them.

As a result, there are two orbits: the CNES precise orbit ephemeris (CNES POE) and the NASA POE provide by the Service d'Orbitographie DORIS (SOD) and the GSFC respectively.

The CNES POE and NASA POE are continuously exchanged. The two orbits are systematically compared by the SOD expert system before final orbit delivery. The RMS orbit differences in three components over each cycle represents a key test at the present time.

Figure n° 4 from [6] shows the mean RMS difference value over the T/P life time is 2.5 cm for the radial component. It is around 6 cm for the cross-track component and around 10 cm for the along-track component. These comparisons do not reflect the tracking data measurements (DORIS and/or laser) for they are used in common. Sometimes, there are jumps in the comparisons which correspond to the fact that with the same data the orbits are computed with different strategies and software. However, accompanied by excellent fits to the data, these results reveal a high level of orbit performance.

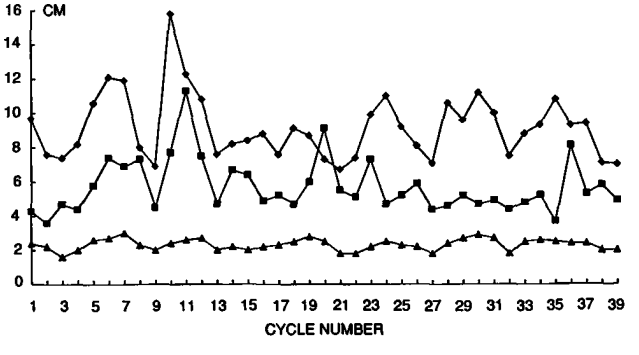


Figure 4: CNES and GSFC comparisons RMS in along-track (diamonds), cross-track (squares), and radial (triangles) components.

2.4 Precise Positioning Ground Beacons performances

Since 1990, DORIS is particularly well suited for global kinematic studies of large scale tectonic plates. 19 DORIS GLBs (figure 5) was installed on 9 major tectonic plates. Each GLBs was located using DORIS system during three years of data of SPOT 2, and one year of data of TOPEX-POSEIDON. [4].

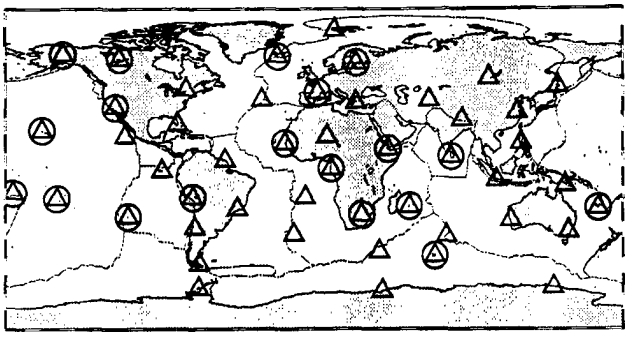


Figure 5: Map of the DORIS network. The circles indicate the 19 sites considered in this study.

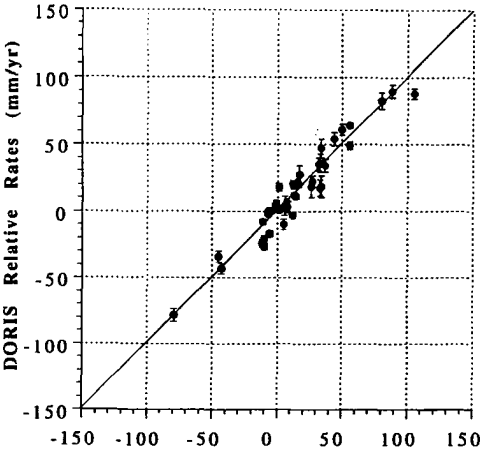


Figure 6: Interpolate velocities of the stations on the African plate relatively to the stations on the Eurasian, Pacific, North American, South American, Indian and Australian plates, computed with DORIS and compared to NUVEL-1.

L. Soudarin and A. Cazenave from GRGS-CNES have estimated the relative rates in millimeters per year of several tectonic plates.

Figure 6 from [4] shows the interplate velocities of the DORIS stations on the African plate relatively to the others plates. 40 to 50 mm per year are observed for these stations.

2.5 Real time on-board orbit determination

A real time on-board orbit determination software (DIODE) has been developed by CNES, integrated on DORIS-SPOT 4 instrument, and tested using DORIS-SPOT 2, 3 and DORIS-TOPEX measurements. For imaging purposes, SPOT required an accuracy better than 50 m, the figure reached by DIODE is less than 5 m rms on 3 axes.

This accuracy is directly connected with the complexity of the dynamical model used which is limited by the on-board processor characteristics.

3. DORIS USOs SPECIFICATIONS AND PERFORMANCES

3.1 DORIS USO Specifications

The main characteristics required for the on-board (OB) and ground USOs are defined in the following Table 1.

Parameters	OB. USO	ODB. USO	GLB. USO
Frequency	10 MHz	5 MHz	5 MHz
DC. Supply variation	$\pm 1.10^{-11}$ (1 %)	$\pm 5.10^{-11}$ (10 %)	$\pm 5.10^{-11}$ (10 %)
Load variation	$\pm 1.10^{-10}$ (10 %)	$\pm 2.10^{-11}$ (10 %)	$\pm 2.10^{-11}$ (10 %)
Pressure variation	$\pm 1.10^{-8}$/bar	$\pm 1.10^{-12}$/mb	$\pm 1.10^{-12}$/mb
Acceleration variation	$\pm 1.10^{-9}$/g	10^{-9}/g	10^{-9}/g
Vibrations (sine, random)	$\pm 5.10^{-9}$	-	-
Magnetic variation	$\pm 5.10^{-11}$/gauss	-	-
Temperature variation	$\pm 2.10^{-10}$ (air, vacuum 0 to 40°C)	$\pm 2.10^{-10}$ (0 to 60°C)	$\pm 1.10^{-10}$ (-20 to 60°C)
Short term stability $\tau = 1$ s	5.10^{-13}	5.10^{-13}	5.10^{-13}
(ALLAN variance) $\tau = 10$ s	5.10^{-13}	5.10^{-13}	5.10^{-13}
DORIS mid Term Stability			
with $\Theta = 0.1$°C/mn	0 to 40°C	0 to 60°C	- 20 to 60°C
R.M.S. SLOPE	$\pm 4.10^{-13}$/mn	$\pm 4.10^{-13}$/mn	$\pm 4.10^{-13}$/mn
Residual (1 σ)	1.10^{-12}	1.10^{-12}	1.10^{-12}
Long term Aging after 1 month			
1 day	$\pm 1.10^{-10}$	$\pm 3.10^{-11}$	$\pm 3.10^{-11}$
1 month	$\pm 1.10^{-9}$	$\pm 5.10^{-10}$	$\pm 5.10^{-10}$
Phase Noise dBc/Hz			
10 Hz	-120	-130	-130
1 kHz	-140	-145	-145
DC. Power			
air, 0°C	6 W	6 W	6 W
Vacuum, 0°C	3,5 W	-	-
Power supply	20 VDC	24 VDC	24 VDC
Random vibration	18 grms	-	-
Sine vibration	20 g - 10 to 100 Hz	2 mm 5 - 25 Hz	2 mm 5 - 25 Hz
Radiations	TOPEX	-	-

Table 1: DORIS USO SPECIFICATIONS

OB USO frequencies are 10 MHz, ODB and GLB frequencies are 5 MHz. Frequency sensitivities versus DC supply, load, pressure, acceleration are very classical. The very important characteristics of DORIS USOs are:

1) The short term stability, expressed in ALLAN variance : $\sigma y(\tau) \leq 5 \cdot 10^{-13}$ over $\tau = 10$ sec.
2) The medium term stability, defined for DORIS by slopes S (min^{-1}) and Residuals (R) at 1 sigma, of 90 samples of frequency measurements made over 10 seconds every 10 seconds. S and R are calculated by the least square method, and must be less than $4 \cdot 10^{-13}$/min and $1 \cdot 10^{-12}$ respectively. Each USO must be compliant to these specifications over all environmental conditions: temperature, magnetic field, radiation in orbit, temperature pressure, humidity on the ground.
   Temperature variations in orbit is only 1 to 3°C peak to peak due to a good thermal regulation of SPOT and TOPEX platform.
   Magnetic variations to take into account are:
   - The magnetic field variations generated by the Earth, at the altitude of the satellite (0,44 gauss peak to peak at poles and 0.22 gauss at equator).
   - The magnetic field variations generated by magnetic torquers of the spacecraft, depending of the intensity of the torquers and on the distance between USO and torquers. In the case of SPOT, the USO "see" about ± 1 gauss. In the case of TOPEX, the spacecraft magnetic variations are negligible.
3) The long term stability (aging) is not a constraint for DORIS. We need performance less than $1 \cdot 10^{-10}$/day.

3.2 DORIS USO procurement

3.2.1 The DORIS OB flight model USOs were procured from FEI (USA) for SPOT 2, then from THOMSON-CEPE (France) for SPOT 3, TOPEX, SPOT 4 and ENVISAT. FEI USOs use two oven for thermal control and Dewar technique. CEPE use one precise thermal control gigogne structure, and BVA-4 (also called QAS) space qualified swept SC crystal resonator.

3.2.2 ODB and GLB USOs (100 units, including spare models) were provided from Oscilloquartz. Models used are 8 600.03 type, using BVA AT 5 MHz crystals, two oven and glass Dewar technique. Theses models are the state of the art products, with extra screening during manufacturing to ensure they met the stringent DORIS specs.

3.3 DORIS USOs performances before integration

All OB and ODB-GLB USOs are tested in detail before integration, first at the suppliers, then at the CNES Time and Frequency Lab, and during all equipement and satellite test in near real life conditions. These included vibrations, thermal vacuum, to access performances before launching.

3.3.1 OB USOs performances

Table 2 gives typical performances, measured by suppliers and CNES.

Parameter	OB USO (CEPE)	GLB USO (OSA)
Short term $\sigma y(\tau)$ $\tau = 10$ s	$3 \cdot 10^{-13}$	$3 \cdot 10^{-13}$
DORIS medium with gradient		
Slopes	$\pm 5 \cdot 10^{-13}$/min	$\pm 5 \cdot 10^{-13}$/min
Residual	$1 \cdot 10^{-12}$	$1 \cdot 10^{-12}$
Aging 1 day	$1 \cdot 10^{-10}$/day after 40 days	$2 \cdot 10^{-11}$/day after 30 days
Phase Noise dBc/Hz		
10 Hz	-142	-142
1 kHz	-155	-155
Temperature Sensitivity	$1 \cdot 10^{-10}$ pp -10 to +50°C in vacuum	$5 \cdot 10^{-11}$ pp -20 to +55°C in air
Magnetic sensitivity	0.2 to $2 \cdot 10^{-11}$/ gauss	NA
DC power at 20°C		
in air	6.5 W	2.4 W
in vacuum	2.0 W	NA

Table 2: Typical DORIS USO performances

CEPE OB USOs are compliant in respect with DORIS medium term specs with the thermal gradient specified (0.1°C/min) figure 7. Due to very low platform temperature variations in orbit (1 to 3°C peak to peak), the thermal component of OB USO is less than $2 \cdot 10^{-13}$ /min. OB CEPE and FEI USOs were not compliant in respect with magnetic environment. The solution was to put USO inside magnetic shielding to assure medium term specs. At the same time, several studies were conducted in France by ONERA [7] and LPMO Besançon [8]. Residuals sensitivity, after replacing ferrite core inductors by air core inductors, is given by crystal mounts. This leads CNES to maintain the shielding solution for SPOT 3, SPOT 4 and ENVISAT. Figure 7 gives medium term performances of FM3 CEPE USO in thermal vacuum test.

3.3.2 ODB and GLB USOs performances

Table 2 gives typical performances of OSA 8600.03 DORIS USOs. Figure 8 gives typical medium term performances measured by CNES.

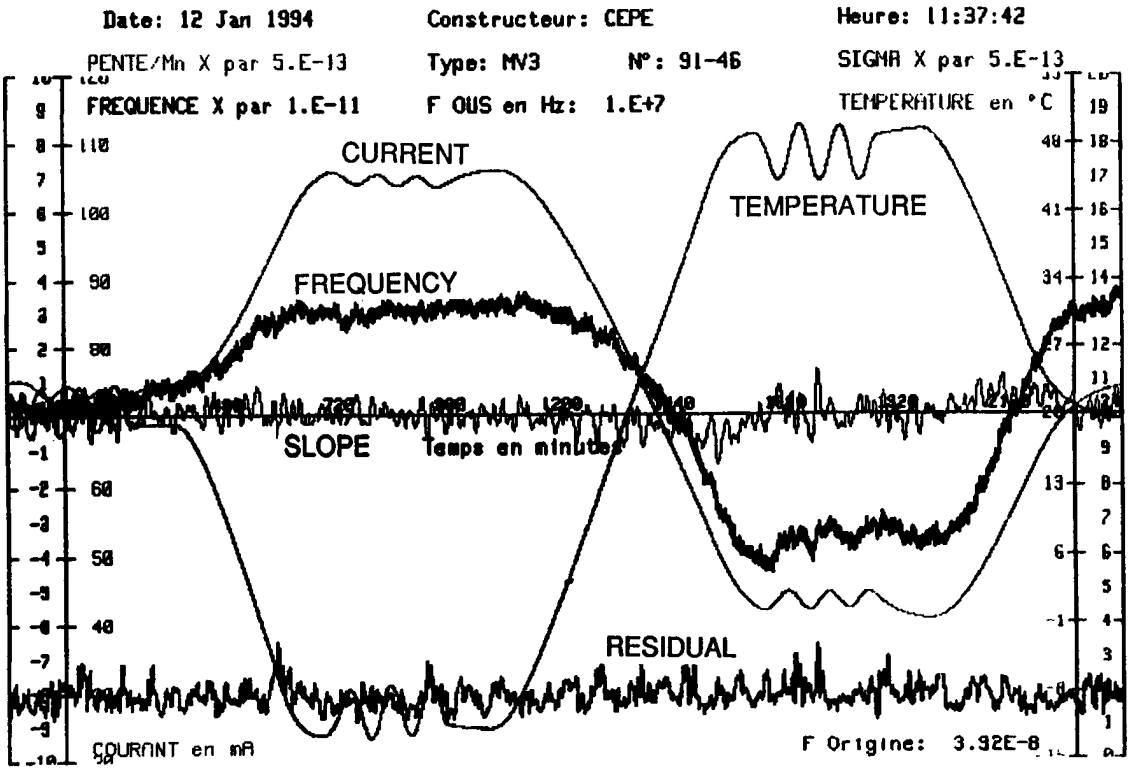


Figure 7: DORIS medium term performances of CEPE FM3 USO

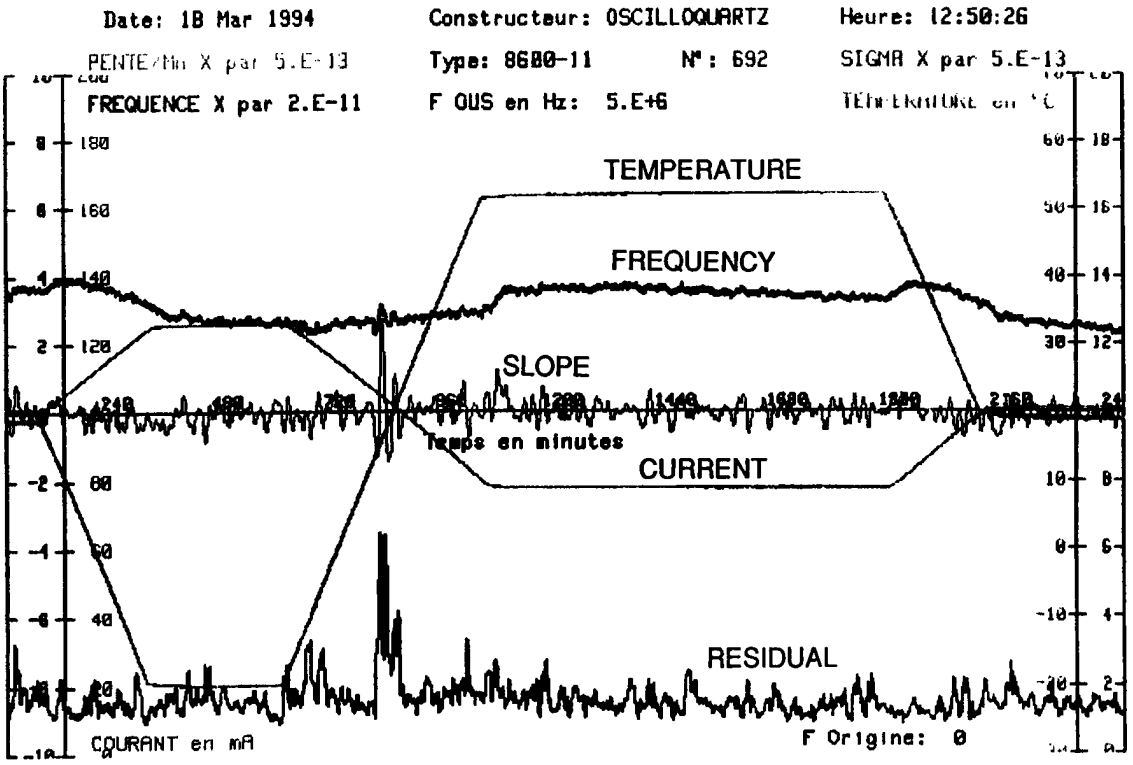


Figure 8: DORIS medium term performances of OSA GLB USO

3.4 DORIS USO performances via orbit determination

SOD processes on-board data using the ZOOM software. REF [6]. Several parameters are adjusted: orbital elements, dynamic parameters, and frequency deviation between OB and ODB, for each pass. The absolute deviations can be determined when satellites are over Master Beacons of Toulouse and Kourou, because the frequency of each beacon is known within $\pm 1 . 10^{-13}$. Using SAMOUS software, SOD calculates the absolute frequency of each ODB and GLB USOs. In order to test accuracy of SAMOUS software, one experiment was conducted during one month in France, between two separate beacons: one in Toulouse, and the other in Observatoire de la Côte d'Azur-(OCA)-Grasse, 600 km apart from each other. The two beacons were located very accurately, and the OCA DORIS ODB USO frequency was measured each 10 seconds, with an accuracy of $1 . 10^{-13}$ using an hydrogen maser as reference. 77 common views were collected, and experimental measurements were compared to the computed values. They agree within $4 . 10^{-12}$ 1 sigma.

3.4.1 In orbit DORIS USO performances

Figure 9 gives results of FEI DORIS-SPOT 2, figures 10 and 11 give results of CEPE DORIS-TOPEX and SPOT 3, summarized in the following table 3.

	SPOT 2	TOPEX	SPOT 3
After 1 month	$+1.1 . 10^{-11}$/day	$+33 . 10^{-11}$/day	$+8 . 10^{-11}$/day
After 1 year	$+6 . 10^{-12}$/day	$-45 . 10^{-12}$/day	$-21 . 10^{-11}$/day
After 2 years	$+6 . 10^{-12}$/day	$-3 . 10^{-12}$/day	-

Table 3: In orbit DORIS USO drift

3.4.2 ODB USOs performances

The drift of 50 OBD USO are between $2 . 10^{-10}$ and $10 . 10^{-10}$ per year. Figures 12 and 13 give examples of the very low aging BVA USO. The best performance is less than $1 . 10^{-10}$ per year, which is the performance of a rubidium frequency standard.

Some BVA USOs inhibits frequency jumps. Very few jumps were observed at CNES in stable conditions. In operation, frequency variations observed were not identified among several causes (Temperature variations applied to USO, internal crystal, ...). The main important point for DORIS is that these jumps can be "absorbed" by ZOOM, and do not affect, on the first order, the location accuracy.

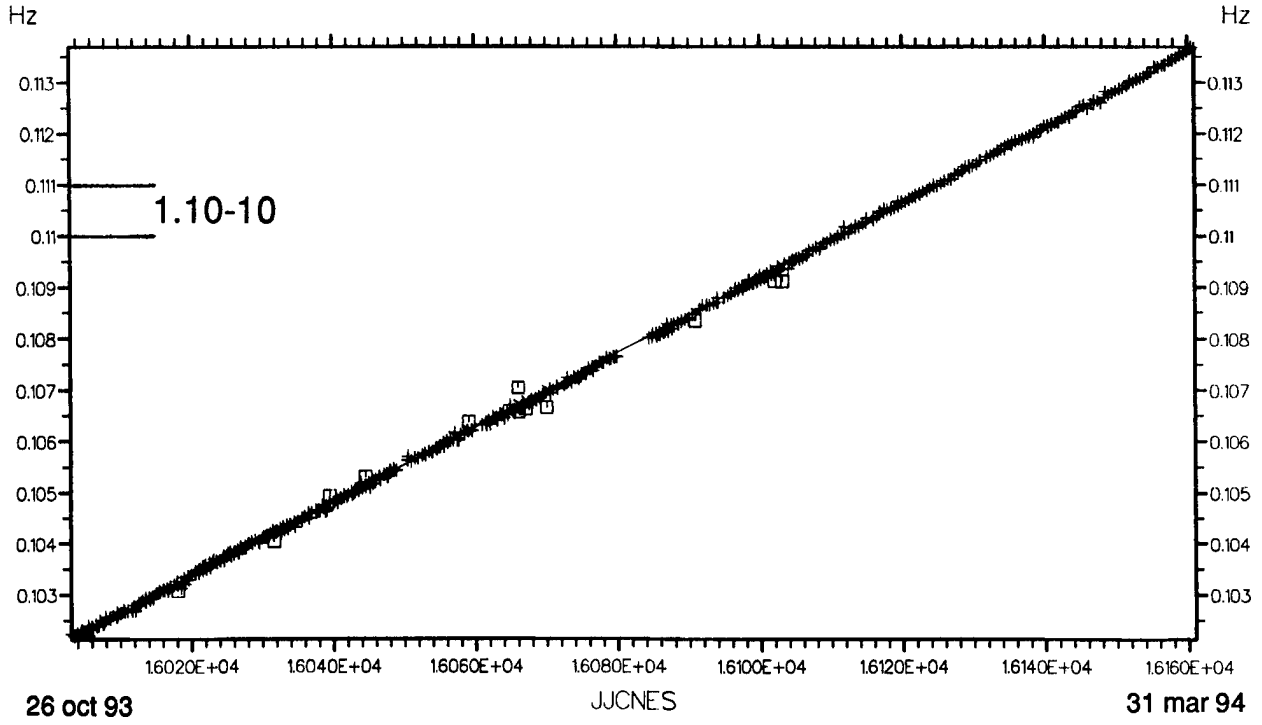


Figure 9: In orbit drift of DORIS-SPOT 2 USO

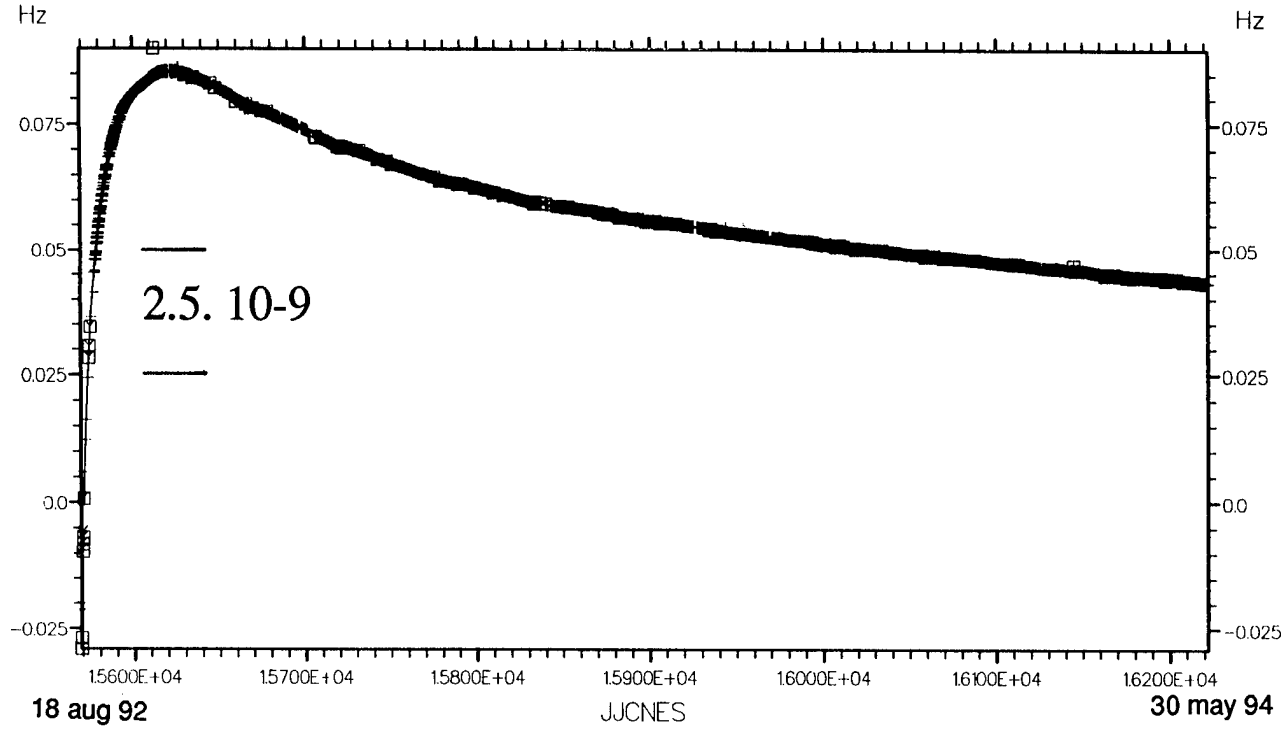


Figure 10: In orbit drift of DORIS-TOPEX USO

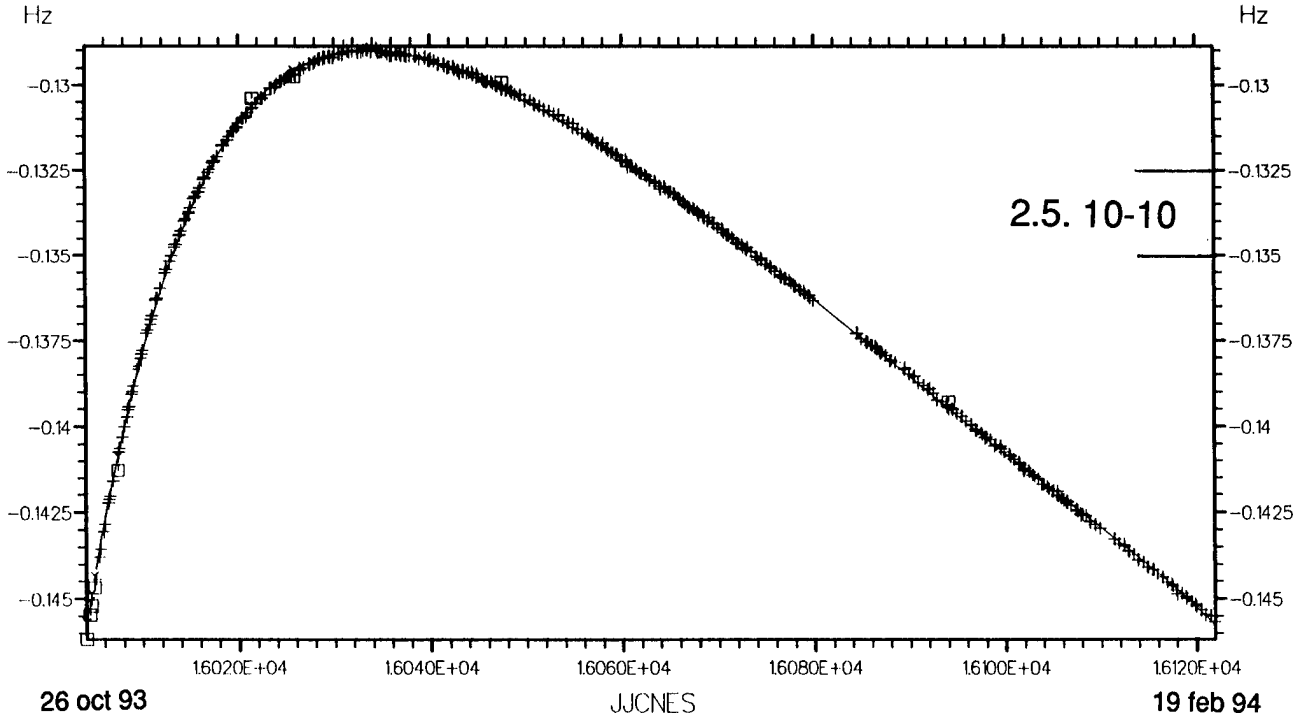


Figure 11: In orbit drift of DORIS-SPOT 3 USO

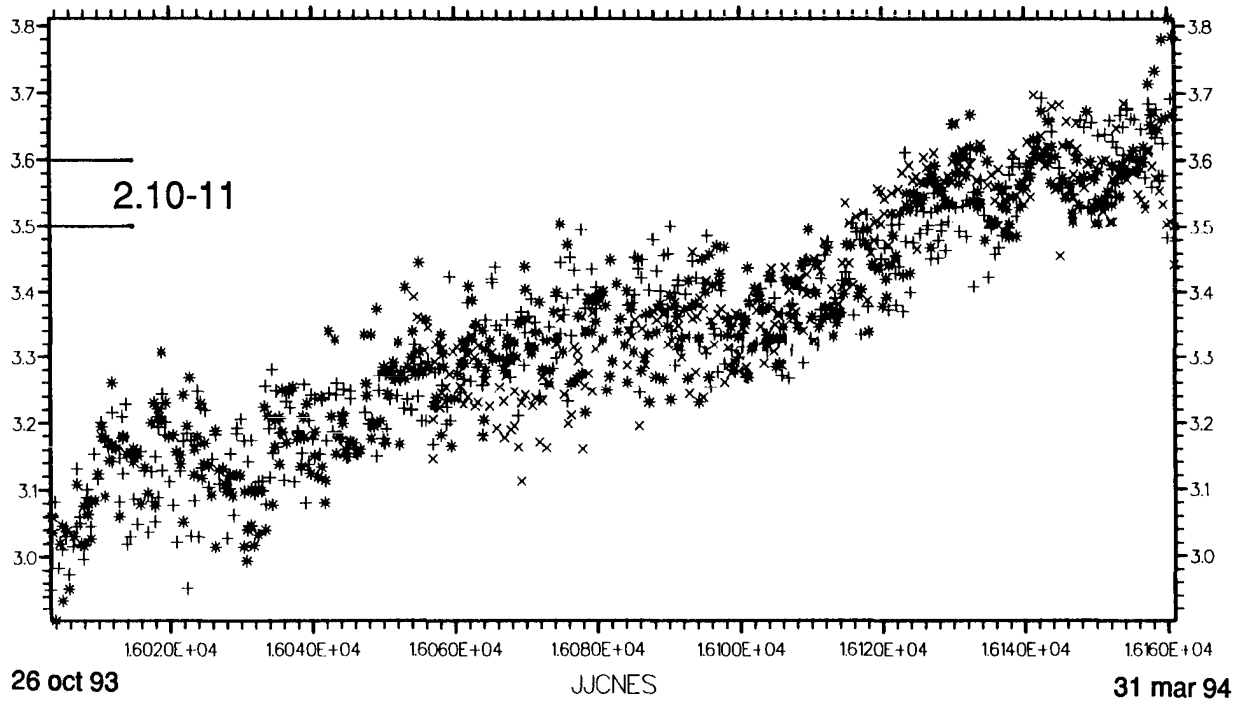


Figure 12: Drift of Hartebeesthock ODB USO

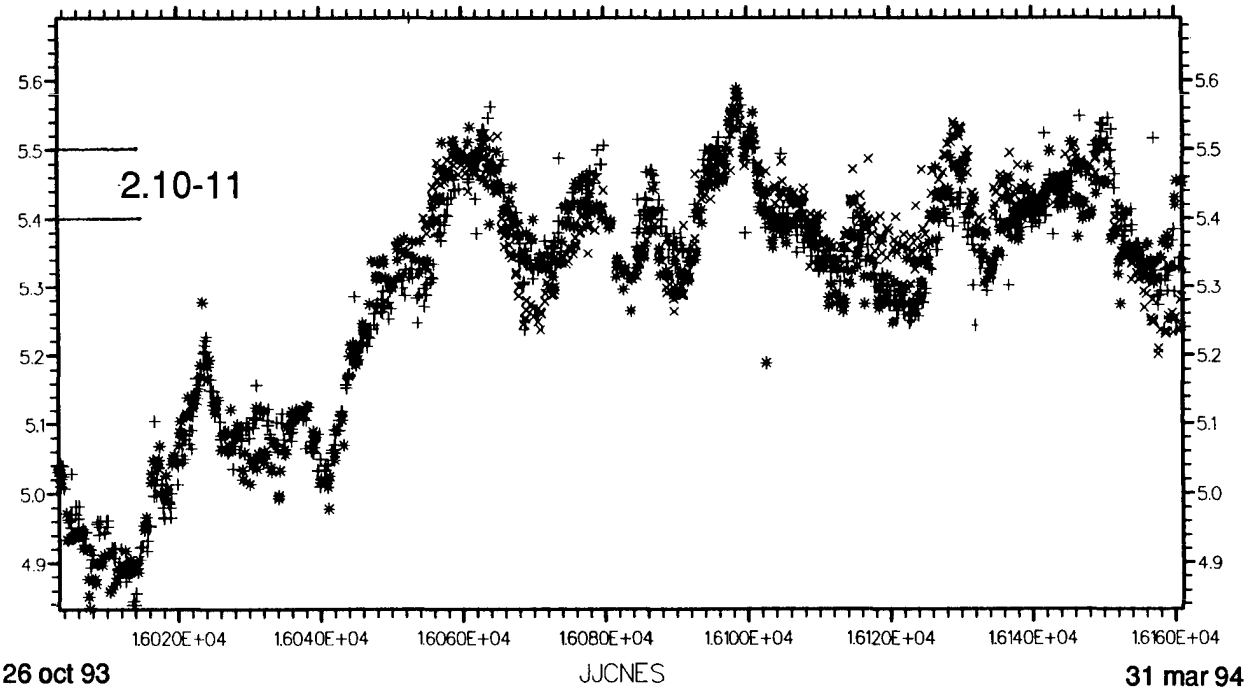


Figure 13: Drift Reykjavik ODB USO

4. CONCLUSIONS

CNES has designed and developed the DORIS system, for several applications, mainly very precise orbit determination and positioning of ground beacons. DORIS payload have been integrated on three French SPOT satellites, on TOPEX-POSEIDON NASA satellite, and will be integrated on ENVISAT ESA satellite.

After several years of orbit data, and using state of the art software, it was demonstrated that an centimetric absolute accuracy can be reached at a relatively low cost compared to other systems.

Doppler one way measurements accuracy strongly depends on the short term and medium term stability of OB and ground USOs.

DORIS uses the state of the art BVA crystals oscillators, and needs very low thermal and magnetic sensitivities. Studies are going on to reduce short term stability in the range of 10^{-14} τ = 10 s and $\tau = 10^3$ s and to decrease thermal and magnetic sensitivity too (10^{-13}/°C and 10^{-13}/gauss).

ACKNOWLEDGEMENTS

The authors are grateful to many collegues, especially A. Piuzzi, M. Deleuze, ... (Orbit Metrology Department), A. Cazenave, L. Soudarin (CNES-GRGS) and J.F. Dutrey, J.B. Laporte (Time and Frequency Department) for all results and valuable discussions.

REFERENCES

[1] M. DORRER, B. LABORDE, P. DESCHAMPS
DORIS: System Assessment results with DORIS on SPOT 2.
IAF-90-366- Dresden - Germany, October 1990.

[2] F. NOUEL et al
Precise orbit determination of SPOT platform with DORIS.
AAS-AIAA conference, paper 91-409.

[3] L. SOUDARIN, A. CAZENAVE
Global geodesy using DORIS data on SPOT 2.
Geophysical Research Letters, Vol 20, n° 4, Feb. 1993.

[4] L. SOUDARIN, A. CAZENAVE
Large scale tectonic plate motions measured with the DORIS space geodesy system.
Geophysical Research letters, Vol 22, n° 4, Feb. 1995.

[5] P. ESCUDIER, A. AURIOL
DORIS system, AAAF Int symposium, September 1993.

[6] F. NOUEL et al
Precise CNES orbits for TOPEX-POSEIDON: is reaching 2 cm still a challenge?
Journal of Geophysical Research, TOPEX-POSEIDON special issue October 1993.

[7] F. DEYZAC
Magnetic sensitivity of quartz oscillators.
4th EFTF, March 1990.

[8] R. BRENDEL et al
Magnetic sensitivity of oscillator components.
5th EFTF, March 1991.

[9] J.B. LAPORTE
Specifications and performances of ultra stable oscillators in the DORIS orbitography and positioning system.
2nd EFTF, March 1988.

[10] M. BRUNET
DORIS precise orbit determination and location system - Performances of ultra stable oscillators.
6th EFTF, March 1992.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PERFORMANCE OF GLOBAL POSITIONING SYSTEM (GPS) ON-ORBIT NAVSTAR CLOCKS

Thomas B. McCaskill
Wilson G. Reid
Orville J. Oaks
Ronald L. Beard

U.S. Naval Research Laboratory
Washington, D.C., 20375

James A. Buisson
Hugh E. Warren
SFA, Inc.
Washington, D.C.

ABSTRACT

The Naval Research Laboratory (NRL) analyses the performance of the on-orbit Navstar clocks in the Global Positioning System (GPS). This work is sponsored by the GPS Joint Program Office and is done in cooperation with the GPS Master Control Station. NRL has analyzed the performance of more than 50 Navstar atomic clocks with a total lifetime of more than 150 years. The results presented in this paper are a summary of the performance of the 25 Navstar atomic clocks that were operational during 1995.

NRL analysis of on-orbit Navstar clocks is performed using a multi-year database that includes data from initial activation of each clock through the first quarter of 1995. The Navstar pseudorange data was observed using single and dual-frequency GPS Precise Positioning Service (PPS) receivers that correct for the effects of selective availability. Quality of clock operation is calculated using both broadcast and post-processed precise ephemerides.

This presentation includes a description of the Navstar clocks that are currently operating in the GPS constellation. The Navstar clock performance in the time domain is characterized by frequency-stability profiles with sample times that range from 1 to 10-days, or more. The results presented show an improvement in the frequency stability over earlier Block I Navstar clocks with the majority of the currently operating Navstar clocks having a frequency stability for a sample time of one day of 1 $pp10^{13}$ *or better.*

BACKGROUND

The Global Positioning System is a Department of Defense (DoD) space-based navigation system that provides precise navigation, time, and frequency capabilities to users located anywhere in the world and to spacecraft in low-Earth orbits. The current constellation of 25 satellites consists of 24 Block II Navstar space vehicles (SV) and one remaining Block I SV. Each Navstar SV broadcasts time-coded signals and a GPS navigation message. The navigation message provides information that is used to compute the position of the SV and to relate the time of each of the Navstar clocks to a common GPS system time. Each GPS user passively receives the time-coded signals from four or more Navstar space vehicles and uses information contained in the signals to compute position, velocity, and time. GPS system time is steered so that a user can determine the time offset of his reference clock from Coordinated Universal Time (UTC) by using an additional correction provided in the navigation message.

The Global Positioning System is operated from the Master Control Station (MCS) located at the Combined Space Operations Center (CSOC) at Falcon Air Force Base in Colorado Springs, Colorado. The GPS control segment consists of the Master Control Station plus five GPS monitor stations located at Colorado Springs, Hawaii, Kwajalein Island, Diego Garcia Island, and Ascension Island. *Figure 1* presents the information flow from a single Navstar space vehicle to each of the GPS monitor stations.

0-7803-2500-1/95 $4.00 © 1995 IEEE

The GPS monitor stations passively track each Navstar space vehicle in view, collecting pseudorange and pseudorange-rate data from each SV. This tracking data is sent to the MCS where the Navstar ephemeris and clock parameters are estimated and predicted. The MCS periodically uploads to each Navstar SV the ephemeris and clock data which is then included in the broadcast GPS navigation message. Additional measurements—collected at the U.S. Naval Observatory (USNO) precise-time station—are referenced to the DoD Master Clock and are used for determining the steering of GPS time to UTC (USNO).

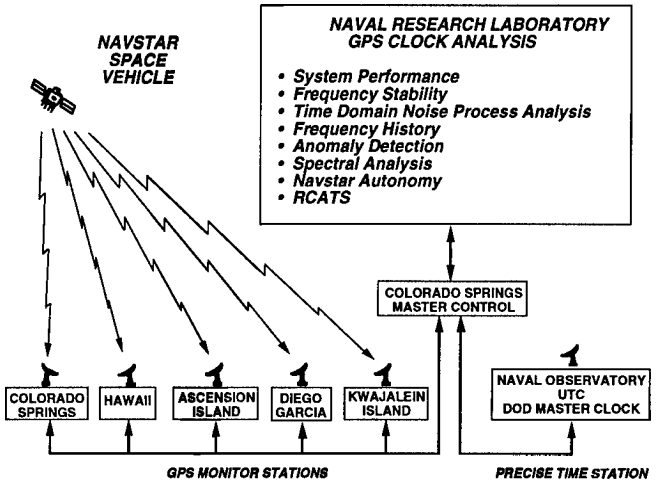


Figure 1

To appreciate the importance of the frequency stability of the Navstar clocks [1–6] consider the following. A navigation solution for a user's position and clock offset is obtained by making four simultaneous measurements of pseudorange using the precise timing signals broadcast by each of the four Navstar space vehicles. A single measurement of pseudorange taken from an unknown location results in four unknowns: the offset between the user's clock and the Navstar clock and the three position coordinates (x, y, z) of the user. The three additional simultaneous measurements result in a total of seven unknowns: the offset between the user's clock and each of the four Navstar clocks and the three position coordinates of the user. By synchronizing each measurement to a common GPS time, the number of unknowns is reduced from seven to four: the offset between the user's clock and GPS time and the three position coordinates (x, y, z) of the user. After an upload by the MCS, the burden for keeping GPS time is borne entirely by the Navstar clock until the next upload. The time prediction uncertainty of a Navstar clock is related to the frequency stability of the clock and the amount of time since the last clock update.

Other measures of performance, such as the time and frequency offset from GPS system time and the useful life span of the Navstar clocks are also considered in the analysis.

PSEUDORANGE MODEL

The pseudorange measurements are related to the time difference between the Navstar clock and the monitor station reference clock by the following equation:

$$PR = \rho + c(t_{SV} - t_{MS}) + ct_A + ct_r + e$$

where PR is the measured pseudorange, ρ is the geometric range from the Navstar at the time of transmission to the monitor station at the time of reception, c is the speed of light, t_{MS} is the time of the monitor station reference clock, t_{SV} is the time of the Navstar clock, t_A accounts for delays due to the ionosphere, troposphere, antenna, and equipment delay, t_r is a user-calculated relativity correction, and e is the pseudorange measurement error. The phase offset between a Navstar clock and the reference clock is obtained by solving the pseudorange equation for $(t_{SV} - t_{MS})$. Given two clock offset measurements, denoted by $x(t_j)$ and $x(t_k)$, sampled at times t_j and t_k, the sample time τ is defined as $\tau = (t_j - t_k)$. The average fractional frequency offset $\overline{y}(t)$ is computed by:

$$\overline{y}(t) = \frac{x(t_j) - x(t_k)}{\tau}$$

The frequency offset of a Navstar clock is computed using successive clock offset measurements separated by one sidereal day.

Data for the Navstar Block I and Block II clocks was collected at the five GPS monitor stations using dual-frequency, PPS receivers which measure the ionospheric delay and automatically correct for the effects of selective availability. Data collected at the USNO precise-time station used a single-frequency, PPS receiver with ionospheric corrections derived from a model of the ionosphere using coefficients obtained from the GPS navigation message. The precise, post-processed ephemeris was used to obtain the monitor station clock offset, and the broadcast ephemeris was used to obtain the offset of the DoD master clock at USNO.

STABILITY MODEL

The frequency stability of a clock is defined as an infinite time average [7]. For a finite number of uniformly sampled frequency-offset measurements the stability is estimated using the following equation:

$$\sigma_y^2(\tau) \simeq \frac{1}{2(M-1)} \sum_{k=1}^{M-1} (\overline{y}_{k+1} - \overline{y}_k)^2$$

The average frequency offset is computed from pairs of clock offset measurements separated by a sample time τ. The square root of this two-sample variance, $\sigma_y(\tau)$, is known as the frequency stability.

This time-domain, frequency-stability measure is independent of initial clock offset and initial frequency offset. The response to a periodic signal varies with sample time but is invariant to a periodic signal sampled at multiples of the fundamental period. Hence, the two-sample variance for a sample time of one day is invariant to any residual periodic error resulting from the GPS orbit which has a nominal period of half a sidereal day.

CURRENT CONSTELLATION

Table 1 presents the location of the Navstar space vehicles in the GPS constellation at the end of the first quarter of 1995. The constellation consists of 24 space vehicles located in the six planes A–F with four space vehicles per plane in slots 1–4. Two types of atomic clocks—rubidium and cesium—are currently in use on the Navstar space vehicles. In *Table 1* the currently operating clock on all of the Navstar space vehicles is cesium except for those SV numbers which have been circled indicating that the currently operating clock is rubidium.

Table 1

GPS CONSTELLATION
SPACE VEHICLE LOCATION

Plane	Slot 1	Slot 2	Slot 3	Slot 4
A	39	25	27	19*
B	22	20*	13	35
C	36*	28	31*	37
D	24*	15	17	34
E	14	21	16	23
F	32*	26	18	29

* RUBIDIUM CLOCKS
ALL OTHER CLOCKS CESIUM

As of 31 March 1995

Figure 2 shows for each of the space vehicles the number of clocks that have been placed in operation since the space vehicle was inserted into the constellation. The only remaining Block I clock currently in operation is the Navstar 10 rubidium frequency standard. All of the Block II clocks in operation at the end of the quarter were cesium frequency standards except for the rubidium frequency standards on Navstars 19, 20, 24, 31, 32, and 36. Seven of the Block II space vehicles have had two of their four clocks activated. Four of the Block II space vehicles have had three of the four clocks activated. And one of the Block II space vehicles—Navstar 19—has had all four clocks activated.

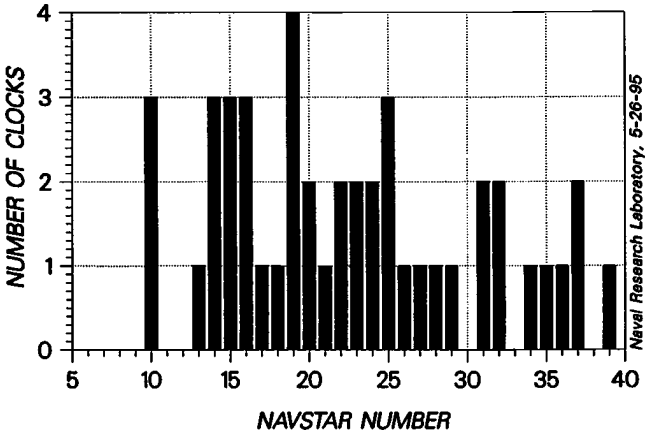


Figure 2

Figure 3 shows the operating lifetime, or length of service, in years of the clocks that were operating at the end of the quarter. The rubidium clock on Navstar 19 was switched to the second rubidium clock after only 75 days of operation. The cesium clock on Navstar 24 was switched to the first rubidium clock after 658 days of operation. The cesium clock on Navstar 32 was switched to the first rubidium clock after 833 days of operation.

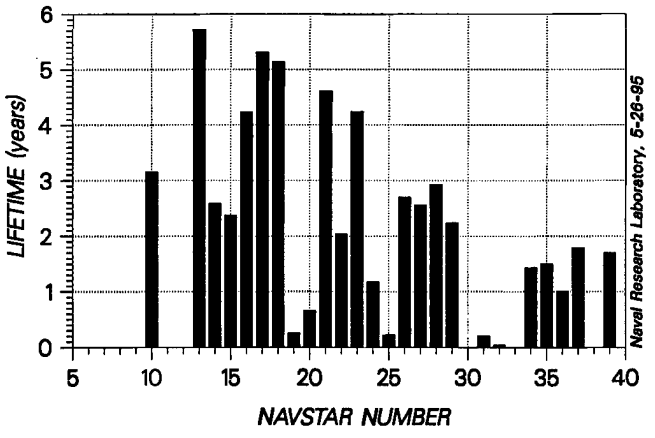


Figure 3

Figures 4–9 show the frequency standard operating schedule for planes A-F for each of the clocks in the constellation. *Figure 4* shows Navstar 19 to be on its last clock which was activated 30 December 1994.

FREQUENCY STANDARD OPERATING SCHEDULE
PLANE A

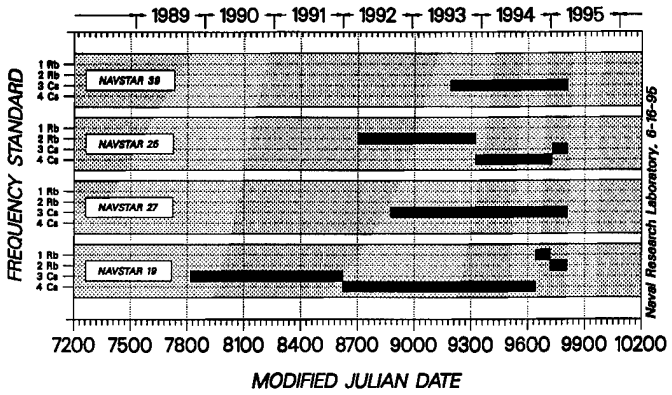


Figure 4

FREQUENCY STANDARD OPERATING SCHEDULE
PLANE B

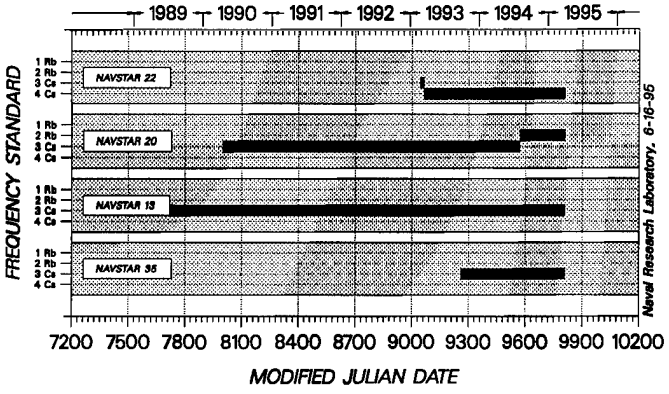


Figure 5

FREQUENCY STANDARD OPERATING SCHEDULE
PLANE C

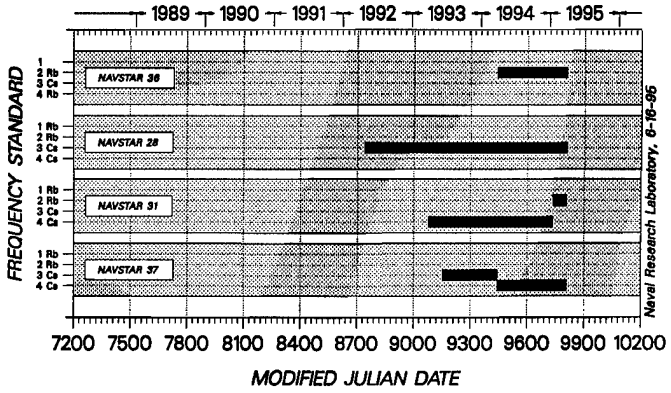


Figure 6

FREQUENCY STANDARD OPERATING SCHEDULE
PLANE D

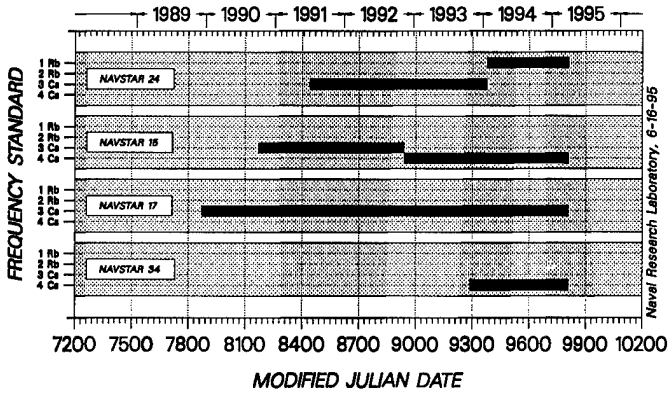


Figure 7

FREQUENCY STANDARD OPERATING SCHEDULE
PLANE E

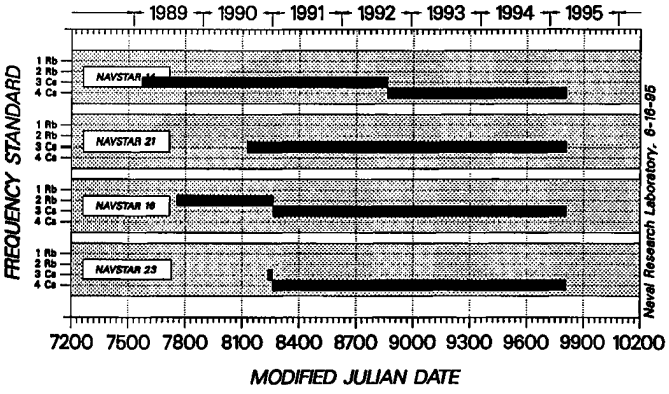


Figure 8

FREQUENCY STANDARD OPERATING SCHEDULE
PLANE F

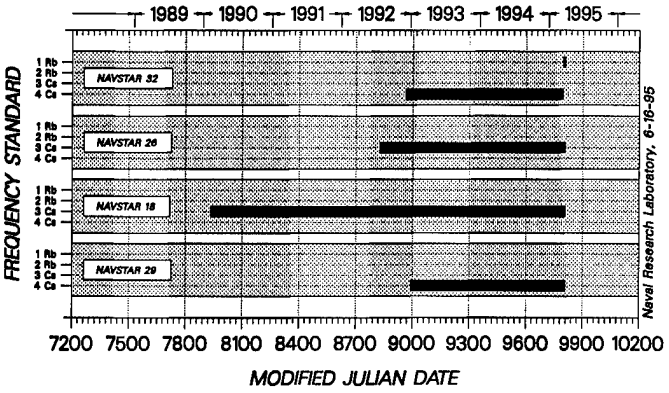


Figure 9

Figure 10 presents the frequency offset from UTC (USNO) in $pp10^{11}$ for each of the clocks that were operating at the end of the quarter. The six clocks, including the Block I Navstar 10 clock, having the largest frequency offset were all rubidium frequency standards.

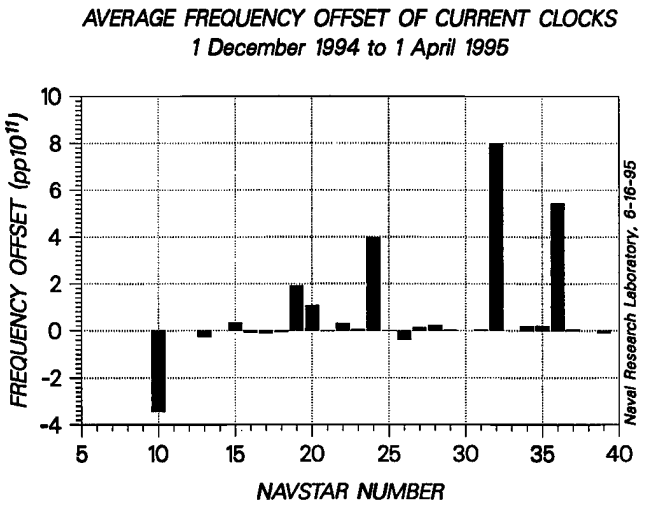


Figure 10

Figure 11 is a plot on an expanded scale ($pp10^{12}$) of the frequency of the cesium clocks and shows the maximum frequency offset of the cesium clocks with respect to the DoD master clock at the Naval Observatory to be about -3.8 $pp10^{12}$. Ten of the cesium clocks showed a positive offset while the other eight showed a negative offset. The Navstar 14 cesium clock had a negative frequency offset which was too small to be seen on the plot.

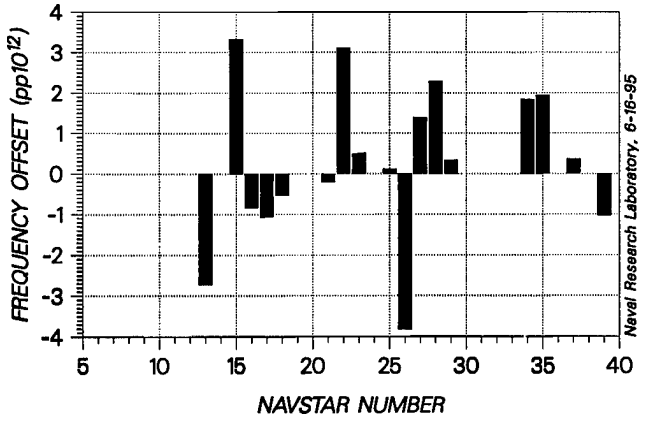


Figure 11

The data in *Figure 11* can be used to check the adequacy of the first order relativistic correction [8] introduced by the frequency synthesizer in the space vehicle. The mean of the distribution of the frequency offset for on-orbit cesium clocks based on this sample of 18 clocks was estimated to be $+2.8$ $pp10^{14}$. The standard deviation of the population was similarly estimated to be 4.5 $pp10^{13}$. The relatively small value of the mean offset validates the relativistic correction to the frequency offset so that the space vehicle clocks will run at nominally the same rate as UTC.

Figure 12 presents the linear drift for each of the clocks in the constellation and shows that the rubidium clocks exhibit the typically large drift rates characteristic of rubidium frequency standards—the largest being Navstar 19 with a drift of -4.6 $pp10^{13}$ per day. The drift for the Navstar 32 rubidium clock was not calculated because insufficient data was available to obtain a meaningful estimate. All of the rubidium clocks exhibit negative drift.

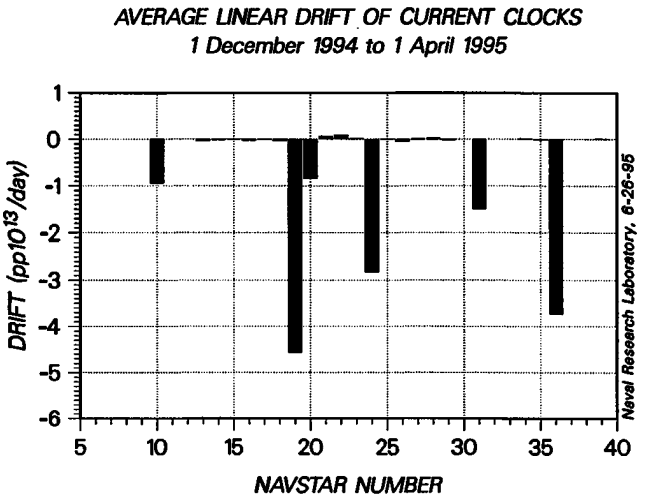


Figure 12

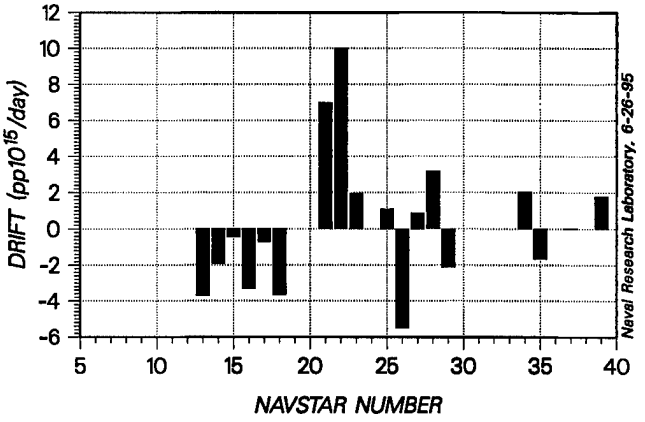


Figure 13

Figure 13 is a plot on an expanded scale of the linear drift of the cesium clocks. The cesium clocks are evenly divided between positive and negative drift. All but three of the cesium clocks exhibit a drift with a magnitude below 5 $pp10^{15}$ per day which is about two orders of magnitude less than the largest drift rate reported for the Block II rubidium clocks.

Figure 14 presents estimates of the frequency stability for each of the Navstar clocks for a sample time of one day using the broadcast ephemeris. Superimposed on *Figure 14* are dashed lines corresponding to the GPS system specification of 2 $pp10^{13}$ and 5 $pp10^{13}$ for the one-day stability of the cesium and rubidium clocks.

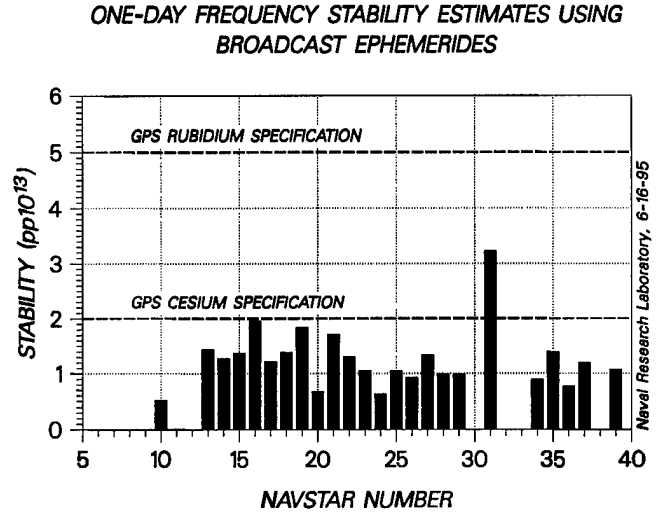


Figure 14

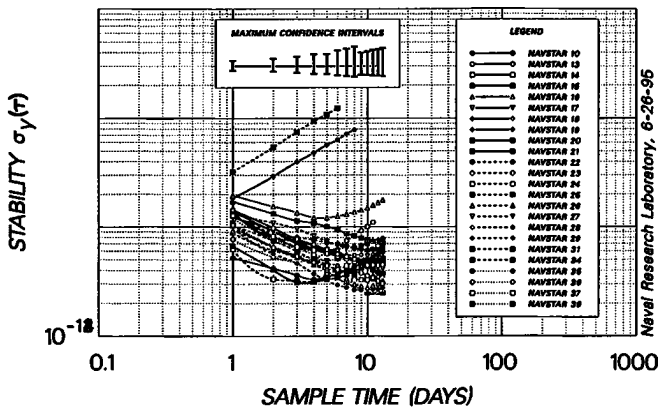


Figure 15

The clocks on Navstars 10, 19, 20, 24, 31, 32, and 36 were rubidium. The frequency stability of all Block II space vehicle clocks can be seen to meet the specifications. Seven of the Block II clocks show a stability at or below 1 $pp10^{13}$. Three of these are the rubidium clocks on Navstars 20, 24, and 36 while the other four are the cesium clocks on Navstars 26, 28, 29, and 34. The Navstar 26 and 28 cesium clocks were built by Frequency & Time Systems, Inc., while Navstars 29 and 34 carry the alternate-source cesium clocks built by Kernco.

Figure 15 is a plot of the frequency stability profile corrected for linear drift for the 120 days covered by this report for 24 of the 25 clocks in the constellation using the broadcast ephemeris. The exception is the Navstar 32 rubidium clock for which insufficient data was available to calculate a meaningful estimate. All but two of the clocks—the rubidium frequency standards on Navstar 19 and Navstar 31—are evenly distributed between 2 $pp10^{14}$ and 2 $pp10^{13}$ at 12 days. The stability of these two exceptions rapidly degrades after one day because of the large random walk frequency noise that is typical of rubidium clocks.

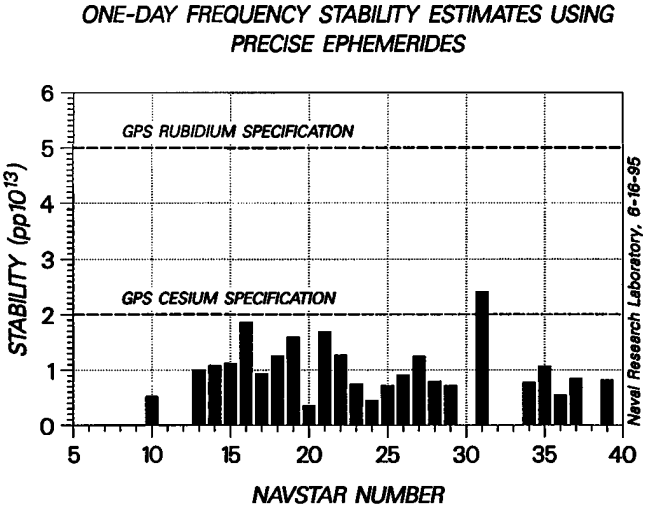


Figure 16

Figure 16 presents the minimum estimates of the frequency stability for each of the Navstar clocks for a sample time of one day using data from the five GPS monitor stations and the precise ephemerides. The stability for the Navstar 32 rubidium clock is not shown because insufficient data was available to calculate a meaningful estimate. In the case of Navstar 36, only the data preceding some erratic behavior which would later precipitate its deactivation was used in the calculation. Superimposed on the figure are dashed lines corresponding to the GPS system specification of 2 $pp10^{13}$ and 5 $pp10^{13}$ for the one-day stability of the cesium and rubidium clocks. With the use of the precise ephemerides 14 of the 25 operational clocks in the constellation are now estimated to have a frequency stability for a sample time of one day at or below 1 $pp10^{13}$.

CONCLUSIONS

Analysis covered the full constellation of 24 operational Block II space vehicles and one remaining Block I space vehicle for the first quarter of 1995. The frequency stability was estimated for each of 24 clocks using both the broadcast and the precise ephemerides. Table 2 presents a summary of the performance.

Table 2

PERFORMANCE OF GPS BLOCK I AND BLOCK II NAVSTAR CLOCKS
1 December 1994 to 1 April 1995

Navstar Number	Clock Type	Frequency Stability 1 day $pp10^{13}$	Frequency Stability 10 days $pp10^{14}$	Frequency Drift $pp10^{15}/day$
10	Rb	0.5	5.0	−95.2
13	Cs	1.0	4.5	−3.7
14	Cs	1.1	4.5	−1.9
15	Cs	1.1	4.3	−0.4
16	Cs	1.9	15.0	−3.3
17	Cs	0.9	4.5	−0.7
18	Cs	1.3	5.5	−3.7
19	Rb	1.6		−457.0
20	Rb	0.4	2.7	−86.7
21	Cs	1.7	7.4	9.1
22	Cs	1.3	6.4	10.0
23	Cs	0.8	3.6	+1.7
24	Rb	0.5	5.9	−283.0
25	Cs	0.7		−13.0
26	Cs	0.9	2.8	−5.5
27	Cs	1.3	5.8	+0.9
28	Cs	0.8	3.5	+4.2
29	Cs	0.7	2.7	−2.1
31	Rb	2.4		−150.0
34	Cs	0.8	2.5	2.0
35	Cs	1.1	6.9	−1.7
36	Rb	0.7	66.7	−253.1
37	Cs	0.9	3.3	+0.0
39	Cs	0.8	5.0	+1.8

An average of less than two Block II Navstar clocks have been activated indicating that an average of at least two spare clocks are available to complete the design lifetime of the Block II space vehicles.

The average operating time of the Block II clocks that have been activated is 2.36 years.

The maximum frequency offset of any Navstar atomic clock during the first quarter of 1995 was 8 $pp10^{11}$ which is almost two orders of magnitude smaller than the maximum frequency offset that can be accommodated by the navigation message.

The mean frequency offset of the on-orbit cesium clocks was estimated to be 2.8 $pp10^{14}$ based on a sample of 18 clocks.

Fourteen Navstar clocks—more than half of those currently operating—have a one-day frequency stability estimated at or below 1 $pp10^{13}$ as measured by the five GPS monitor stations using the precise ephemerides.

REFERENCES

[1] McCaskill, Thomas B., and Buisson, James A., "On-Orbit Frequency Stability Analysis of GPS Clocks and the Importance of Frequency Stability to Precise Positioning", *First International Symposium on Precise Positioning with the Global Positioning System*, 15 April 1985, Vol. 1, pp. 37–50

[2] McCaskill, T.B. and Buisson, J.A., "NTS-1 (TIMATION III) Satellite Quartz and Rubidium Oscillator Frequency Stability Results", *Proceedings of the 29th Annual Frequency Control Conference*, 28–30 May 1975, pages 425–435

[3] McCaskill, Thomas B., Buisson, James A., and Reid, W.G., U.S. Naval Research Laboratory, and Warren, Hugh E., Sachs Freeman Associates, "Analysis of the Frequency Stability of On-Orbit GPS NAVSTAR Clocks", *1993 IEEE International Frequency Control Symposium*, 2–4 June 1993, pp. 23–32

[4] Reid, W.G., McCaskill, T.B., Oaks, O.J., U.S. Naval Research Laboratory, and Buisson, J.A., and Warren, H.E, "Effect of Space Vehicle Temperature on the Frequency of On-Orbit NAVSTAR Clocks", *1994 IEEE International Frequency Control Symposium*, 1–3 June 1994, pp. 782–790

[5] McCaskill, T.B., Reid, W.G., Buisson, J.A., and Warren, H.E, "Effect of Broadcast and Precise Ephemerides On Estimates of the Frequency Stability of GPS NAVSTAR Clocks", *International Journal of Satellite Communications - Special Issue On Navigation and Positioning*, September–October 1994, pp. 435–442

[6] Barnes, J.A., *et al.*, "Characterization of Frequency Stability", *IEEE Transactions on Instrumentation and Measurements*, IM-20, 2, May 1971, pp. 105–120

[7] Allan, D.W., "Time and Frequency (Time Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators", *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, UFFC-34, No. 6, November 1987, pp. 647–654

[8] Easton, R.L., Buisson, J.A., and McCaskill, T.B., "Initial Results of the NAVSTAR GPS NTS-2 Satellite", NRL Report 8232, 25 May 1978

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ACTIVITIES OF THE RUSSIAN INSTITUTE OF RADIONAVIGATION AND TIME IN DEVELOPMENT OF GLONASS ON-BOARD Q-ENHANCED OSCILLATING COMPACT H-MASER

Yu.G.Gouzhva, A.G.Gevorkyan, A.L.Myasnikov, G.M.Kryukov, V.Z.Gankin, O.N.Kornishov, S.D.Teplova

The Russian Institute of Radionavigation and Time, 2 Rastrelli sq., 193124 St.Petersburg, Russia

Abstract

The paper describes a new type of on-board atomic frequency standard developed by the Russian Institute of Radionavigation and Time (RIRT). This is Q-enhanced oscillating compact H-maser with dimensions and weight less than those of GLONASS Cs clock. The multifactor optimization of maser's main building block construction and physical parameters has been carried out, which enabled one to lower weight of H-maser physical package down to 10 kg. The forecast enhanced performance has been obtained and attainable daily frequency stability of the H-maser comprised 10^{-14}. The latter is by an order of magnitude higher than that of Cs clock. Reduced dimension of developed maser will allow to substitute Cs clock by H-clock in GLONASS satellite and will provide progress in this field.

Introduction

RIRT activities in development of compact H-maser have been stimulated by a desire to utilize the high stability and accuracy of conventional H-maser in application with severe size and weight constraints.

It is of great importance for improving accuracy of mutual synchronization satellite time for GLONASS signal (it has to be no worse than 1-2 ns for 24 hours) as well as for creating a future European civil navigation system.

Although several prototypes of H-masers have been developed [1] and have been adapted for various future space applications (such as VLBI, high precision time transfer, space tracing, global time synchronization and the testing relativistic gravitation) only NASA/SAO Gravity Probe A at suborbital flight has carried a high-stable H-maser into space [2].

These masers are of very high performance but they are too massive because of use of full size cavity with multilayer shielding.

Analysis has shown that significant decreasing of H-maser dimensions along with providing high metrological performance has been achieved by cavity volume diminishing by inducing into the cavity a dielectric load formed by a wall of the storage bulb made from leucosaphire [3,4].

Another way is a use of magnetron type TE_{011} cavity with ring cutting metal field concentrators. In this case cavity can be produced at any minimization level (see, for example, [5,6]) as far as the volume of the storage bulb makes it possible to provide reasonable high Q_l of spectral line, radiation power P and, as a result, high frequency stability.

These cavities have Q-factor significantly lower than self-excitation threshold, therefore it is used either in passive H-maser devices or in active ones with the aim to enhance its Q. Construction of the magnetron type cavity ensures high mechanical rigidity, reliability and good filling factor value.

A leucosaphire storage bulb 900 sm^3 has been used in our own first design of an engineering model of space version in 1988 [4]. It has been tested during a year and demonstrated a frequency stability near $1.5 \cdot 10^{-14}$ for time interval 10^3-10^5 s over ambient temperature $20 \pm 3^{\circ}C$. This maser and its electronic system satisfied GLONASS requirements with exception of its volume (75 l) and weight (65 kg). Despite its excellent results in passing all kind of ground test the model had some

0-7803-2500-1/95 $4.00 © 1995 IEEE

disadvantages, namely, a large volume of cavity temperature coefficient, high cost leucosaphire single crystals and lack of technological effectiveness for serial production as well.

While taking into account a level of development, cheapness, compactness, working stability of regeneration, cost-saving, reliability and simple realization of H-maser electronics compared to passive one, RIRT gave preference to developing the project of oscillating H-masers with Q-enhanced magnetron type cavity.

The main goal of this work is creating Q-enhanced compact H-maser for GLONASS and establishing the influence of perturbation of its parameters.

Optimization

For providing H-maser prescribed performance, the multi-factor optimization of maser's main building block construction and physical parameters has been carried out by mathematical simulation. The cavity size (D_c, L_c), result coupling cavity with regenerator coefficient (β_1,β_2) and noise factor of its input amplifier F, as well as maximum permissible Q-enhanced cavity Q_r have been taken as the initial data. The size of the construction elements of cavity-storage bulb, emission line quality Q_l, radiation power P, a coefficient of input and output coupling the cavity with regenerator β_1, β_2, the noise temperature of regenerated cavity T_N, filling factor η, an excitation parameter S et al are calculated.

The minimum frequency dispersion of the atomic hydrogen emission line in the cavity $\sigma^2(\tau)$ (with taking into account the cavity noise over short observation time intervals τ) has been accepted as a criterion of optimization. The influence of noise on $\sigma^2(\tau)$ is determined in accordance with [6] as

$$\sigma^2(\tau) = \frac{k T_N}{Q_l^2 2P\tau} + \frac{Fk T_a(1+\beta_1+\beta_2)+\beta_1 k T_N}{\omega^2 \tau^2 \beta_1 P} \cdot B, \quad (1)$$

where k is the Boltzman constant, T_a is the temperature of input amplifier of regenerator, ω is the maser frequency, B is the frequencies band.

$$T_N = T_c[1+\sqrt{(T_a/T_c)(1+\beta_1+\beta_2)/\beta_1}\ (1-1/J)]^2,$$

$$T_a = T_c (F-1), \quad J = Q_r/Q.$$

T_c is the cavity temperature, J is the regeneration coefficient, Q is loaded Q-factor of cavity. The second term in (1) describing the additive noises of amplifier and regeneretor cavity into $\sigma^2(\tau)$ can be neglected because of the fast rate of its decrease with time.

The mathematical maser model takes into account the dependence of Q_r and η on sizes of storage bulb, cavity and its electrodes, that of Q_l on the rate of hydrogen atom escape from the bulb r_b, on the relaxation rates of atoms on the bulb surface and on spin-spin exchange as well as the dependence T_N on T_c, Q_r, β_1, β_2, F and J. Monte-Carlo method has been used. The region of determination of optimized maser parameters has been made more exact after multiple ranging and analyzing of the massive of the calculated data σ_i. As the result the range of permissible size of cavity-storage bulb has been established.

In Fig. 1 the dependence of calculated minimum magnitudes σ_i of varying cavity sizes is given. For further studies the cavity sizes not higher than 75 mm at D_c=L_c have been chosen. Fig. 2 shows the changes in parameters η, Q_c and S on sizes of cavity electrodes. Note, that optimum value of S corresponds to the chosen sizes. The region of optimization for r_b is presented in Fig. 3. The curves are plotted on the base of the analysis [7] of dependencies P, S, Q_l on r_b for definite magnitudes of hydrogen atoms flow into the bulb, size of beam generator and bulb input hole taking into account. The figures and points on the curves correspond to maximum values of $P/10^{-14}$ W, $Q_l/10^8$ and minimum values of $\sigma_{1s}/10^{-12}$. The magnitudes of flow atoms into a bulb $I/10^{11}s^{-1}$ are indicated in the parentheses at optimum chosen values of r_b for flow from source of H-beam I_s. The length of a tube is taken as 25 mm. The upper curve corresponds to r_b optimizing by maximum signal/noise relation. The lower curve corresponds to r_b optimizing by minimum of pulling factor Q_r/Q_l. The most favorable is optimization by minimum σ_{1s} (middle curve). At the working flow into the bulb $I=(4\text{-}6)\cdot10^{11}$ s^{-1} the optimum value of r_b comprises (2-3) s^{-1}, $Q_l=5\cdot10^8$. The prescribed values of σ_{1s}, P, Q, T_N versus J and β_1 are given in Fig. 4,5. In reality, the radiation power P grows with increase J, but frequency stability σ_{1s} drops because of the influence of regenerator circuit (Fig. 4, dashed line). With the increase of β_1 the noise temperature slightly decreases, σ_{1s} reaches $1\cdot10^{-12}$. The physical meaning of decreasing T_N with the increase of β_1 is in redistribution of noise power and signal power in favour of a signal in the regeneration process. Therefore, it is reasonable to choose β_1 in a

range from 0.4 up to 0.6. In this case 30% of power goes to a receiver and T_N lowers down to 1700 K. The curves in Fig. 2-5 are plotted for the followed initial data: Q is 3500, bulb volume is 80 sm^3, cavity volume is 330 sm^3, r_b is 3 s^{-1}, I is $5 \cdot 10^{11}$ s^{-1}, β_1 is 0.5, β_2 is 0.05, η is 1.2, F is 2.3 dB, P is $5 \cdot 10^{-14}$ W, Q_r is 60 000, Q_l is $6 \cdot 10^8$, T_N is 1700 K, σ_{1s} is $(7\text{-}10) \cdot 10^{-13}$. On the base of the prognosis characteristics of generator the recommendation for construction of H-maser main block have been chosen.

Minimization of regenerator modes instability

The cavity auto-tuning system functioning through a response of emission signal to cavity frequency modulation cannot keep up in full with relatively fast frequency shifts caused by fast oscillations of on-board power supply and temperature changes. Therefore one needs to clear up an influence of external disturbances and its decreasing to the level of $5 \cdot 10^{-15}$ on ω_r resonant frequency and Q_r of regenerated cavity using its dependences on G and Θ, regeneration circuit parameters:

$$\omega_r = \omega_0 [1-(1-G\sin\Theta)/(2Q)],$$

$$Q_r = Q/(1-G\cos\Theta),$$

where ω_0 is the non-regenerated cavity intrinsic frequency, Θ is the phase lag angle ($\Theta = \omega t_l + \psi$), ω is the H-maser emission signal frequency, t_l is the lag time, ψ is the phase shift, $G=1-1/J$ is the module of regeneration circuit transfer coefficient, G is and Θ is the value fluctuations under external disturbances give rise to deviations of cavity frequency and Q_r.

$$d\omega_r = \omega_0 (G \cdot \cos\Theta \cdot d\Theta + \sin\Theta \cdot dG/(2Q),$$

$$dQ_r = J^2 Q(\cos\Theta \cdot dG - G \cdot \sin\Theta \cdot d\Theta).$$

A change of phase incursion $\Delta\Theta$ and of transfer coefficient ΔG depending on changes of temperature ΔT and supply voltage ΔU can be expressed as follows: $\Delta\Theta = \alpha\Delta U + \beta\Delta T$ and $\Delta G = \gamma\Delta U + \xi\Delta T$, where $\alpha, \beta, \gamma, \xi$ are values of coefficients experimentally determined: α=0.53 rad/V, $\beta = 5 \cdot 10^{-3}$ rad/°C, γ=1.1 1/V, $\xi = 2.3 \cdot 10^{-2}$ 1/°C. While presetting a tolerable Allan variance at the level of $5 \cdot 10^{-15}$ and taking into account a pulling coefficient of $2 \cdot 10^{-4}$, G=0.95, a phase tuning error of regeneration circuit ±3%, what is in accordance with $\sin\Theta = \pm 0.05$, then tolerable changes of power supply voltage and temperature are $\Delta U = 1 \cdot 10^{-7}$ V and $\Delta T = 1 \cdot 10^{-5}$ °C. The computations show that, for a tolerable cavity Q instability, the values of ΔU and $\Delta\Theta$ are less stringent in one-two orders compared to requirements for cavity frequency instability. In order to achieve a power supply voltage instability of regenerator up to 0.1 μV when on-board main changes over a range of ±10%, two voltage stabilizers are installed in series, and the second stabilizer is placed within a regenerator ovenized zone. Besides, it is necessary to decrease an influence of regenerator mode instability by means of reduction of redundancy for phase and amplification regulation ranges in one order of value, that is to diminish a regulation slope. In order to avoid a situation when the other resonances within a cavity (but not the operating resonance) could cause an instable regenerator operation mode at the operating resonance, the cavity is realized as a frequency-selective amplifier with a bandpass of 10-20 MHz at the hydrogen emission frequency. Frequency selectivity aids to the rejection of non-artificial interferences. The regenerator amplifies a hydrogen emission signal with a gain of Jg, where g is module of regenerator gain in power. The regenerator is sensitive to external electromagnetic interferences and, in particular, to interferences caused by its own amplified signal penetrating into a cavity through non-controlled propagation paths. The level of this signal should be lower than a thermal noise level within a bandpass of emission line. The degree of filtering and shielding (D) should be:

$$D = \pi k T_N/(r_2 J g P),$$

where P is the emission power; r_2 is the rate of atomic oscillating magnetization relaxation. When Jg=60 dB, $P = 5 \cdot 10^{-14}$ W, r_2=3 s^{-1}, T_N=2000 K, the degree of filtering and shielding should be minus 120 dB. This condition is practically realized using ring disk capacitors within control/power supply circuits, as well as multilayer shielding of flexible microwave cables.

Decrease of magnetic relaxations and frequency shifts by means of Q-enhanced

Emission line Q_l, frequency shift of operation transition δ and emission power P are changing under magnetic disturbances as follows:

$$Q_l = Q_{lO}/[1+4x_z r_2/(r_2^2+x_m^2)],$$

$$\delta = -2r_2^2 x_z^2 \delta_z (I_1-I_3)/\{I_0 \cdot [(r_2^2+x_m^2-\delta_z^2)^2+4r_2^2\delta_z^2]\},$$

$$P/P_0 = (I_2-I_4)/I_0-1+ +[(I_1+I_3-2I_2)/I_0]\{x_z(2r_2+r_1)/[r_1(r_2^2+x_m^2)]\}- -2x_zr_2^{-2}\{1-(4r_2^2+4r_1r_2+r_1^2)/[r_1^2(1+r_2^2x_m^{-2})]\},$$

where Q_{10} is the emission line quality when there are no magnetic disturbances, $x_z=\mu_B\mu_0H_z/2h$, $x_m=\mu_B\mu_0H_m/2h$, H_z is the field strength of magnetic disturbance, oscillation field strength $H_m^2=2PQJ\eta/(\mu_0\omega_0V_c)$, r_1 and r_2 are relaxation rates of population difference and oscillating magnetization respectively, δ_z is the difference between a magnetic disturbance frequency and intrinsic Zeeman frequency multiplied by 2π, μ_B is the Bohr magneton, μ_0 is the magnetic constant, h is Plank constant divided by 2π, threshold values $P_0=h\omega I_0$ and $I_0=hV_cr_1r_2/(\mu_B\mu_0\pi JQ)$, ω_0 is the non-disturbed transition frequency, V_c is the cavity volume, I_1, I_2, I_3, I_4 are the atomic flow components in states $(F,m_F)=(1,1)$, $(1,0)$, $(1,-1)$ and $(0,0)$ correspondingly. The last expression is given in a simplified form proceeding from the assumption that $\delta_2<<r_2$. Expressions for δ and P/P_0 are taken from [8], and only significant terms are remained for clearness purpose. It follows from given expressions that an influence of magnetic disturbances x_z is reduced by Q_1 and P in the first power and by δ in the second power when J/V_c ratio increases.

Minimization of cavity temperature instability

H-maser long term frequency instability is caused to a great degree by cavity frequency temperature instability. It is possible to solve a problem of its minimization in such a case only when one can measure a cavity frequency with a relative error no more than $1\cdot10^{-11}$. It is convenient to perform cavity frequency measurement with a regenerator used in a self-excitation mode at the cavity operating resonant frequency. In such a case it takes a role of cavity frequency sensor.

When T_N=2000 K, regenerator self-excitation signal power P_r=0.1 mW, Q=2800, frequency instability $\sigma(\tau)$ is $4\cdot10^{-12}$ over 1 s. For example, self-exciting oscillator permit to measure a temperature of aluminum cavity with an error of 10^{-7} °C (when a regenerator power supply voltage changes are equal to zero). Such a cavity frequency sensor was used for tuning of cavity ovenizing system with a compensating heater. While using supplementary resistors, a relation between powers of main and compensating heaters was selecting by means of successive approximation technique which provides an independence of controlled cavity frequency from environmental temperature change. In this case, a temperature on the compensating heater exceeded that of regulating thermoresistor in such a degree that a temperature was reached in a certain point of the cavity body which was equal to that of the regulating thermoresistor, and the cavity temperature/frequency coefficient became equal to zero with an error caused by frequency instability. Cavity temperature/frequency coefficient is near to zero when slow change of environmental temperature takes place. During a dynamic behavior, thermocompensation is retarded against to the disturbing action. Such a dynamic error of the ovenized system can be decreased using an external step of ovenizing, for example, regenerator oven. Then it is possible to place a compensating heater on the internal step, but to connect it to the external step, in order to reduce a dynamic error to zero limit.

H-maser physical package

The design of physical package is given in Fig. 6.

The intermetallic hydrige $LaNi_4CoHx$ is used to supply the molecular hydrogen. Its 350 g contain 30 l of H2 at normal pressure, that is quite enough for 4 years of continuous maser operation. The hydrogen pressure is (1-2) atm at the room temperature and it is practically constant. The metal bulb with H_2 is vacuum - tightly connected to PdAg purifier. The hydrogen flow is controlled by purifier temperature the Pirani gauge is used in loop to stabilize a pressure at hydrogen flow. Purifier power consumption is about 1.5 W.

The discharge bulb of the atomic hydrogen source is made of pure quartz glass. Its volume is 0.1 l. The atomic beam is generated by multichannel collimator (300 channels of 20 μ diameter, 1.0 mm length each). The source is connected to the vacuum pump by indium gasket output. The discharge is excited by 115 MHz. Its power consumption (6-9) W.

The magnetic state selection is performed by means of a six-pole magnet from UNDK alloy. The bore diameter magnet is 1.5 mm magnetic induction of the pole tips is 0.9 T. Hydrogen and residual gas within volumes of beam and cavity is realized with two Ti getter pumps and two 0.9 l/s ion pumps.

The Al cylindrical magnetron-type cavity had a volume 0.3 l. Two capacitive electrodes in a shape of ring cutting metal field concentrators are located within the cavity. The surface of cavity and electrodes were covered by silver. Cavity Q-factor is 3500. Such a cavity

construction ensures high mechanical rigidity, shocks and atmospheric pressure changes. The cavity is provided with two coupling loops and two tuning varactors. The resonance frequency of tuning varactor is chosen to be lower than the cavity frequency for the linearization its retuning.

The storage bulb is located in the middle of cavity. The cylindrical bulb is made from thin wall quartz glass, its volume comprises 0.08 l, and it has escape time 0.35. The inner surface of the bulb is coated with three layers of fluorine - based polymer (F4MD or F10) with the help of specially developed technique. Its ensures small wall shift and spectral line boarding ($8 \cdot 10^{-12}$ and $6 \cdot 10^{8}$, respectively). Thermal stabilization of cavity and regenerator is provided by compensated one-zone ovens, using placed in vacuum. The temperature is controlled within a range 10^{-3}-10^{-4} °C. The 4-layered permalloy 81 NMA magnetic shield system is used. For improving the shielding factor the "lamp connection" of shielding elements is realized. Shielding factor is 10^{5} for a change of the field strength for B of $5 \cdot 10^{-5}$ T. The three shields are located within vacuum system and fourth one covers all physical package construction. A weak magnetic fields is generated by solenoid with correcting coils. H-maser's physical package weight is 10.5 kg, diameter is 170 mm, length is 370 mm.

Electronics

H-maser block diagram is given in Fig. 7. The main unit is H-maser physical package with Q-enhanced cavity (HMPP). This unit generates a signal with 1420405751.8 Hz - frequency and no less than 100 dBm power. The HMPP mode is provided by means of Pirani gauge, regenerator/cavity heater controllers (CRHC) i. e. overnizing system, and autonomous cavity auto-tuning (CAT).

The flow controller (FC) and Pirani gauge maintains a constant pressure of molecular hydrogen at the input of the discharge bulb while external disturbing factor change by means of changing voltage at the heater of the palladium purifier. A thermoresistor is used as a pressure sensor.

CRHC maintains a prescribed temperature for HMPP providing a temperature/frequency coefficient no more than $1 \cdot 10^{-12}$/°C without CAT.

CAT provides a lock of cavity frequency to the H-maser line excitation frequency, thus eliminating frequency shifts caused by destabilizing factors, such as temperature drifts, cavity mechanical deformations, load impedance changes and others.

CAT uses an amplitude technique which gives an error signal with information on cavity detuning. This signal is formed as a result of synchronous detection of H-maser transformed signal. A reference signal is that of cavity tuning frequency modulation. An integrated error signal is put on cavity tuning element, thus compensating a detuning.

Phase-lock loop (PLL) transfers HMPP frequency stability on Q-oscillator standard frequency. For signal transformation such units are used as receiver (REC), multiplier (M) and synthesizer (S). PLL has a correction of H-maser frequency with the adjustment step of $\pm 5 \cdot 10^{-14}$, as well as taking into account relativity correction for various space vehicle orbits.

The telemetric system (T) gives data on H-maser main parameters using amplitude and threshold sensor.

Control system (CS) provides through radio commands (RC) H-maser switching on, mode selecting and processing of other commands.

Power supply system (PS) generates a grid of necessary voltages for H-maser units from on-board mains (27V). PLL/CAT optimization was considered in [9]. The main original decisions were as follows:

- PLL provides an automatic achievement of synchronous mode. A search action is generated by a frequency/phase detector from receiver noises over all the range of initial detunings;
- PL-loop is the first order as tatil. It has a threshold circuit for accumulated error elimination due to QO frequency drift.

Frequency multiplier being a main source of H-maser frequency destabilization is realized with a circuit providing minimum phase temperature shifts.

CAT uses a working algorithm eliminating linear drifts of device elements.

Due to the change of spacing cavity frequency modulation signal, along with a compensation of spin-exchange frequency shift, an influence of cavity Q-factor is excluded.

General view of H-maser is shown in Fig. 8. Its dimensions are 500x365x365 mm, volume is 65 l, weight is 40 kg.

Test results and future trend

Strong maser oscillation was obtained with enhanced cavity Q_r of 56.000 and atomic line $6 \cdot 10^8$. Two H-maser sets have been manufactured in 1994 and demonstrated short term stability $\sigma_y(\tau)$ is $1.5 \cdot 10^{-12} \tau^{-1/2}$ over the interval for 10^4 s is $(1.5\text{-}2) 10^{-14}$ for 24 hours. Fig. 9 represents its behavior under normal laboratory conditions. The results obtained are closed to calculated ones. Over the interval (1-100) s the main contribution to instability of H-maser gives regenerator noises, while in the interval up to 10^6 s Q-enhanced cavity temperature drift is added. Flicker floor depends on the amplifier regenerator, phase shifter and temperature sensitivity. The investigated sets had excessively large range regulation of phase and amplifier regenerator (250^0 and 25 dB, respectively). Their lowering should result in decreasing noises caused by instability of chain supply power.

Despite the use of CAT in H-maser, the temperature sensitivity of cavity can be decreased by thermal compensation of static regulation error. It is possible to get using optimum relation between the power or main and additional heaters (RIRT reference CAT test are demonstrated daily frequency drift less than $1 \cdot 10^{-16}$). Compact H-maser meets all GLONASS requirements, passed through all kind of ground tests, suitable for serial production. Since 1995 the new model of H-maser is developing to improve technical performance without increase the size and weight. We changed a storage bulb up to (0.2-0.3) l with the same Q-enhanced magnetron-type cavity having higher loaded Q. We also modernized the cavity and regenerator ovenizing systems. It made possible with the help of above mentioned technique to reach temperature stability at the level of $5 \cdot 10^{-5}$ ^{0}C. We expect to have daily frequency stability near $(3\text{-}5) 10^{-15}$.

Conclusion

The basic design of proposed Q-enhanced oscillating compact H-maser for GLONASS is described. It is pointed out that the temperature sensitivity of Q-enhanced cavity and flicker noise of regenerator is expected to be difficult problem to solve. However, if it can be solved, the compact H-maser would be capable of approaching its theoretical stability out to $(10^5\text{-}10^6)$ s close to that of conventional maser. We have shown the viability of the concept of Q-enhancement in H-maser. High long term stability H-maser with a properly designed cavity auto-tuning has been demonstrated. Over the short term ($\tau \leq 1$ day) the environmental effects and short term noise dominate wich are characterized by Allan variance measurement as white frequency noise ($\tau^{-1/2}$). The stability performance can be enhanced by from improved packaging to minimize environmental sensitivity and technological advancement in low noise electronic components. The achieved small size, low weight , low power consumption (included warm up time) and high stability performance will provide system designers with on additional option in the choice of high stability frequency sources for modern navigation and communication systems.

Reference

1. R.Vessot et. al. "A hydrogen maser for space." Proc. of the 6th EFTF, 1992, pp.19-25.

2. R.Vessot et. al. "Test of relativistic gravitation with a space-born H-maser." Phys. Rev., 1980, v.45, pp.2081-2084.

3. L.Bernier. "Basic design of a space borne H-maser." Proc. of the 6th EFTF, 1992, pp.283-286.

4. V.Gankin et. al. "A frequency standard based of hydrogen generator with Q-enhanced cavity." Voprosi Radioelectroniki OVR, 1992, N6, pp.74-80.

5. H.T.M.Wang. "Subcompact hydrogen maser atomic clock." Proc. of the IEEE, 1984, v.77, N7, pp.982-992.

6. A.Myasnikov et. al. "Design and characteristics of small-size H-maser." Radionavigation and Time, 1993, v.1,2, pp.88-93.

7. D.Kleppner et. al. "Hydrogen maser principles and techniques." Phys. Rev., 1965, v.138, second series, N4A, pp.A972-A983.

8. G.Busca et. al. "Zeeman effects on H and Rb maser." Proc. of the 31th Annual Symp. on Frequency and Control, 1979, pp.535-542.

9. Yu.Gouzhva et. al. "Automatic control systems for ground and space-based hydrogen frequency standard." Proc. of the 9th EFTF, 1995, (in press).

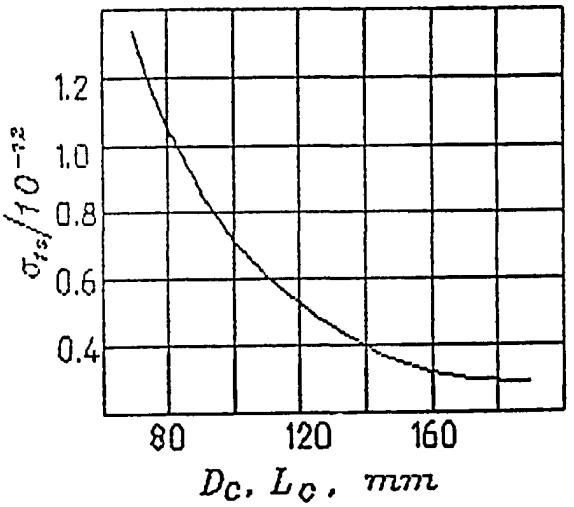


Fig.1. The dependence of short term frequency stability σ_{1s} on the cavity sizes.

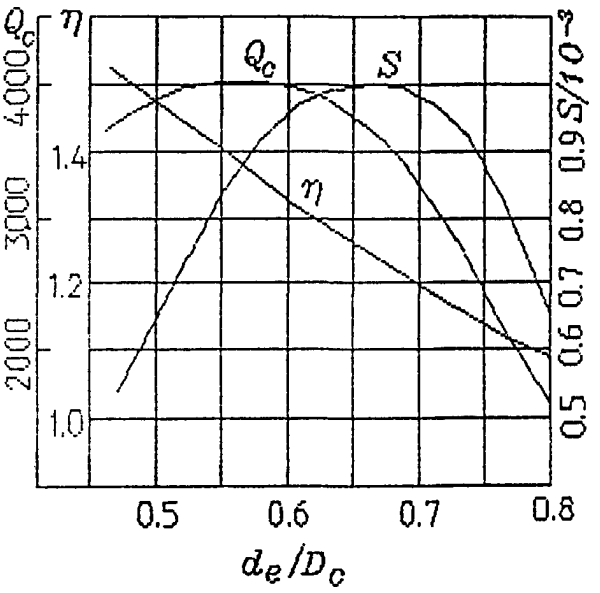


Fig.2. The dependences of unloaded Q-factor cavity Q_C, filling factor η and excitation parameter S on the relation of electrode diameter d_e to cavity diameter D_C

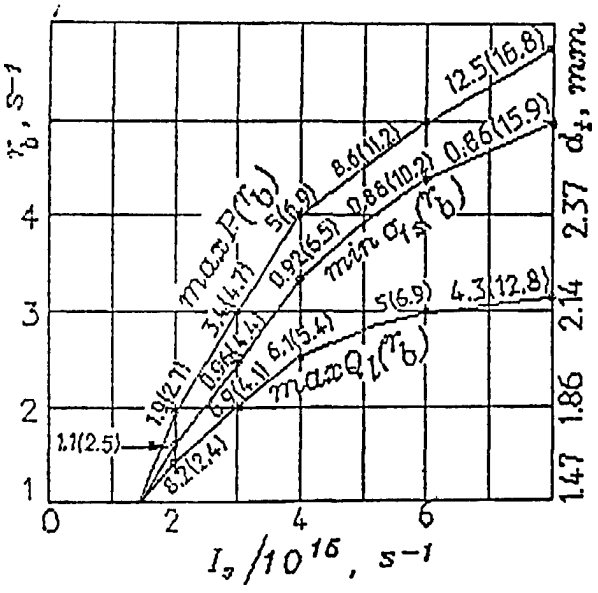


Fig.3. Optimization escape rate of hydrogen atoms r_B from a bulb.

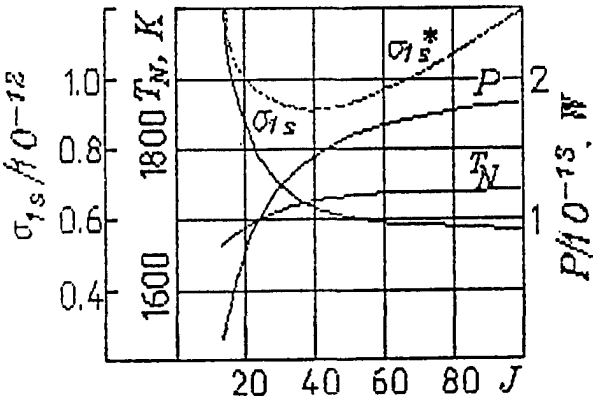


Fig.4. Short term frequency stability σ_{1s}, flicker noises σ_{1s}^*, noise temperature T_N, radiation power P versus regeneration coefficient J.

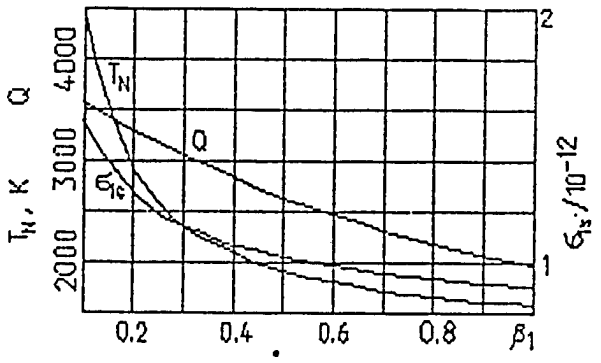


Fig.5. The dependence of noise temperature T_N, loaded Q-factor of cavity Q and short term frequency stability σ_{1s} on input cavity coupling β_1.

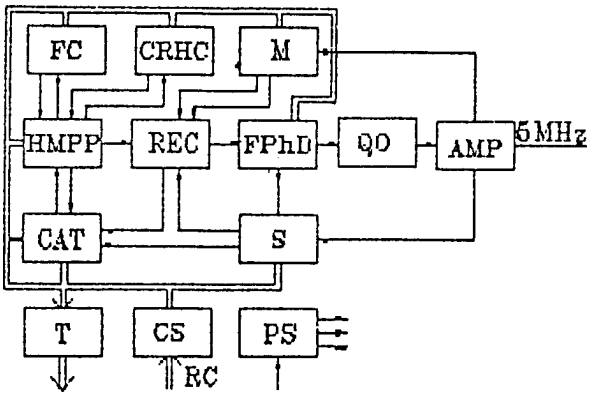


Fig.7. Block-diagram of compact H-maser.

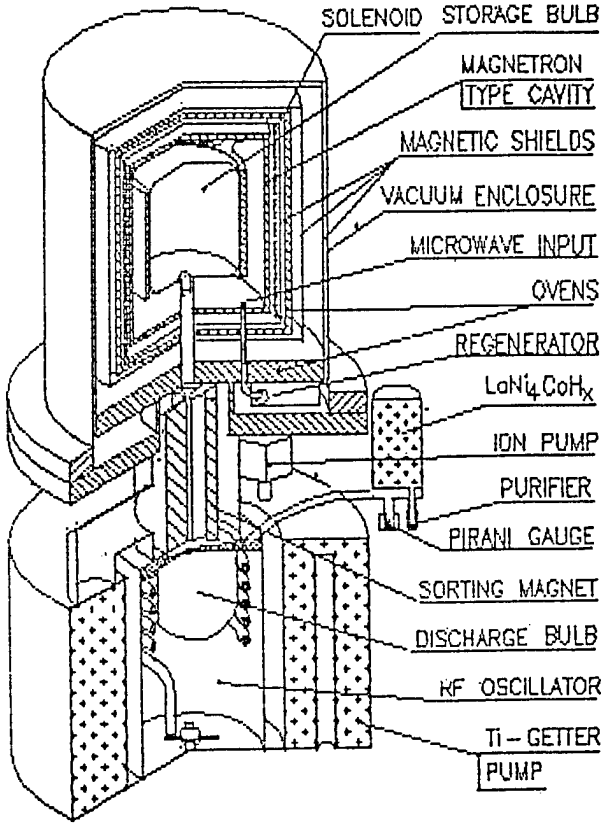


Fig.6. H-maser physical package.

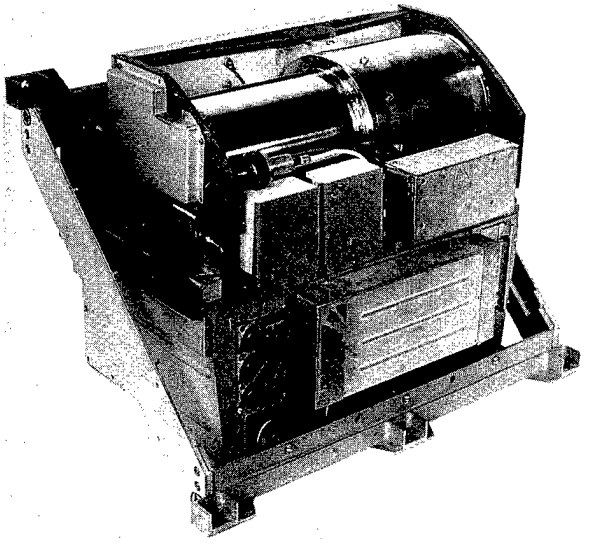


Fig.8. General view of compact H-maser.

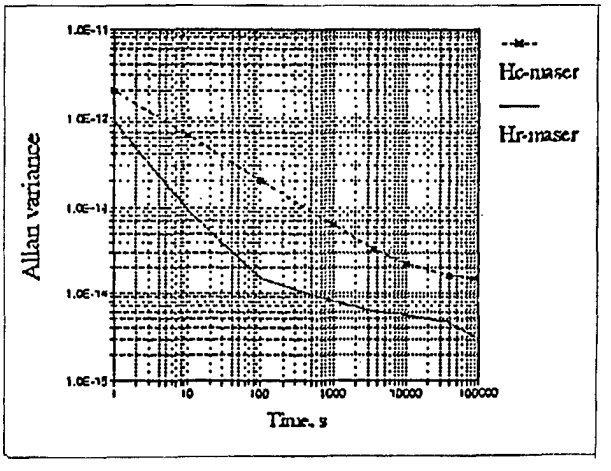


Fig.9. Compact H_C-maser frequency stability.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

INFLUENCE OF SATURATION BEAM PARAMETERS ON LASER FREQUENCY LOCKED TO THE CESIUM CYCLING TRANSITION

Ramiz Gamidov, Ismail Taşkin, and Vladimir Sautenkov

Ulusal Metroloji Enstitüsu (UME), P.K. 21, 41470 Gebze-Kocaeli, Turkey
FAX: (90) 262 646 59 14

In this work, external cavity laser diodes were stabilized on the cesium cycling transition (852 nm). The influence of the laser beam parameters (power, diameter and polarization) on the frequency of the laser have been studied. Contributions of light-pressure effects on the laser frequency shift were observed.

Introduction

Atomic clocks based on optically pumped and cooled cesium atoms require laser sources with high frequency stability [1-3]. Usually the lasers diode are locked to the cycling transition at 852 nm. Different stabilization techniques have been used [4-6] and frequency stabilities laser diodes at better than 10^{-12} has been achieved [6]. However, in the literature there is not enough information about possible frequency shifts of the stabilized laser.

In this work, we have studied the dependence of the laser diode frequency on the optical beam parameters which include optical power P, diameter D, and polarization experiments were performance on the cesium D_2-line by using laser diodes.

Experimental results

A block diagram of the experimental setup is presented in Fig.1. In our experiment, two external cavity diode lasers are used. The laser is the AlGaAs type AC-SDL-5410. The output of the external cavity diode laser is zero-order diffracted beam, with contains 40% of the optical power incident off of a 1200 lines/mm diffraction grating. We detected the Doppler-free resonance of the Cs cycling transition by using standard weak probe wave (WPW) and standing wave (SW) methods with linear or circular polarized laser beams. The ratio between pump and weak probe beam powers was 25 in the WPW experiments. The

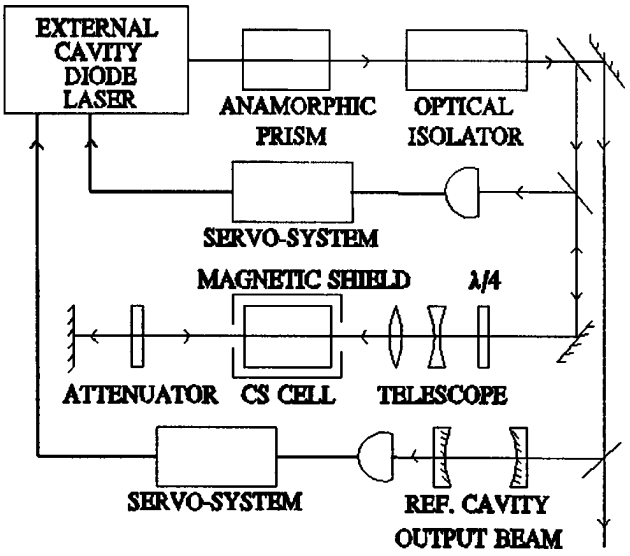


Fig.1. Experimental setup.

cesium cells (length 3 cm, temperature 21 C) are surrounded by magnetic shields to prevent frequency drift due to magnetic field fluctuation. The beam diameter, D, of the first laser in the cesium cell can be varied from 0.2 cm to 1.2 cm by changing the magnification of the telescope. Two sets measurements were performed. In the first set, lineshapes of the Doppler-free resonances were investigated. In the second set, we had measured the frequency difference between two lasers which were locked to atomic Doppler-free resonances. In both cases, the laser frequency was modulated at 20 kHz and the standards synchronous demodulation technique (lock-in amplifier and integrator) were used for the frequency stabilization on the top of the reference resonance [6]. The modulation and the error signal were sent to a PZT which tuned the external cavity length. The laser

0-7803-2500-1/95 $4.00 © 1995 IEEE

frequency deviation (2 MHz) was considerably less then the Doppler-free resonance width. In this condition, the signal after the lock-in amplifier is proportional to the first derivative of the Doppler-free resonance.

Spectral lineshapes

The spectral shapes of the saturated Doppler-free resonances were measured (the derivative of the Doppler-free resonance). In this experiment we locked the laser frequency to the reference cavity. The signal from the lock-in amplifier was recorded by the computer. In Fig.2 typical Doppler-free resonances observed at different pump beam intensities, I, are presented.

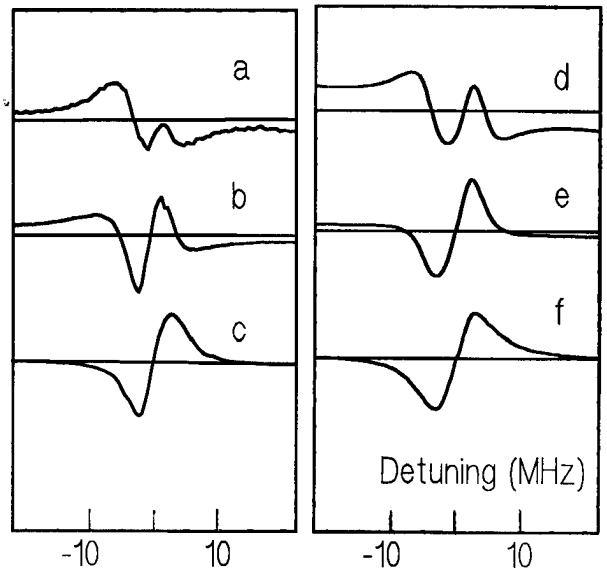


Fig.2. Doppler-free resonances in WPW (a,b,c) and SW (d,e,f) experiments (D=1 cm) with π polarized light: (a,d) - I=0.2 mW/cm^2, (b,e) - 0.9 mW/cm^2, (c,f) - 5 mW/cm^2.

Observed resonances show complicated spectral shapes which depend on the laser beam intensity. The intensity dependent distortion introduces a frequency shift of the Doppler-free resonance.

Frequency shift and stability

In order to investigate the laser frequency shift and the Allan variance (the laser stability) we used heterodyne technique. The lasers were locked to the Doppler-free saturation resonances on the cesium D_2 line. The first laser was stabilized on the cesium cycling transition, F=4 - F=5. The second laser was locked to a neighboring cross-over resonance between F=4 - F=5 and F=4 - F=4. The light beams from both lasers superimposed on the fast photodiode. The beat signal was amplified and counted.

In Fig.3, the frequency difference, Δf ,versus the pumping beam power in the WPW experiment are presented for different beam geometries.

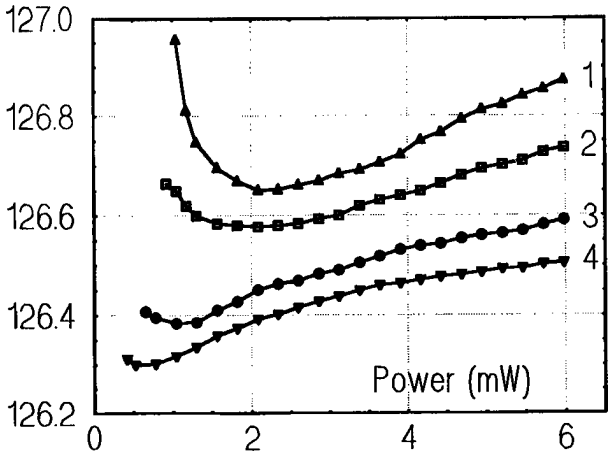


Fig.3. The frequency difference (MHz) vs pump beam power (mW) in WPW experiments: 1 - D=1 cm, σ polarization, 2 - D=1 cm, π polarization, 3 - D=0.4 cm, σ polarization, 4 - D=0.4 cm, π polarization.

In Fig.4, the frequency difference, Δf , versus the pumping beam power in the SW experiments are shown.

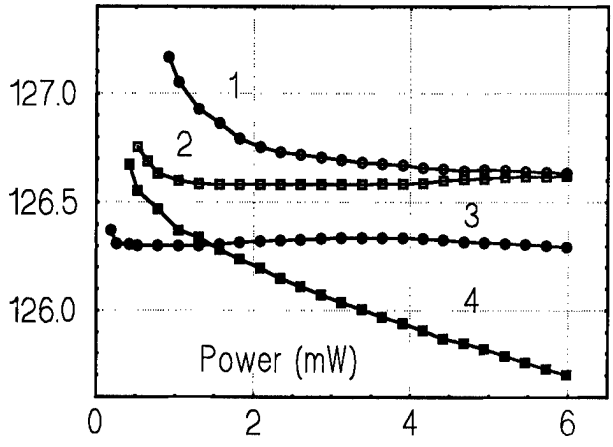


Fig.4. The frequency difference (MHz) vs pump beam power (mW) in SW experiments: 1 - D=1 cm, σ polarization, 2 - D=1 cm, π polarization, 3 - 0.4 cm, π polarization, 4 - D=0.4 cm, σ polarization.

One can see that for reasonable beam diameters it is possible to choose a laser pump power, P, and a ratio between counter-propagating beams in a such way that the laser frequency will be practically power independent.

We would like to measure the frequency stability of our lasers under optimum conditions. However, the Allan variance was limited by frequency drifts of the second laser which was locked to the crossover resonance. This resonance is on the Doppler-contour slope and very sensitive to the laser intensity and cell temperature fluctuations. Typical Allan variances of the beat note versus the averaging time are presented in Fig.5.

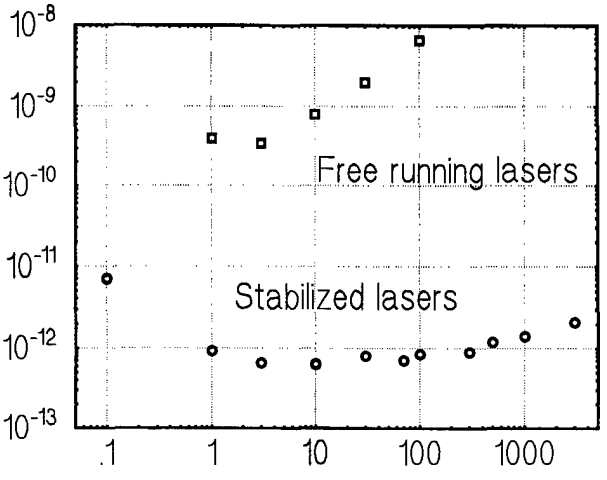


Fig.5. Square root of the Allan variance.

When the lasers are stabilized, the measured Allan variance shows a flicker floor at 7×10^{-13}.

Discussions

In many-level atomic systems the saturated absorption can be due to a redistribution of the population between the ground and the excited states [7], and also due to the ground state optical pumping [8]. In the first case, the saturation is reached when the excitation rate, $\sigma I/hf$, is comparable with the relaxation rate of the excited state, γ. For typical atomic cycling transitions the saturation intensity is $I_s \sim 10$ mW/cm^2 ($\gamma \sim 10^{-8}$ s). The relaxation time of the ground state population is described by the integration time, T=D/V, which is approximately 10^{-5} s, for D=0.4 cm. In this case the saturation intensity, $I = hf/\sigma T$, is nearly two orders of magnitude less. At low laser beam intensity ($I \ll I_s$) observed resonances show complicated spectral line-shapes. The Doppler-free saturation resonance on the cycling transition is a superposition of two resonances with different signs and width. One of them (a narrow resonance) is typical for two-level atomic system, another (a broad resonance) is due to optical pumping of ground state magnetic sublevels. At high laser beam intensity ($I \gg I_s$) practically all of the atoms are optically pumped and the Doppler-free saturation resonance due to excited state population will dominate. We locked the laser frequency to "the narrow resonance".

Now we would like to discuss a light pressure effect which depends mainly on laser beam diameter. For simplicity we consider only weak probe wave experiments with circularly polarized beams. The light-pressure-induced frequency shift of the Doppler-free resonance in a two-level atomic system was predicted by Kazantzev et al [9]. The calculated frequency shift, δf, is equal to $\varepsilon T\gamma/2$, where ε is the recoil frequency shift, T is the interaction time and γ is the linewidth. A light-pressure-induced asymmetry of the Doppler-free resonance was observed in experiments with two co-propagating beams [10-12].

Here we report the first observation of the light-pressure-induced frequency shifts in the counter-propagating beams scheme. We have changed the beam diameter, D, from 0.4 cm to 1.2 cm with steps of 0.2 cm. We have observed that the laser frequency shift is proportional to beam diameter, D. The coefficient is estimated as 0.5 MHz/cm, and this value is in reasonable agreement with the theoretical prediction for the cesium transition ($\varepsilon = 2\pi$ 0.5 kHz, $\gamma = 2\pi$ 5.3 MHz).

It is complicated to give explicit analysis of the observed intensity dependence of the frequency shift at high intensities ($I >$ 1 mW/cm^2). The lineshape and frequency shift of the Doppler-free resonances in many-level atomic systems can depend on many parameters. For instance, different saturations of neighboring hf-transitions causes strong intensity dependence of the Doppler background. We plan to make additional measurements by using third derivative approach (synchronous demodulation on 60 kHz). By using this method one can diminish "the slope effects" (influence of the intensity dependent Doppler background).

Conclusions

Influence of laser beam parameters (power, diameter and polarization) on the laser diode frequency

has been studied. Light-pressure-induced frequency shift were observed. Frequency stabilities of better than 10^{-12} were demonstrated. The stabilized lasers will be used for optical cooling and trapping of cesium atoms.

Acknowledgement

We are grateful to Sevilay Uğur for valuable contribution in the data acquisition and Ibrahim Akı for technical assistance.

References

[1] R.E.Drullinger, D.J.Glaze, J.L.Lowe, and J.H.Schirly, "The NIST optically pumped cesium frequency standard", IEEE Trans.Instrum.Meas., vol.40, pp. 162-164, April 1991.

[2] E. de Clercq, G.D.Rovera, S.Boizid, and A.Clairon, "The LPTF optically pumped primary frequency standard", IEEE Trans.Instrum.Meas." vol. 42, pp. 437-461, April 1993.

[3] S.N.Lea, A.Clairon, S.Salomon, P.Laurent, B.Loinis, J.Reichel, A.Nadir, and G.Santarelly, "Laser cooling and trapping of atoms: new tools for ultra-stable caesium clocks", Physica Scripta, vol.T51, pp. 78-84, 1994.

[4] A.M. Akulshin, V.V.Nikitin, V.A.Sautenkov, V.V.Vasiliev, V.L.Velichansky, E.K.Yurkin, and A.S.Zibrov, "Frequency stabilization of highly coherent AlGaAs diode lasers", in Proceedings of 4th Symp. Frequency Standards and Metrology (Ancona, Italy, 1988) Berlin: Springer-Verlag, 1989, pp. 236-241.

[5] T.Ikegami, S.Oshima, and M.Ohtsu, "Frequency stabilization of laser diode on the Cs-D2 line with Zeeman modulation method", Jap.J.Appl.Phys., vol. 28, pp. L1839-1841, October 1989.

[6] G.D.Rovera, G.Santarelly, and A.Clairon, "A laser diode system stabilized on the Caesium D2 line", Rev.Sci.Instrum., vol.65, pp. 1502-1504, May 1994.

[7] Y.R.Shen, The principles of non-linear optics, New-York: John Willey & Sons, 1984, ch. 13, pp. 216-227.

[8] M.S.Feld, and D.E.Murnick, "Saturation spectroscopy with laser optical pumping in atomic barium", Phys.Rev., vol. A21, pp. 1955-1968, June 1980.

[9] A.P.Kazantzev, G.I.Surdutovich, and V.P.Yakovlev, "Influence of light pressure on non-linear susceptibility of resonance atoms", Pisma Zh.Eksp.Teor.Fiz. (in Russian), vol. 43, pp. 222-224, 1986.

[10] R.Grimm, and J.Mlynek, "Light-pressure-induced line-shape asymmetry of saturation dip in an atomic gas", Phys.Rev.Lett., vol. 63, pp.232-235, July 1989.

[11] A.M.Akulshin, V.L.Velichansky, R.G.Gamidov, A.P.Kazanzev, V.A.Sautenkov, G.I.Surdutovich, and V.P.Yakovlev, "Effect of radiation light pressure on shape of saturated absorption resonance of cesium vapor"' JETP Lett., vol.50, pp. 187-190, August 1989.

[12] R.Gamidov, A.C.Ismailov, and H.Uğur, "On the structure of sub-Doppler absorption resonance on atomic transition from ground degenerate state", Optics and Spectroscopy, vol. 77, pp. 6-10, January 1994.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

EXTERNAL CAVITY DIODE LASER HIGH RESOLUTION SPECTROSCOPY OF THE Ca AND Sr INTERCOMBINATION LINES FOR THE DEVELOPMENT OF A TRANSPORTABLE FREQUENCY/LENGTH STANDARD.

A.CELIKOV,* P.KERSTEN, F.RIEHLE, G.ZINNER, L.D'EVELYN**, A.ZIBROV*, V.VELICHANSKY,* J.HELMCKE

Physikalish Technische Bundesunstalt, D-38116 Braunschweig, Germany
*P.N. Lebedev Physics Institute, 117924, Leninsky 53, Moscow, Russia
**NIST, 325 Broadway, Boulder, CO 80303, USA

ABSTRACT

Capabilities of a high resolution diode laser spectrometer have been demonstrated by performing spectroscopy with thermal Ca and Sr atoms in an atomic beam and with laser cooled Ca atoms in a magneto optical trap. The minimum spectral linewidth of the observed fringes is as narrow as 1.5kHz, allowing a small scale diode laser transfer frequency/length standard with a very low level of uncertainty.

INTRODUCTION

The now widely used wavelength-frequency standard in the visible based on a I_2 stabilized He-Ne laser has limitations in performance due to the cell dependent systematic shifts and a relatively large width of the reference transition ≈(1MHz). The intercombination transitions in the even isotopes of Sr, and Ca prove to have been a very promising wavelength/frequency reference in the visible [1], given their very small radiative linewidth, insensitivity to electrical and magnetic fields as well as the absence of the hyperfine structure. However, their widespread use has until now been restricted due to some practical considerations, namely the necessity of using a dye laser to excite these transitions. And though a lot of work has been invested to carefully study and understand the system [2,3], developing a transportable standard based on a Ca stabilized dye laser seems impossible. Relative new-comers to the area of high resolution spectroscopy, diode lasers might offer an alternative to dye lasers and enable high accuracy transfer standards [4,5,6]. In order to achieve the ultimate performance of a frequency standard, the influence of velocity dependent shifts and broadening should be reduced by employing laser cooling. The strong $^1S_0$-$^1P_1$ singlet transition allows laser cooling and trapping atoms, which reduces the influence of the Doppler effects and other systematic errors and increases the observation time. Such laser cooled atoms in the case of Ca would provide a frequency/length standard with the uncertainty of $3*10^{-15}$ [22] provided the clock laser is good enough to match the 400 Hz radiation linewidth of the Ca transition and all the necessary conditions of optimal excitation of the clock transition, such as power and the quality of the wavefront are met.

Most laser cooled systems require a long interrogation time of the resonant light with the absorber to achieve high spectral resolution and a sufficient signal to noise ratio necessary for high accuracy and stability, which implies that the laser oscillator exciting the clock transition should be stable. As an alternative to dye lasers, diode lasers can also provide a precisely controllable local oscillator with extremely high spectral purity. In this paper we demonstrate a diode laser system capable of kHz resolution of the Ca and Sr intercombination lines.

0-7803-2500-1/95 $4.00 © 1995 IEEE

As the output power of nowadays commercially available red diode lasers is somehow restricted., instead of using an AOM to control the frequency of a diode laser as in the case of a dye laser [7]., a two laser system suggested and realized in [8-10] might be more convenient. Here one laser is stabilized to a non tunable cavity and an offset phase locked slave laser is utilized to perform precise control and frequency tuning over atomic transitions. Such systems have been developed by several groups around the world and are successfully used in a number of ongoing experiments.

DIODE LASER SYSTEM

We used two systems: a two laser spectrometer, and a diode laser stabilized to a tunable reference resonator The block-diagram of a two laser spectrometer is depicted in fig 1.

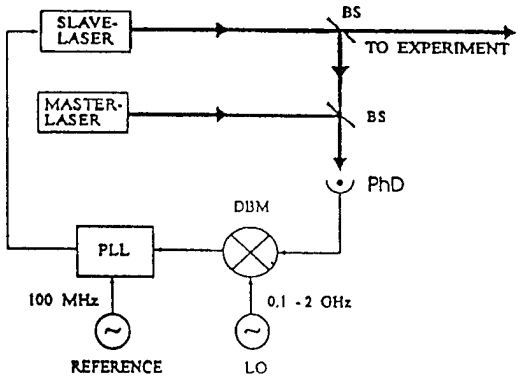


Fig.1 Block diagram of a two laser spectrometer. The slave laser is phase locked to the master laser with a controllable offset. PPL-phase locked loop, DBM- doble balanced mixer, BS- beam splitter, PhD- photodiode

Two diode lasers are operated in an external configuration [11], allowing enough resolution and feedback level to maintain a stable single mode regime. In this scheme, the reflection of the laser diode output mirror is reduced by putting an anti reflection (AR) coating on it and the output is loaded with a grating in the Littman configuration, so that the first diffraction order of the grating is retro-reflected from a PZT driven mirror and after diffracting on the grating again goes back into the laser. Coarse tuning is performed by adjusting the mirror and fine tuning is achieved by changing the length of the resonator by PZT. An etalon is placed inside the resonator to enhance the selectivity and suppress spurious modes of the extended resonator. This double pass configuration was chosen mostly for convenience, as the output beam does not change its position in space while the frequency of the laser is tuned. The linewidth of a diode laser in an extended cavity configuration is significantly reduced compared to a solitary laser diode [12]. As measured the linewidth of an ECDL in a free runninng way is about 200kHz, with the fast linewidth determined by the spontaneous noise level being about 40kHz and the main contribution the linewidth is made by low frequency Fourier components of the noise spectrum depending on frequency in 1/f and $1/f^2$ fashion. Therefore as it was shown [13,14] it is possible to control the frequency with the smaller bandwidth electrical feedback than in the case of a solitary laser diode.

The radio frequency phase modulation technique [7,15] provides a good means for reducing the frequency fluctuation of a laser forming a broad band error signal pretty much immune to the offset drift of electronics and an unwelcome AM modulation of the laser light. We have used this method to lock the master laser to a nontunable nonconfocal resonator, which was suggested in [16] as a reference for ultimately high short term stability, and low drift. High finesse (F=3000) mirrors (focal lengths are 1m and 10m) were optically contacted to a finished facet zerodur spacer, which was enclosed into a copper shielding suspended in vacuum with four springs. This allowed acoustical insulation, minimized coupling to the basement and significantly

reduced frequency jitter. The outer chamber was wrapped with a copper enclosure which was temperature stabilized and isolated from the outside with a piece foam material. To damp the pendulum modes, soft indium wires were used. The linewidth of the resonator was approximately 300kHz providing a steep slope and a good signal-to-noise ratio frequency discriminator.

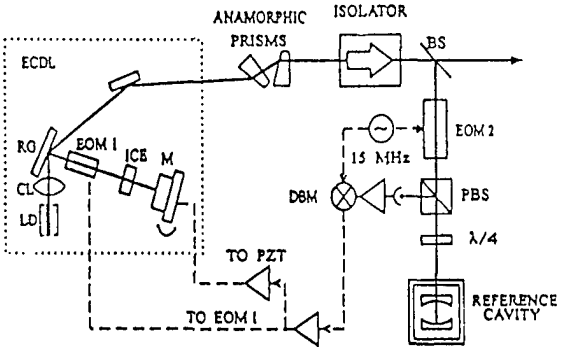


Fig.2. Schematic diagram of the frequency stabilized diode laser. LD-laser diode, CL-collimating lens, RG-reflection grating, EOM-electro-optical modulator, BS-beam splitter, PBS-polarizing beam splitter, DBM-double balanced mixer

A usual rf heterodyne set up fig.2 with an external phase electro optical modulator (EOM2) was used, with the modulation frequency of 15MHz well above the linewidth of the resonator was chosen to provide a broad band acquisition range. The error signal was integrated and fed back via two channels. A slow (5kHz) feedback loop to the PZT driving the mirror of the laser resonator centered the frequency of the laser to the reference line and helped restrict the dynamic range of the fast channel. Fast fluctuations were compensated by putting the feedback signal into an intracavity EOM1 [17], a $LiNbO_3$ crystal 4*10*40mm in size. The crystal was mounted on a lead substrate to suppress pieso resonances. Fast performance of the EOM allowed high gain at low frequencies which was rolled off with tripple double and single pole units to a unity gain of about 2.3MHz. Transfer function correction was introduced for the phase shift of the reference cavity, the bandwidth was limited mainly due to a prominent piezo resonance of the EOM1 at 2.3 MHz. The linewidth of the master laser was checked by a heterodyne method against a dye laser stabilized to a high frequency recoil component of the Ca intercombination transition. The beat note between the two lasers is plotted in fig.3.

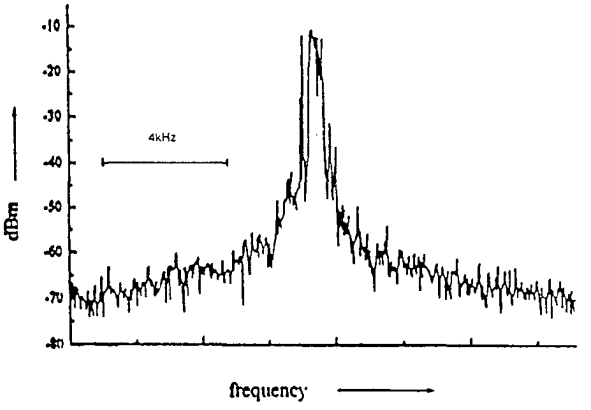


Fig.3 Beat note spectrum between a frequency stabilized diode laser and a dye laser. RBW is 100Hz, Span time is 15s.

The drift of the laser frequency, due to the thermal change in the reference resonator length was measured to be 50 Hz per second.

As high resolution spectroscopy requires both frequency stability and a good quality of the wave front, it is necessary to use spatial filtering. Sending a beam through a single mode fiber might solve the problem of the wave front but at the cost of power, for due to poor wave front quality the coupling rate into the fiber is relatively low and as a consequence much power is lost. Therefore we used another diode laser (ECDL) slaving to the master laser. Under the phase locked conditions the stability of the master laser is transferred to the slave laser and achievable resolution completely determined by

the stability of the master laser. Most of the power of the slave laser was saved for spectroscopy while 10μW was split off to get the beat note with the master laser stabilized to the reference resonator. The offset of the two lasers was adjusted by changing the frequency of the rf local oscillator which was mixed with the beat note signal to come down to 100MHz reference frequency. The error signal from the output of the phase/frequency detector was integrated in the appropriate bandwidth and fed back into the slave laser. Fast fluctuations were corrected by changing the bias current of the laser and slow ones by PZT. The necessary lag lead corrections were added to compensate for diode laser frequency response roll off. The attainable servo bandwidth of about 1.5MHz of the phase locked loop was mainly limited by the frequency dependent ECDL modulation response.

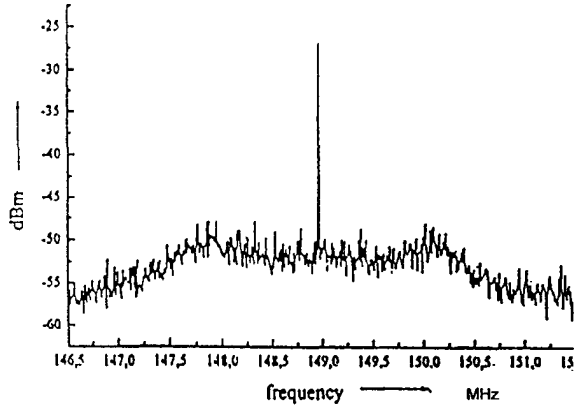


Fig.4 Beat note spectrum between two external cavity diode lasers under the phase locking conditions. RBW is 3kHz, Span time is 15s.

Fig.4 shows a beat note spectrum of the slave laser phase locked to the master ECDL stabilized to a reference resonator. This laser system was used in the Ca experiment.

To perform spectroscopy of Sr we locked the frequency of an external cavity diode laser in the Littrow configuration to the slope of a reflection resonance of a confocal resonator, having the line width of 7MHz (FWHM) and the FSR of 1.5 GHz.

CA SPECTROSCOPY

1. HOT ATOMS

We used an atomic beam apparatus which is schematically diagrammed in fig.5, and described in details in [18].

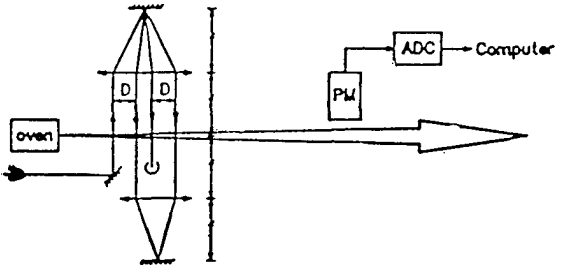


Fig.5 Experimental set up for separated field excitation of Ca atoms with traveling waves.

The fluorescence was detected one decay length downstream of the excitation region with a photo multiplier. The residual Doppler broadening determined by the degree of collimation of the atomic beam and by the temperature of the oven was 1.7MHz. A small magnetic field in the interaction region parallel to the electric vector of the laser field was applied to allow only the $\Delta m=0$ transition. Two pairs of counter-propagating running waves crossing the atomic beam perpendicularly formed an optical Ramsey excitation scheme. We chose this excitation scheme to get an optimum contrast [19]. To the first order of approximation the resolution (FWHM) depends on the separation between the excitation zones and the effective longitudinal velocity V_x of the atoms [20]:

$\Delta\nu=V_x/4D$

Nonlinear optical Ramsey fringes obtained at a separation 2D of 28mm between the two zones are presented in fig.6.

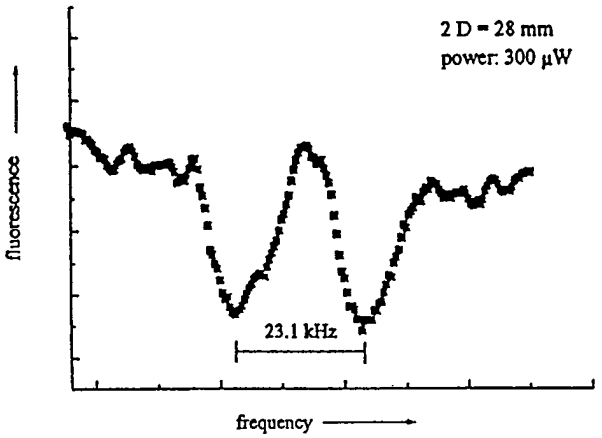


Fig.6 Ramsey fringes in an effusive Ca beam obtained with excitation by four running waves, at the separation between the interaction zones 2D = 28mm, The diameter of the beam is 3mm; integration time is 1s.

The frequency of the slave laser was scanned over the transition by changing the offset frequency between the two lasers. The resolution of 10 kHz was limited by the separation 2D determined by the geometrical size of the windows. In striving to get maximum accuracy with a thermal beam, after increasing the separation distance, and optimize the phase of the excitation field, we would almost immediately come to the restriction imposed by the second order Doppler. Reducing the second order Doppler effect leads us to using cooled atoms whose velocity distribution is narrower, which minimizes the uncertainties associated with the optical phase shift and second order Doppler.

2. COLD ATOMS

The atomic beam apparatus for cooling deflection and trapping Ca atoms shown in fig.7 is described in details elsewhere [21], therefore we will present a short description here.

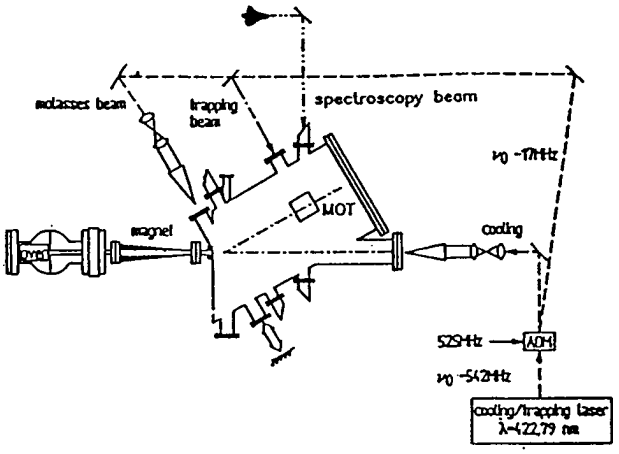


Fig.7 Schematic diagram of the apparatus for cooling and deflecting a calcium atomic beam and for trapping the atoms in a magneto optical trap.

The light at the wavelength of 423nm for atom manipulations was produced by a ring dye laser. An appropriate AOM technique provided necessary frequency detuning of the stopping deflecting and trapping beams.

The Zeeman shift in a spatially varying magnetic field applied to the decelerating atoms compensated for the Doppler shift during the cooling process. To get rid of the systematic light shifts caused by the stopping radiation the beam of atoms with the longitudinal velocity distribution centered at 40cm/s was deflected by approximately 30 grad. with respect to the direction of the atomic beam by a stading wave forming a flux of cold atoms with a density of 10^{10} atoms/s. To further reduce the velocity of atoms, we used a magneto optical trap, where atoms are confined in a small volume. The

optimal storage time of the trap as measured in [22] was about 100ms and the amount of trapped atoms was 10^7.

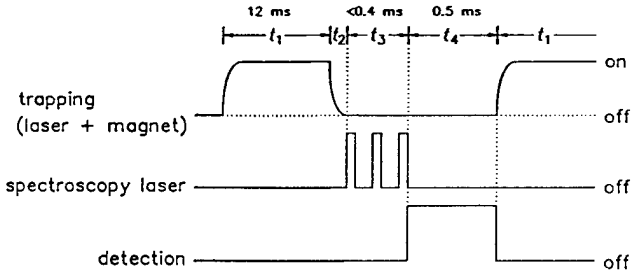


Fig.8 Scheme of the time sequence used for spectroscopy with trapped atoms.

We performed our spectroscopic measurements in a pulsed regime, the sequence of pulses is shown in fig.8, where the cold atoms are prepared by sending a pulse of blue light for cooling, deflecting, and trapping the atoms lasting about 12ms and after all the magnetic coil switching off transit processes died away, three successive pulses of a standing wave from the clock laser interrogated the intercombination transition, and after that the fluorescence of the atoms decaying from the P level was detected during the time corresponding to its lifetime. Despite the fact that the standing wave Ramsey excitation scheme does not give the optimal contrast of the Ramsey fringes, we chose it for simplicity of the experimental set up and, besides it helps avoid unwelcome optical phase error shifts [23]. To make all the atoms comprising the Doppler line contribute to the signal we chose clock pulse duration to be $1/\Delta\nu_D$, consequently the power necessary for atoms to get a $\pi/2$ pulse was approximately 4mW. As we could not get this much power by sending the light from the diode laser through a single mode fiber, we slavedthe dye laser to the master diode laser and used the same offset technique to scan the laser over the transition, so that the phase stability was taken from the diode laser system. The spectroscopic diagrams corresponding to the different time interval between the pulses are shown in fig.9

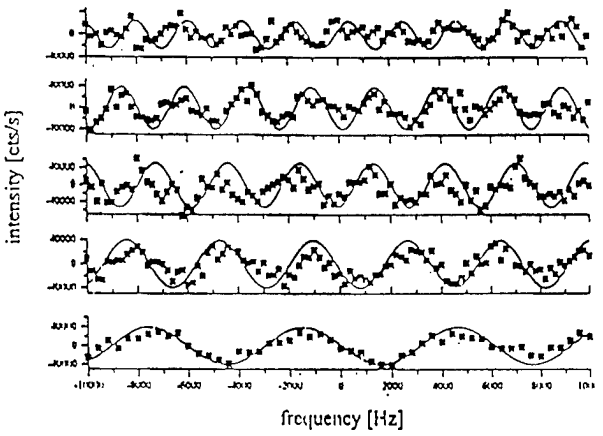


Fig.9 Ramsey fringes obtained in a magneto optical trap by applying a series of three consecutive pulses of a standing wave. The spectral resolution is determined by the separation between the pulses.

During the integration time of 1s nearly 100 trapping-excitation-detection processes were performed, and then the frequency of the offset laser was changed and the measurements were repeated and so on. The attainable resolution mainly restricted by the thermal drift of the reference cavity was about 1.5kHz.

SPECTROSCOPY OF Sr

Doppler-free resonances of saturated absorption in Sr beam are shown in Fig.10.

The spectra were taken during approximately 1 second by applying a ramp to the PZT of the reference cavity. For details see ref. [24].

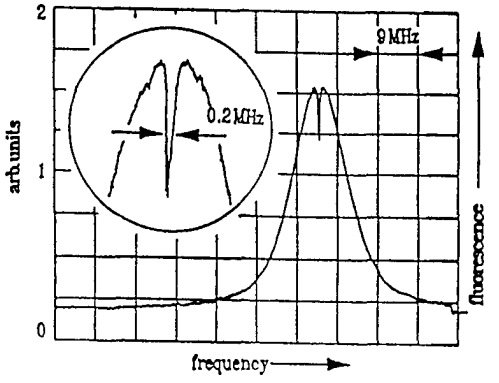


Fig.10 A Doppler free resonance in a Sr atomic beam obtained with the excitation by two running waves.

When the magnetic field is aligned with the E vector of the laser field the σ transitions are prohibited and only the Δm=0 transition remains. The contrast of the Doppler free resonance with respect to the Doppler line is about 27 percent and the line width is 200kHz, where the angle broadening makes the main contribution to the observed linewidth.

As the amplitude of the frequency scan is increased, the resonances corresponding to ^{86}Sr, and the F=9/2-F'=7/2 hyperfine component of the odd ^{87}Sr isotope are displayed. The Doppler-free resonances on top of each are visible Fig.11.

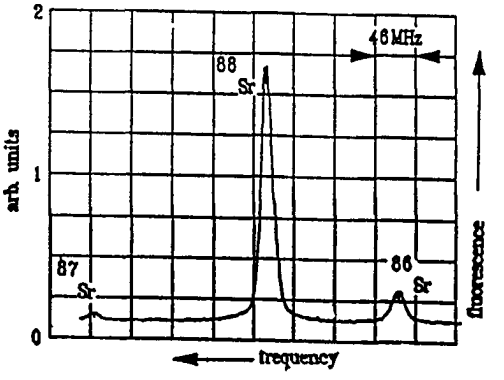


Fig.11 Resonances corresponding to the ^{86}Sr,^{88}Sr and of the F=9/2-F'=7/2 hyperfine component of ^{87}Sr isotopes

The isotope shifts are measured to be 155MHz for ^{86}Sr and 206MHz for the hyperfine component of ^{87}Sr, with respect to ^{88}Sr. Each of them reveals a Doppler free structure.

CONCLUSIONS

We have built a high resolution diode laser spectrometer and checked its performance against the Ca intercombination line in spectroscopic experiments with thermal and laser cooled Ca atoms. A resolution of 1.5kHz would lead to an accuracy of 10^{13} of a frequency standard. We have demonstrated a possibility to use an all diode laser system with an ultimate stability for enhanced resolution paving a way to the development of a highly precision, low cost, transfer standard in the visible. Advantages in diode laser frequency doubling techniques allows using an all diode laser frequency standard based on cooled atoms..

We have performed spectroscopy of Sr atoms in a thermal atomic beam with an eye towards using the Sr intercombination line as an optical frequency reference in the visible.

REFERENCES

1. Report of 81st meeting of the Comite International des Poids et Mesures, Bureau International des Poids et Mesures, Metrologia 1993/94, 30 523-541, J.L.Hall ,M.Zhu,and P.Buch, 1989, J.Opt. Soc.Am., 6, 2196,

2. R.L.Barger, J.C.Bergquist, T.C.English, D.J.Glase, 1979, Appl. Phys. Let., 34, 850

3. J. Helmcke, A.Morinaga, J Ishikawa, F.Riehle, 1989, IEEE Trans. Instr. Meas., 38, 524

4. A.S. Zibrov, R.W. Fox, R. Ellingsen, C.S. Weimmer, V.L. Velichansky, G.M. Tino, L.W. Hollberg, 1994, Appl. Phys. B, 59, 327

5. A. Celikov, F.Riehle, V.Velichansky, and J. Helmcke, 1994, Opt. Com. 107 54

6. T. Kurosu, J. Ishikawa, N. ito, R.W. Fox, Photonics West, SPIE, 1995 p.2378

7. J. Helmcke, J.J.Snyder, A. Morinaga, F. Mensing, and M.Glaser, 1987, Appl. Phys., B 43, 85

8. G.Santarelli, A.Clarion, S.N.Lea, G.M.Tino, 1994, Opt. Com. 104, p. 339

9. H.R.Telle, D.Meschede, T.W.Hansch, Optics Lett. 15, (1990), 532

10. H.Telle, H.Li, Electron. Lett. 26, (1990), p.858

11.V.L. Velichansky, A.S.Zibrov, V.S.Kargopol'tsev,

V.I.Molochev, V.V.Nikitin, V.A.Sautenkov, G.G.Kharisov, D.A.Turikov, 1978,Sov., Tech. Phys. Lett., 4, 438

12.E.M. Belenov, V.L. Velichansky, A.S. Zibrov, V.V. Nikitin, V.A. Sautenkov, A.S.Uskov, 1983, Sov. J. Quantum Electron., 13, 792

13.A.M Akulshin, V.V. Nikitin, V.A. Sautenkov, V.V. Vasilev, V.L.Velichansky, E.K.Yurkin, A.S. Zibrov, in: Frequency Standards and Metrology, Proceedings, of the Fourth Symposium, Ancona, Italy, ed. De. Marchi (Springer, Berlin, 1988) p. 236

14.C.H. Shin, and M. Ohtsu, 1990, Optics Lett. 15 1455

15. R.W.P.Drever, J.L. Hall, F.V. Kowalski, J.Hough, G.M. Ford, A.J.Munley, and H.Ward 1983, Appl. Phys. B 31, 97

16 J.Helmcke, S.A. Lee, J.L. Hall, Appl. Optics 21, (1982), p. 1686

.17.R.W. Fox, H.G. Robinson, A.S.Zibrov, N. Mackie, J. Magyar, and L.W. Hollberg, 1992, SPIE Proceedings of the Boston Conference, Vol. 1837, 360

18.F.Riehle, J.Ishikawa, J. Helmcke, 1988, Phys. Rev. Lett. 61, 2092

19.J. Helmcke, D.Zevgolis, B.U.Yen., 1982, Appl. Phys. Lett. B 28

20. R.L. Barger, Opt. Let.6, 1981, 145

21.A. Witte, Th. Kistters, F. Riehle, J. Helmcke, 1992, J. Opt. Soc. Am. B 9 1030

22.Th. Kisters, K.Zeiske, F.Riehle, J.Helmcke, 1994, Appl. Phys.,B 59, 89

23.See for example A. Morinaga, F.Riehle, J. Ishikawa, J.Helmcke

Appl.Phys. B, (1989), 165

24.A.Celikov, A. Akulshin, V. Velichansky, A Zibrov Laser Pysics, to be published.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Characteristics and Performance of a Novel Photonic Oscillator

Steve Yao and L. Maleki
Jet Propulsion Laboratory, California Institute of Technology,
4800 Oak Grove Drive, Pasadena, California 91109

1 Introduction

With the invention of the vacuum tube in the early part of the century, electronic oscillators were soon developed and used in various applications. In these oscillators a time varying flow of electrons across a load is detected as the oscillating signal. The solid state counter-parts of these "valve" oscillators are now pervasive in virtually every application of electronic devices and systems. Despite the widespread use of electronic oscillators these devices, whether of vacuum tube or the solid state variety, are relatively noisy and lack adequate stability for applications where high stability and spectral purity are required. In this paper we introduce a novel photonic oscillator which represents the extension of the electronic oscillator to the field of photonics and shows a potential for spectral purity and frequency stability rivaling and exceeding the best crystal oscillators.

In photonic devices photons perform the same physical function that electrons perform in electronic devices. Thus in the photonic oscillator the oscillation of the intensity of the photons detected with a photodetector provides the reference signal, much in the same fashion that the oscillatory current across a load resistor comprises the signal of the electronic oscillators. Despite this close similarity, significantly different features of the photonic oscillator provide for much improved noise and stability performance, as compared with the electronic oscillator.

In this paper we describe the physical basis of the photonic oscillator, and present preliminary experimental and analytical results for a number of such oscillators developed at JPL[1]. The paper will conclude with a discussion of the applications of the photonic oscillators, and an assessment for its potential stability performance.

2 Physical Description

We will describe the physical basis of the photonic oscillator by drawing the appropriate parallels with the operational basis of the electronic oscillator, which is well understood. In way of a short review, consider the classic Van der Pol oscillator of Figure 1. Here electrons are generated by a cathode in a triode, where they traverse through a potential generated by a grid, and are collected at the anode. The anode current flows through a resonance filter for frequency selection, and is detected as a voltage across a load resistor. Part of the anode signal is detected by an inductor and is fed back on the grid, the function of which is to provide the voltage that modulates the flow of the electron current. The current-voltage characteristics of a typical triode is also shown in the Figure 1.

As is true for all self-sustaining oscillators, a transient signal starts a signal which in the process of feedback generates the self oscillation. The presence of the frequency selecting filter in the circuit ensures that a single frequency of oscillation is sustained and all other frequencies due to the transient are suppressed. Finally, the nonlinear current-voltage characteristics of the triode further ensures the buildup of the fedback signal up to the saturation level. At this point it is important to note three points concerning the operation of the oscillator. First, the losses in the circuit are compensated with the gain of the triode, thus ensuring that the requirement of self sustaining oscillations, *viz*, the loss is less than or equal to the gain, is met. The second point is that the buildup of the oscillations in the circuit depends on the *inphase* superposition of the partial waves in the feedback process. Finally, the Q of such oscillators is determined by the Q of the inductors in the circuit.

In the photonic oscillator light, generated by a laser is modulated by an external electrooptic modulator,

0-7803-2500-1/95 $4.00 © 1995 IEEE

Van der Pol Oscillator

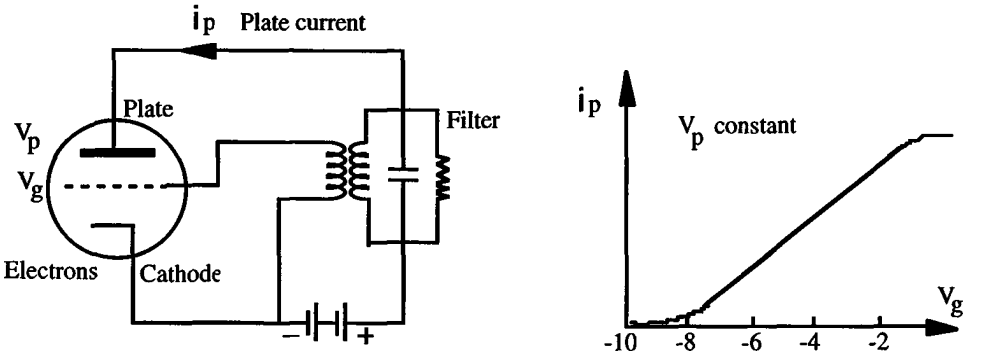


Figure 1.

and passed through a fiber optic link before detection by an optical detector. The electrical signal at the output of the detector is fed back to the modulator through an amplifier and a filter. In such a scheme, shown in Figure 2, the modulator plays the role of the grid, and the long fiber delay is the frequency selective element. The self sustained oscillation of this device are also initiated by a transient, which is frequency selected and amplified before feedback for buildup. The transmission response curve of the modulator is also shown in the figure to complete the parallels drawn between the electronic and the photonic oscillator.

Despite their strong similarities, the photonic oscillator has differences due to the use of photonic components which profoundly influence its operation and performance. To begin with, the fiber delay selects frequencies that preserve a multiple of 2π of phase for a trip around the feedback loop. Delay lines have been studied extensively and are known to be similar in function to cavities, in which a time τ corresponds to the inverse of the decay rate of the amplitude of a mode. In a real microwave cavity, however, losses are dispersive, so that different modes at different frequencies have different Q's. The same is true for electrical delay lines, where dispersion limits the highest Q's to a narrow range of frequencies.

By contrast in the photonic oscillator the loss mechanisms only affect the optical carrier, and are thus fixed in value for all frequencies of the intensity modulation. In this fashion the fundamental frequency ω and all its 2π multiples survive with the un-degraded Q. In fact the function of the filter shown in Figure 2 is to limit the oscillation frequency to a single mode.

Finally, it is known that the length of the delay is proportional to the degree of spectral purity obtained in a delay line oscillator. This fact, which will be verified below, is used to advantage in the photonic oscillator where many kilometers of the optical fiber may be used with little loss of power. Typically, the loss of single mode fibers at 1330 nm is less than 0.3 dB/km, an extremely low value as compared to electrical delay lines.

The photonic oscillator can be represented by a simple functional block diagram shown in Fig. 3. It is a six-port device, with both optical and electrical injection ports, both optical and electrical output ports, and two voltage controlling ports. One of the controlling ports is simply the bias port of the electrooptic (E/O) modulator and the other one is connected to a fiber stretcher for controlling the loop length.

Figure 2

As will be shown below, the two injections ports can be used to injection lock the photonic oscillator to a reference source either optically or electrically. The two output ports provide outputs with an RF carrier in both optical and electrical forms. Finally, the two controlling ports can be used to tune the oscillation frequency and to make the photonic oscillator a voltage controlled oscillator (VCO). The six ports collectively make interfacing the oscillator with photonic or electronic systems very simple.

3 Oscillation Characteristics

We now provide an analysis of the characteristics of the photonic oscillator. We may determine the oscillation threshold of the oscillator by comparing the loss and the gain in the circuit. The open loop gain of the photonic oscillator is equivalent to the electrical (RF) power gain of an externally modulated photonic link. The expression for this gain has been derived elsewhere [2] and is given by:

$$G^2 = G_a^2 \pi^2 \frac{I_p^2 R_L}{V_\pi^2 / R_m} \qquad (1)$$

where G is the open loop voltage gain, G_a is the gain of the amplifier, I is the photocurrent in the detector, R_L is the load resistor at the detector output, and R_m and V_π are the input impedance and the halfwave voltage of the modulator, respectively. Here we have assumed an E/O modulator of Mach-Zehnder type biased at quadrature. The oscillation condition of the photonic oscillator then requires that the open loop gain, G, be larger than 1. Thus if we set R_L=R_m=R, this condition together with the above equation yields:

$$IRG_a \geq V_\pi / \pi. \qquad (2)$$

Equation 2 immediately points to an important feature of the photonic oscillator: The presence of an amplifier is not a necessary condition for oscillation. We can eliminate the amplifier, i.e. set G=1, and as long as IR$\geq V_\pi/\pi$ is satisfied, the oscillation condition is met. This property is extremely important since it provides for the elimination of the flicker noise of the amplifier from the oscillator noise budget. The amplifier noise floor is usually the ultimate limitation in the noise floor of oscillators. The condition of oscillation without an amplifier in the loop is practical when we consider that for a modulator with V_π=3.14 volts and impedance of 50 Ω, a photocurrent of 20 mA will be required. This is achievable with high power detectors and small solid state lasers commercially available today.

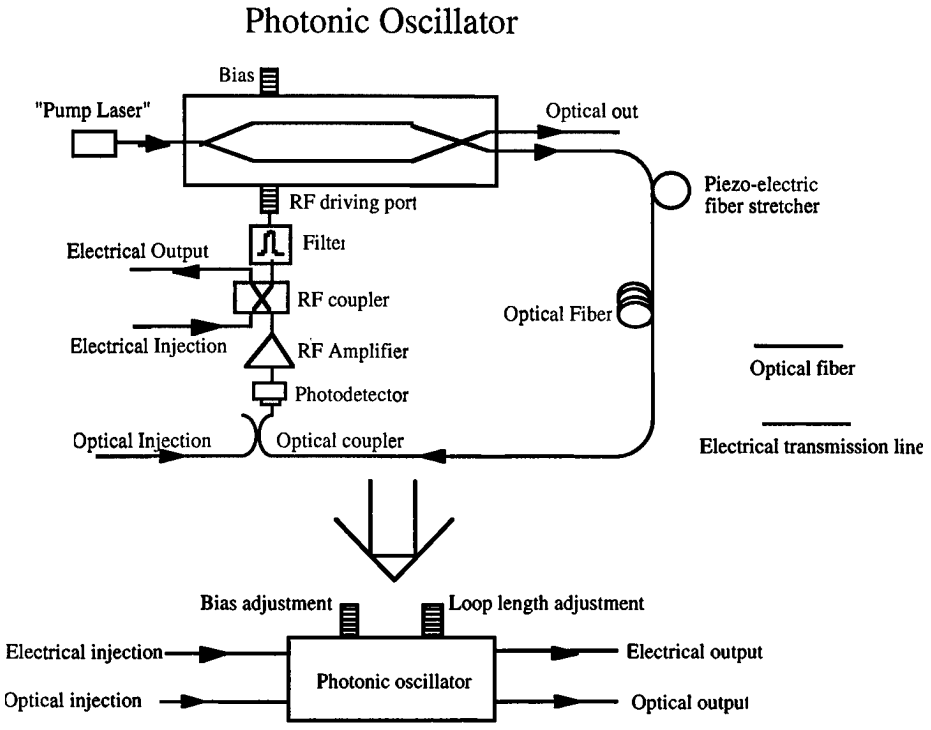


Figure 3

4 Noise Considerations

An example of the signal with one of the first photonic oscillator demonstrated at JPL is shown in Figure 4. The oscillation frequency of 9.22 GHz was limited to the frequency response of the modulator used. The lower diagram in the figure compares the spectrum of a free running photonic oscillator at 100 Mhz to that of a signal generated with a commercial synthesizer (Hewlett-Packard 8656 A). It is clear from the data that the photonic oscillator output is qualitatively less noisy than the synthesizer. This demonstration was encouraging enough to prompt us to develop a noise model for the photonic oscillator.

The noise spectrum of the oscillator may be calculated by considering the circulating partial waves around the feed back loop, and the influence of the noise as the power builds up. A simplified picture may be obtained by referring all noise, consisting of the thermal noise, the shot noise of the detector, and the laser noise to a

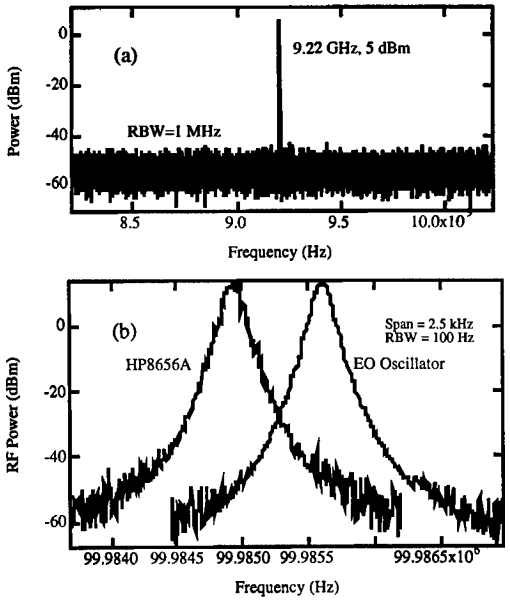


Figure 4.

noise power at the output of the photodetector. The signal V(t) at this point is given by $V_0 e^{i\omega t}$. Upon circulation in the loop, and with fiber delay time of τ, the signal is:

$$V(t) = A(\omega)V(t-\tau), \tag{3}$$

where A includes the contributions of noise sources. Considering the gain of the amplifier upon the circulation of the partial waves in the loop, the final signal, V_{out} is given by a sum :

$$V_{out} = GV_0 \sum_{n=0}^{\infty} A^n e^{i\omega(t-n\tau)}. \tag{4}$$

Note the approximation of allowing n to go to infinity. Based on this scheme, the noise power spectrum may be shown [3]to be given by:

$$S(f) = \frac{\delta}{(\delta/2\tau)^2 + (2\pi f\tau)^2}, \tag{5}$$

where the noise power, S, is a function of the frequency offset, f, and δ=ρG/P, with P as the oscillating electric power, and ρ the input noise including all contributions of the thermal noise, shot noise and the laser relative intensity noise (RIN) terms. For frequencies centered at f_0 it can be shown that the width of the spectrum, Δf, is equal to

$$\Delta f = \frac{1}{2\pi}\frac{\delta}{\tau^2}. \tag{6}$$

It is then clear that the analysis predicts that spectral purity of the photonic oscillator is independent of the oscillation frequency, and improves with the inverse of the square of the delay time, τ.

Figures 5, 6 show comparisons of experimental data with this analysis. They clearly demonstrate that the spectrum improves with the power, as well as the delay, by the predicted factor of 20 dB/decade, and is independent of frequency for the measured values of 100, 200, and 800 MHz.

5 Oscillator Stabilization

As mentioned briefly in the introduction, the oscillation frequency of the photonic oscillator can be tuned for VCO operation. The frequency tuning may be acconplished by changing the loop

Phase noise as a function of oscillation frequency

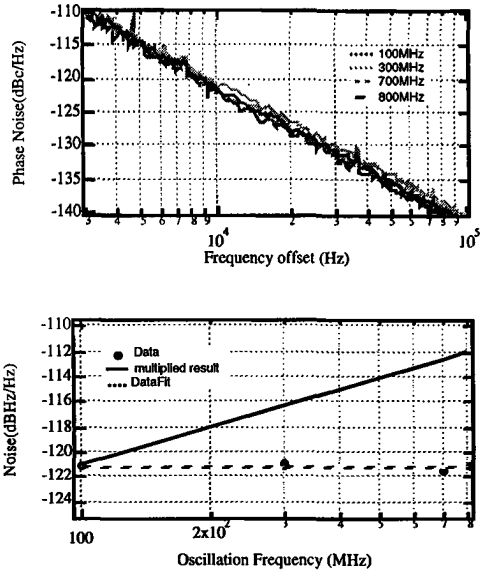
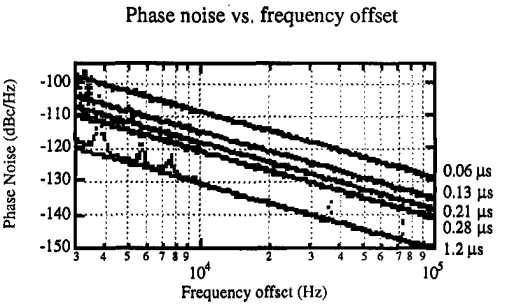


Figure 5

length using a piezo-electric stretcher, or by adjusting the bias voltage. The frequency change Δf is given by $\Delta f = -f_0 \Delta L/L$, where L is the loop length, ΔL is the loop length change, and f_o is the nominal oscillation frequency. However, the tuning sensitivity (Hz/volt) is expected to be small, and the technique somewhat unwieldy due to its mechanical nature.

The oscillation frequency can also be tuned by changing the bias voltage of the E/O modulator. Fig. 8 shows experimental data for variations of the oscillation frequency as a function of the bias voltage. As shown by this data, the frequency tuning of the oscillator is linearly proportional to the bias voltage, with a slope of 38.8 kHz/volt. In this experiment, the output power of the oscillator remained relatively unchanged in a wide voltage range, as shown in Fig. 8. This result is significant because it provided a simple way to tune the oscillation frequency with high sensitivity, and is instrumental for realizing a phase locked loop (PLL) using the novel photonic oscillator, as will be discussed below.



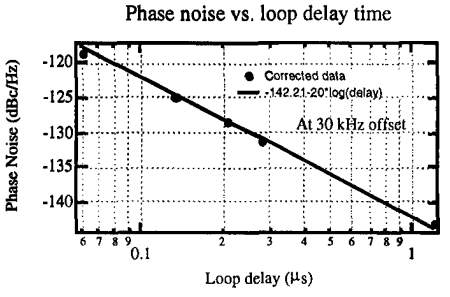
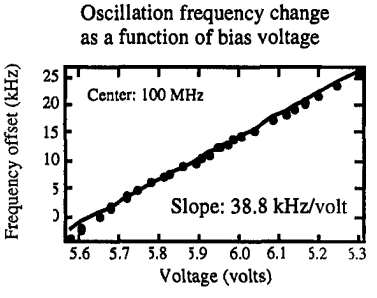


Figure 6



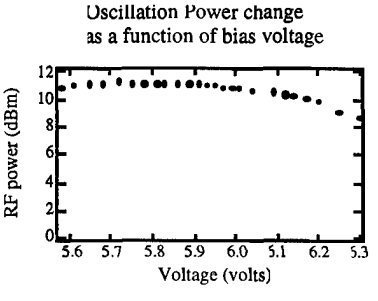


Figure 8

It is well known that the stability of an oscillator may be improved by injection locking its signal with the signal from another oscillator.

This approach may also be used with the photonic oscillator. However, a unique aspect of the photonic oscillator is that it can be injection locked by either an optical signal or an electrical signal, as shown in Fig. 3.

The ability to injection lock an oscillator optically is important because it allows remote synchronization. This function is critical for high frequency RF and microwave systems which require many oscillators locked to a single master, as in a phased array radar. Another advantage of optical injection locking is that the locking oscillator is electrically isolated from the locked oscillator, eliminating the need for impedance matching between the oscillators. Fig. 9 shows the experimental results of injection locking the photonic oscillator with a maser reference at 100 MHz through the electrical injection port. Similar results are expected for optical injection since the optical injection signal will first be converted to an electrical signal by an internal photodetector in the loop before entering the E/O modulator. As shown in Fig. 9, with an injection power of -5 dBm, the phase noise of the photonic oscillator is almost identical to that of the injecting maser signal.

Note however that the output RF power of the photonic oscillator is 13 dBm, resulting in a gain of 18 dB. As the injection power decreases, the phase noise of the photonic oscillator increases somewhat. However, the output RF power level remains the same and therefore the gain is effectively increased. In these experiments we were able to injection lock the photonic oscillator to a maser reference with an injection power as low as -50 dBm. As expected, the locking range was linearly proportional to the square-root of the injection power, agreeing well with Adler's injection locking theory [4].

Although injection locking is an effective way for synchronizing and stabilizing oscillators, it requires a low noise and high stability source to begin with. Making a high frequency and high stability source itself is a difficult task. By using a novel scheme called self- injection locking to stabilize the photonic oscillator, the photonic oscillator may be made as the frequency reference, as illustrated in Fig. 11. In this scheme, we derive a small portion of the output optical signal from the E/O oscillator and send it through a long fiber delay line.

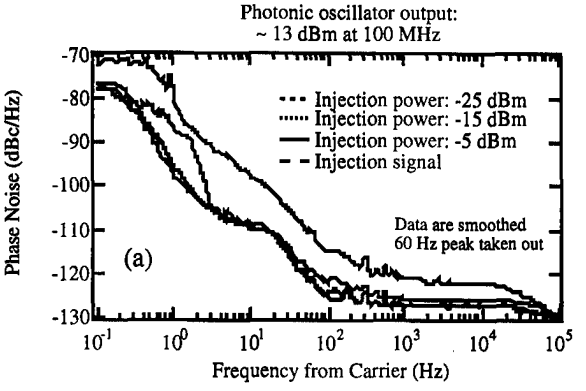


Figure 9. Injction locking to H-maser

The output from the fiber delay line is then converted to electric signal and is fed back to the RF driving port of the E/O modulator. Note that the open-loop gain of this feedback loop should be kept much below unity so that no self-oscillation can be started. Basically, what we do here is to inject a delayed replica of the photonic oscillator's output back to the oscillator and force the oscillator to lock to its "past". This will prevent the oscillator from changing its frequency and phase, and hence reduce the frequency and phase fluctuations. The frequency stability of the oscillator then is expected to be proportional to the length fluctuation $\Delta L/L$ of the fiber delay line. Fig. 10 is the experimental results showing the efficacy of the self-injection technique in reducing the frequency noise of the photonic oscillator. The length of the delay line used in the experiment is 12 km and the feedback injection RF power is -19.23 dBm.

It is evident that self-injection locking greatly reduced the frequency fluctuations of the E/O oscillator. Further noise reduction is expected if we reduce the length fluctuation of the fiber delay line by placing it in a temperature controlled environment and isolating it from acoustic vibrations. More experiments are under way to further reduce the noise of the self-injection locked photonic oscillator.

The unique optical output of the photonic oscillator may be employed yet in a different manner to stabilize it. With the optical output, we can make a self-phase locked loop as shown in Fig. 10. Similar to self-injection locking described earlier, the self-phase locked loop forces the oscillator to have reduced fluctuations.

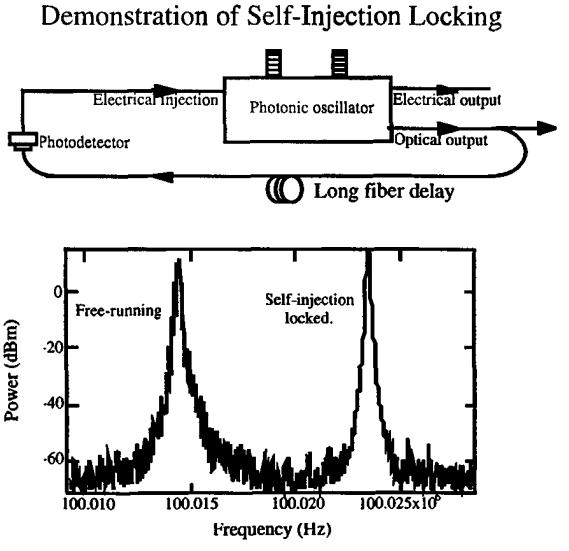


Figure 11

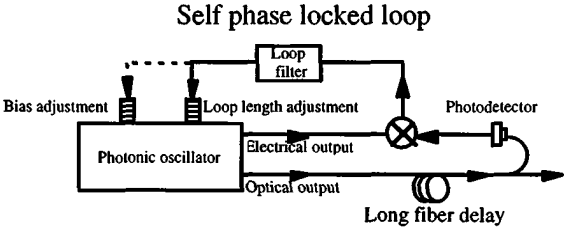


Figure 10

The use of the delay line discriminator technique to stabilize an oscillator is well known in the field. But to effectively stabilize an oscillator, a delay of many kilometers is needed and thus has been considered impractical before the emergence of the photonic technology. At JPL we have previously demonstrated the use of a fiber optic delay line to stabilize a traditional VCO, and obtained excellent results.[7] However, in that setup, the fiber optic delay line included a laser transmitter to convert the VCO's electrical output into optical signal and then transmit the optical signal through a few kilometers

of fiber Since the photonic oscillator automatically contains an optical output, it is ideal for using the fiber delay line technique to stabilize itself without the need of electrical to optical signal conversion. Consequently, the device is simple, low loss, and less expensive. Experiments are underway to extend the technique of self-phase locking to obtain ultra-stable performance with the photonic oscillator.

6 Applications

Because of its great simplicity and flexibility, the photonic oscillator may be applied to a large variety of important problems in frequency control. These applications may be categorized as follows. With simple components, the photonic oscillator offers moderate performance, comparable to commercial synthesizer outputs, at a small, low mass, and low cost package. A low mass version of the oscillator based on optical integrated components will be of great interest for spacecraft applications. Its capability to be used as a VCO combined with its potential for stabilization make it a very attractive candidate as the LO for ultra-stable atomic standards, such as the trapped ion standard and the laser cooled cesium fountains. Its low noise at very high frequencies, and its immunity to degradation with frequency multiplication make it desirable for microwave frequency generation and synthesis. This is an important application for a number of functions required for high resolution, and high agility radar. The frequency synthesis properties of the photonic oscillator may be used to develop comb generators, multipliers, and dividers, all of which are useful for the development of an all-photonic frequency synthesizer. Finally its ability to amplify injected signals may be used in clock and carrier recovery, as well as other related applications. Most of these functions have already been demonstrated in our laboratory [5, 6], while others are under development.

In conclusion, we have developed a novel oscillator based on the photonic components. The oscillator has many desirable features including low noise, high stability, and for certain configurations, low mass and cost. The oscillator may readily be applied to a number of problems in frequency control including reference generation, frequency synthesis, and all applications requiring VCO's. Further development of the photonic oscillator holds the potential for much improved performance.

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with National Aeronautics and Space Administration.

References

[1] X. Steve Yao and Lute Maleki, 'High frequency optical subcarrier generator,' Electron. Lett. 30 (18), pp1525-1526 (1994)

[2] X. S. Yao and L. Maleki, 'Influence of an externally modulated photonic link on a microwave communications system," TDA Progress Report 42-117, Jet Propulsion Laboratory, pp16-28 (1994).

[3] X.Steve Yao and Lute Maleki, *tobepublished.*

[4] R. Adler, 'A study of locking phenomena in oscillators,' Proc. IRE, 34, pp351-357 (1946)

[5] X. Steve Yao and George Lutes, 'A high speed photonic clock and carrier regenerator,' TDA Progress Report , Jet Propulsion Laboratory, 42-121, pp202-210 (1995).

[6] X. Steve Yao and Lute Maleki, A novel photonic oscillator, To be published in TDA Progress Report, Jet Propulsion Laboratory.

[7] R. Logan, L. Maleki and M. Shadaram, "Stabilization of oscillator phase using a fiber optic delay-line," in Proceedings of the 45th Annual Frequency Control Symposium, May 29-31, Los Angeles, CA, 1991.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

New Tests for Variations of the Fine Structure Constant

John D. Prestage, Robert L. Tjoelker, and Lute Maleki

California Institute of Technology, Jet Propulsion Laboratory
Frequency Standards Laboratory
4800 Oak Grove Drive, Bldg 298
Pasadena, California 91109

Abstract

We describe a new test for possible variations of the fine structure constant, α, by comparisons of rates between clocks based on hyperfine transitions in alkali atoms with different atomic number Z. H-maser, Cs and Hg^+ clocks have a different dependence on α via relativistic contributions of order $(Z\alpha)^2$. Recent H-maser vs. Hg^+ clock comparison data improves laboratory limits on a time variation by 100-fold to give $\dot{\alpha}/\alpha \leq 3.7\text{x}10^{-14}$/yr. Future laser cooled clocks (Be^+, Rb, Cs, Hg^+, etc.), when compared, will yield the most sensitive of all tests for $\dot{\alpha}/\alpha$.

Introduction

Since Dirac's large number hypothesis (LNH)[1], the search for a time variation of the fundamental constants has been the subject of much work[2]. Dirac noticed that the ratio of the electrostatic to gravitational forces between an electron and proton (~$2\text{x}10^{39}$) was close to the age of the universe expressed in units of the light transit time across the classical electron radius, $R_e/c = e^2/(m_e c^3)$. He conjectured that these two very large quantities were proportional, hence, the ratio $e^2/(Gm_p m_e)$ would vary with the age of the universe. A fractional change $\delta G/G \approx -5\text{x}10^{-11}$/year would result assuming a universe $2\text{x}10^{10}$ years old. Teller and other authors [2,3] have postulated a relationship for the fine structure constant $\alpha^{-1} \sim \log[hc/(Gm_e^2)]$ where $[hc/(Gm_e^2)]^{1/2}$~(electron Compton wavelength) /(Planck Length). Taken with the Dirac hypothesis of a time varying G, α may vary $\dot{\alpha}/\alpha \sim \alpha(\delta G/G) \sim 3.6\text{x}10^{-13}$/yr.

The fine structure constant $\alpha = 2\pi e^2/hc$ characterizes the strength of the electromagnetic force which holds atoms together. Biology, chemistry, solid state physics, etc., are in a fundamental way, determined by the size of $\alpha(\approx 1/137)$. Because α involves the elementary unit of charge, e, the speed of light, c, and Planck's constant, h, it combines electromagnetism, relativity, and quantum mechanics. It is *the* primary expansion parameter used in quantum electrodynamics (QED) to accurately model certain measurable properties of the electron to the parts per billion level. The value for α, however, is not predicted by QED but should be result of a unified theory of the four basic interactions — electromagnetic, gravitational, strong and weak. Such grand unified theories are being developed and, surprisingly, many have cosmological solutions where α and other "constants" actually change over time. For example, the string theory prediction from reference [5] gives $\dot{\alpha}/\alpha \sim 10^{-3}(\delta G/G) \sim 10^{-17}$/yr or smaller.

Variation of the non-gravitational constants is forbidden in General Relativity and other metric theories of gravity, where gravitational fields are

0-7803-2500-1/95 $4.00 © 1995 IEEE

described as a geometrical property of space-time. The equivalence principle forms the basis for all metric theories and requires local position invariance: in local freely falling frames the outcome of any local non-gravitational test experiment is independent of where and when in the universe it is performed [4]. A changing fine structure constant, α, as predicted in some cosmological string theories [5], would violate the equivalence principle signalling the breakdown of gravitation as a geometrical phenomena and, as we show in this paper, would lead to a drift in the relative frequencies of H-masers, Rb, Cs, Hg^+, etc. clocks. Clock comparisons between the next generation of ultra-stable laser-cooled clocks now being developed may be of sufficient sensitivity to reveal equivalence principle violations at the boundary of gravitational and quantum physics.

Several analyses of paleontological, geophysical and astronomical data were made apparently ruling out the LNH variation [2] though there have been conflicting claims for a measured variation of the gravitational constant[6]. The paleontological arguments were based upon the realization that even a small departure of the gravitational constant ,G, from the present day value would make the earth inhospitable to life. Arguments of this sort have arisen largely as a response to Dirac's LNH and have led to the development of the Anthropic Cosmological Principle (ACP). Accordingly [7], the large number ratio (LNR) values are not a consequence of the above proportionality postulated by Dirac but rather, the present day LNR values are one of a relatively small subset (of all possible LNR values) which will lead to the development of observers, i.e., physicists, astronomers, etc.

The experimental search for a temporal variation of α is divided into what might be called cosmological and modern measurements. For example, a stringent limit on α variation follows from an analysis of isotope ratios $^{149}Sm/^{147}Sm$ in the natural Uranium fission reaction that took place some $2x10^9$ years ago at the present day site of the Oklo mine in Gabon, West Africa[2,8]. This ratio is 0.02 rather than 0.9 as in natural samarium from the neutron flux onto ^{149}Sm during the uranium fission. It is thus deduced that the neutron capture cross-section in ^{149}Sm has not changed significantly in $2x10^9$ years from its present day value. Recent modelling [8] of this process has relaxed the original stringent limits by 100-fold to $\dot{\alpha}/\alpha \leq 10^{-15}$/yr. This limits the integrated change in α over the cosmological time period of $2x10^9$ yrs. In a similar way, astronomical measurements of multiple spectral lines (with different dependence on α and other atomic constants) from a common source with a large cosmological red shift, have been used to place limits on variations of α over cosmological time periods of $\dot{\alpha}/\alpha \leq 4x10^{-12}$/yr[9].

Modern or laboratory measurements are based on clock comparisons with ultra-stable oscillators of different physical make-up such as the superconducting cavity oscillator vs. Cesium hyperfine clock transition [10] or the Mg fine structure transition vs. the Cesium hyperfine clock transition[11]. Unlike the results inferred from phenomena taking place over cosmological time scales, the clock comparisons are repeatable and are of duration months to years. These measurements rely on the ultra-high stability of the atomic standards and set limits a few orders of magnitude less stringent than the cosmological measurements [2,8]. The modern clock comparisons are really complementary to the cosmological determinations because they place a limit on a present day variation of α[12].

The string theory prediction [5] for a temporal variation of the fundamental constants has provided an incentive for improved tests of the constancy of α. This paper describes a new method for determining limits on the variation of α by comparing rates for clocks based on atoms of different atomic number Z. The method is based on the increasing importance of relativistic contributions to the hyperfine energy splitting as atomic number Z increases in the group I alkali elements and alkali-like ions. The contribution is a function of αZ which grows faster than $(Z\alpha)^2$ for the heavier atoms and thus differs for hydrogen

(Z=1), beryllium ion (Z=4), rubidium (Z=37), cesium (Z=55), and mercury ion (Z=80). Any variation in α, whether a cosmological time variation or a spatial variation via a dependence of α on the gravitational potential[13], will force a variation in the relative clock rates between any pair of these clocks.

Hyperfine Structure in Alkali Atoms

We begin by comparing the theoretical expressions for the hyperfine splitting (hfs) in hydrogen and the alkali atoms and ions. All continuously operated microwave atomic frequency standards (H, Rb, Cs, and Hg^+) are based on transitions between ground state hyperfine levels determined by the interaction of a nuclear magnetic moment with the magnetic moment of an $S_{1/2}$ state valence electron. The hydrogen hfs is the simplest and to lowest order in α and m_e/m_p, the splitting used as the clock transition in the H-maser is

$$a_s = \frac{8}{3}\alpha^2 g_p \frac{m_e}{m_p} R_\infty c$$

where g_p is the proton g factor, m_e and m_p are the electron and proton masses, and $R_\infty c$ is the Rydberg constant in frequency units.

The theory of the hyperfine splitting in alkali atoms and ions is not so well developed as that for hydrogen but much work has been done and the theoretical expressions predict the splittings for the Cs and Hg^+ clock transition frequencies to the 1% level[14]. The full expression for the hyperfine interaction constant A_s [14,15] is

$$A_s = \frac{8}{3}\alpha^2 g_I Z \frac{z^2}{n_*^3}\left(1-\frac{d\Delta_n}{dn}\right) F_{rel}(\alpha Z)(1-\delta)(1-\epsilon)\frac{m_e}{m_p} R_\infty c$$

The transition frequency between the I±1/2 states is $(I+1/2)A_s$, where I is the nuclear spin angular momentum quantum number.

This expression is composed of several factors. The value of the valence electron wavefunction at the nucleus, obtained by solving the non-relativistic Schrodinger equation, is given by the semi-empirical Fermi-Segrè formula[16]

$$\psi_n^2(0) = \frac{Zz^2}{\pi a_0^3 n_*^3}\left[1 - \frac{d\Delta_n}{dn}\right]$$

where Z is the atomic number, z is the net charge of the remaining ion following the removal of the valence electron, n_* is the effective quantum number chosen to match the measured energy levels, E_{n*}, for the alkali atom according to the Rydberg formula $E_{n*} = -z^2 Ry/n_*^2$. $\Delta_n = n - n_*$ is the quantum defect for the n'th state. The term $(1-\delta)$ is the correction for the departure of the atomic potential from pure Coulomb as the electron enters the relatively large high Z nucleus with $\delta \approx 4\%$ for Cs and 12% for Hg [14]. $(1-\epsilon)$ is a similar correction for the finite size of the nuclear magnetic dipole moment with $\epsilon \approx 0.5\%$ for Cs and 3% for Hg [14].

The Casimir correction factor $F_{rel}(\alpha Z)$ [14,15,17] is obtained when the relativistic wave equation is solved to evaluate the electron wavefunction in the vicinity of the nucleus. For an $S_{1/2}$ state electron $F_{rel}(\alpha Z) = 3[\lambda(4\lambda^2-1)]^{-1}$ where $\lambda = [1-(\alpha Z)^2]^{1/2}$ showing F_{rel} is a strong function of α for high Z nuclei. For $\alpha Z \ll 1$, $F_{rel} \approx 1+11(\alpha Z)^2/6$ but with heavier atoms this approximation breaks down since for Cs, $F_{rel} = 1.39$ and for Hg , $F_{rel} = 2.26$.

Relative Clock Rate Sensitivity to $\dot{\alpha}/\alpha$

A time variation in α will therefore induce a change in the frequency of an H-maser relative to the frequency of a heavy atom hfs transition according to

$$\frac{d}{dt}\ln\left(\frac{A_{alkali}}{a_{hydrogen}}\right) = \alpha \frac{d}{d\alpha}\ln(F_{rel})\left(\frac{1}{\alpha}\frac{d\alpha}{dt}\right)$$

We have assumed the integers z, and Z remain constant. Supposing that α changes, there will be a corresponding change in the effective quantum number n_* since it is determined by the Rydberg levels of the valence electron. However, because

$n_*^2 \sim E_n/(z^2 Ry) \sim (1+\text{higher order in } (z\alpha)^2)$ its changes are small. The finite nuclear volume correction δ does contain terms of order $(\alpha Z)^2$ but its overall sensitivity to α is $\leq$ 10% of that of F_{rel} and is negligible.

The above ratio between hyperfine transitions in different atoms contains no electron to proton mass ratio and the nuclear g-factors enter as a ratio unlike the clock comparisons described in references [10,11]. The above "sensitivity factor" is re-written

$$\alpha \frac{d}{d\alpha} \ln (F_{rel}) = (\alpha Z)^2 \frac{12\lambda^2 - 1}{\lambda^2(4\lambda^2 - 1)} \equiv L_d F_{rel}(\alpha Z)$$

The sensitivity to α variations, $L_d F_{rel}(\alpha Z)$, is plotted against atomic number Z in Fig. 1.

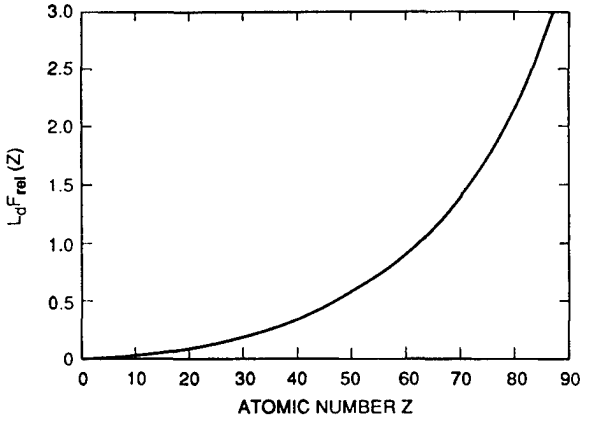


Figure 1: The function $L_d F_{rel}(Z)$ plotted against atomic number Z.

By analogy with a Dirac particle, the ratio g_I/g_p (g values of a bound nucleon to a free nucleon) is relatively insensitive to α. The nuclear g factors are defined as a ratio of the measured nuclear magnetic moment to the nuclear magneton $(eh)/(2m_p c)$ and are determined primarily by the strength of the strong interaction. For an electron bound to a nucleus of charge Z there is a relativistic mass contribution to the electron g-factor of order $(\alpha Z)^2$ [15]. By contrast, the strong force binding a nucleon in a nucleus remains relatively constant with increasing atomic number unlike the electromagnetic binding of an electron to a nucleus. We therefore assume there is no corresponding contribution to the nuclear g-factor ratio which grows with atomic number Z as strong as the $(\alpha Z)^2$ dependence of F_{rel}.

As above, for the comparison of two clocks each based on a transition in different alkali atoms with Z >1, there will be a relative drift in rates

$$\frac{d}{dt} \ln \frac{A_{Alkali1}}{A_{Alkali2}} = (L_d F_{rel}(Z_1) - L_d F_{rel}(Z_2)) \frac{1}{\alpha} \frac{d\alpha}{dt}$$

Table 1 shows the size of the sensitivity $L_d F_{rel}(Z_1) - L_d F_{rel}(Z_2)$ for various clock intercomparisons that might be used to detect a temporal variation in α (or spatial with d/dt replaced by d/dU where U is the solar gravitational potential[13]). A larger sensitivity would cause a larger clock rate difference given a non-zero value for $\dot{\alpha}/\alpha$. Alternatively, given a variation in α, the six distinct drift rates of Table 1 would predict a clear signature which would be useful in discriminating against systematic errors that might show up in any single intercomparison. For example, the Cs vs. Hg^+ rate difference should be $1.4 \div 0.74 \approx 1.9$ times greater than the H-maser vs. Cs rate difference, etc.

	H	Rb	Cs	Hg^+
H	0	0.3	0.74	2.2
Rb	-0.3	0	0.45	1.9
Cs	-0.74	-0.45	0	1.4
Hg^+	-2.2	-1.9	-1.4	0

Table 1: The sensitivity of various clock rate comparisons to a variable fine structure constant. The entry is $L_d F_{rel}(Z_1) - L_d F_{rel}(Z_2)$ and converts fractional changes in α to a drift in clock rates between the two given clocks. For example, if $\dot{\alpha}/\alpha = 10^{-14}$/yr, a frequency drift of 2.2×10^{-14}/yr between an H-maser and an Hg^+ clock would result.

Limits on $\dot{\alpha}/\alpha$ from Clock Comparisons

Several clock comparisons have been made which can be used to search for a variation of α. Long term comparisons of Cs to H-maser clocks are carried out in the generation and maintainence of the worldwide atomic timescale (TAI). A recent comparison carried out over a one year period between two cavity auto-tuned active H-masers and the primary cesium standards, CS1 and CS2, (at PTB in Braunschweig, Germany) showed a $1.5\text{x}10^{-16}$/day relative frequency drift[18]. Similar clock comparisons have been made at the US Naval Observatory[19] with comparable clock rate drifts. Since $L_dF_{rel}(55)=0.74$ we find $\dot{\alpha}/\alpha \leq 1.5\text{x}10^{-16}/\text{day} \div 0.74 = 7\text{x}10^{-14}/\text{yr}$.

We have developed [20,21] an ultra-stable frequency standard based on Hg^+ ions confined to a linear ion trap, and have recently completed a 140 day clock rate comparison [to be published] between it and a cavity tuned H-maser [22]. In that comparison, a limit of $2.1(0.8)\text{x}10^{-16}$ /day was established for the frequency drift between these two long term stable clocks. The Allan deviation of this clock comparison is shown in Figure 2. This is a more sensitive test for α variations than the Cs vs. H-maser comparison since $L_dF_{rel}(80) = 2.2$ and establishes an upper bound $\dot{\alpha}/\alpha \leq 3.7\text{x}10^{-14}$ /yr.

This Hg^+ vs H-maser limit represents a 10 fold improvement over the recent limit[11] and rules out the LNH variation of α ($\sim 3.6\text{x}10^{-13}$/yr) discussed in the introduction. It should be noted that these results are the *only present day laboratory* tests with enough sensitivity to rule out such variations. The limits established in ref [11] on an α variation ($\leq 2.7\text{x}10^{-13}$/yr) were inferred from astrophysical limits placed on $\alpha^2 g_p m_e/m_p$ [9]over a time interval of almost 10^{10} yrs.

The Hg^+ vs. H-maser results presented here represent a 100-fold improvement over the best laboratory limits ($\leq 4\text{x}10^{-12}$ /yr) established in the superconducting cavity vs. Cs frequency comparisons of ref [10]. This improvement follows from the very good long term stability of the atomic Hg^+ and H-maser clocks, with relative drift $\sim 10^{-16}$/ day, as compared to the superconducting cavity oscillator where instrumental drifts can lead to frequency drifts of a few parts in 10^{14}/day [10].

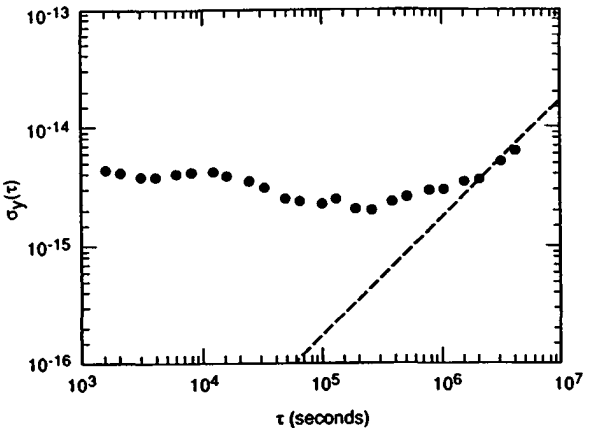


Figure 2: The measured Allan deviation for the 140 day H-maser vs. Hg^+ clock comparison. The dashed line at 45° is the linear drift estimate $2(\pm 1)\text{x}10^{-16}$/day.

Summary

We have developed a new method for detecting variations of the fine structure constant, α, by examining relative drift rates between atomic clocks which are continuously monitored in time scales in several labs worldwide. We have searched for such drifts in a clock comparison between Hg^+ and H-maser clocks and improved modern day limits on an α variation by two orders of magnitude. Further improvements will follow as laser cooled Be^+, Rb, Cs and Hg^+ [23] microwave standards are developed. Comparisons of their clock rates should establish the most sensitive search for any temporal variation of α and may reach a sensitivity approaching the string theory predictions [5]. Finally, this method also shows that comparisons between Cs, Hg^+, Rb, Be^+ and H-maser clocks can be used to improve the complementary search for a dependence of α on the gravitational potential[13].

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] P. A. M. Dirac, *Nature* **139**, 323 (1937)

[2] F. J. Dyson, in *Aspects of Quantum Theory* (Cambridge Univ. Press, 1972), pp. 213-236; in *Current Trends in the Theory of Fields* (AIP, New York, 1983), pp. 163-168; B. W. Petley, *The Fundamental Physical Constants and the Frontier of Measurements* (Adam Hilger Ltd, Boston,1985).

[3] E. Teller, *Phys. Rev.* **73**, 801 (1948); L. D. Landau in *Niels Bohr and the Development of Physics*, ed. W. Pauli (Pergamon Press, London, 1955); B. S. DeWitt, *Phys. Rev. Lett.* **13**, 114 (1964); C. J. Isham, A. Salam, and J. Strathdee, *Phys. Rev. D* **3**, 1805 (1971).

[4] C. M. Will, *Theory and Experiment in Gravitational Physics* (Cambridge University, Cambridge, 1981).

[5] T. Damour and A. M. Polyakov, Nucl. Phys. **B423**, 532-558 (1994).

[6] T. C. Van Flandern in *Precision Measurements and Fundamental Constants II*, NBS Pub. # 617, p. 625 (1981); R. D. Reasonberg, *Phil. Trans. R. Soc.* A **310**, 227 (1983).

[7] J. D. Barrow and F. J. Tipler, *The Anthropic Cosmological Principle*, (Oxford University Press, 1988).

[8] A. I. Schlyakter, *Nature* **25**, 340 (1976); P. Sisterna and H. Vucetich, Phys. Rev. **D41**, 1034 (1990).

[9] J. N. Bachall and M. Schmidt, *Phys. Rev. Lett.* **19**, 1294 (1967); A. M. Wolfe, R. L. Brown, and M. S. Roberts, *Phys. Rev. Lett.* **37**, 179 (1976).

[10] J. P. Turneaure and S. Stein, in *Atomic Masses and Fundamental Constants V* (Plenum, London, 1976) pp. 636-642.

[11] A. Godone, C. Novero, P. Tavella, and K. Rahimullah, *Phys. Rev. Lett.* **71**, 2364 (1993).

[12] B. W. Petley in *Quantum Metrology and Fundamental Physical Constants*, eds. P. H. Cutler and A. A. Lucas, NATO ASI Ser. B. **98** (Plenum, New York, 1983), pp. 333-351.

[13] J. P. Turneaure, C. M. Will, B. F. Farrel, E. M. Mattison, and R. F. C. Vessot, *Phys. Rev. D* **27**, 1705 (1983).

[14] H. Kopfermann, *Nuclear Moments* (Academic Press, New York, 1958), pp. 123-138; H. B. G. Casimir, *On the Interaction between Atomic Nuclei and Electrons* (Freeman, San Francisco, 1963) p. 54.

[15] M. Mizushima, *Quantum Mechanics of Atomic Spectra and Atomic Structure* (W. A. Benjamin, New York, 1970).

[16] L. L. Foldy, *Phys. Rev.* **111**, 1093 (1958).

[17] C. Schwartz, *Phys. Rev.* **97**, 380 (1955).

[18] N. A. Demidov, E. M. Ezhov, B. A. Sakharov, B. A. Uljanov, A. Bauch, and B. Fisher, in *Proc. 6th European Frequency and Time Forum, Noordwijk, NL, 1992* (European Space Agency, Noordwijk, 1992), pp. 409-414.

[19] L. A. Breakiron in *Proc. 25th Annual Precise Time Interval Appl. Meeting* (NASA conference Pub. 3267, 1993), pp. 401-412.

[20] J. D. Prestage, G. J. Dick, and L. Maleki, *J. Appl. Phys.* **66**, p.1013 (1989); J. D. Prestage, R. L. Tjoelker, G. J. Dick, and L. Maleki, *J. Mod. Optics* **39**, pp. 221-232 (1992)

[21]R. L. Tjoelker, J. D. Prestage, G. J. Dick and L. Maleki *Proc. 1994 IEEE Frequency Control Symposium*, pp. 739-743 (1994).

[22] Sigma Tau Standards; Tuscaloosa, AL 35403 USA.

[23] M. E. Poitzsch, J. C. Bergquist, W. M. Itano, and D. J. Wineland, *Proc. 1994 IEEE Frequency Control Symposium*, p. 744 (1994); A. Clairon, et. al., to be published in *IEEE Trans. Instr. Meas.* April, 1995, (special issue on CPEM '94).

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

EXPERIMENTAL TEST OF LOCAL OSCILLATOR–INDUCED PERFORMANCE LIMITATION FOR PULSE–MODE PASSIVE FREQUENCY STANDARDS*

G. J. Dick and D. A. Stowers
California Institute of Technology, Jet Propulsion Laboratory
4800 Oak Grove Drive, Bldg 298
Pasadena, California 91109

Abstract

An experimental test of Local-Oscillator (L.O.) degradation of time–sequential passive oscillator performance was accomplished using the TAC-controlled crystal oscillator described in an accompanying paper. For this quantitative demonstration, an electronic frequency–counter discriminator replaced the trapped ion physics package in a closed–loop experiment. With a 10 second cycle time and 5 second dead time, L.O. limited stability was demonstrated in excellent agreement with theoretical calculations. The results are also in agreement with computer simulations and confirm previous noise analysis for the case of a passive standard with short pulse "Ramsey" interrogation.

Background

A methodology for calculating L.O. induced degradation for pulse-mode atomic frequency standards has been available for some time, representing the first quantitative analysis of this fundamental limitation for any passive atomic frequency standard.[1,2] Passive standards use atomic or ionic transitions with very narrow linewidths to sense and correct the frequency of an ancillary local oscillator. Frequency variability in the L.O. is detected and corrected by counting photon or atoms. Statistical variation of these counts gives rise to a frequency stability (Allan Deviation of fractional frequency $\sigma_y(\tau)$) that improves with increasing measuring time τ as $\sigma_y(\tau) \propto 1/\sqrt{\tau}$. An ideal feedback loop might be expected to show a rapidly improving deviation $\sigma(\tau) \propto 1/\tau$ as it transitions between L.O. performance at shorter measuring times to statistical atomic performance at longer times.

The importance of the L.O. analysis is that it shows that any time-variation in the sensitivity of the feedback signal will alias (down-convert) relatively high–frequency oscillator noise to near zero frequency. The associated $\sigma(\tau) \propto 1/\sqrt{\tau}$ L.O. contribution fundamentally limits and masks the white frequency noise performance of atomic standard itself.

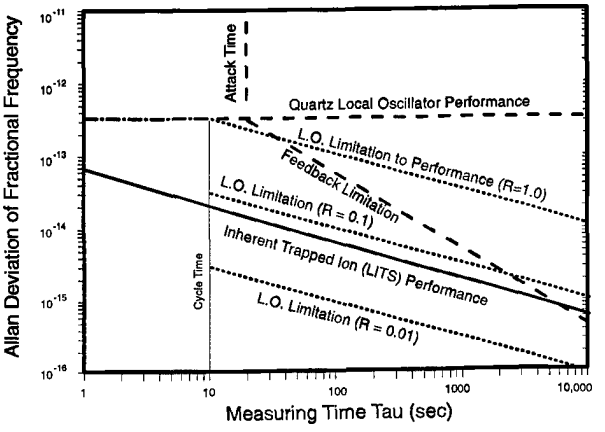


Figure 1: Short– and medium–term frequency stability features for an ion trap standard with cycle time of 10 seconds. Values for **R** are shown in Fig. 2.

The interrelation of these various dependencies is shown in Figure 1 for the JPL mercury–based Linear Ion Trap Standard (LITS) stabilizing a quartz L.O.

*This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

0-7803-2500-1/95 $4.00 © 1995 IEEE

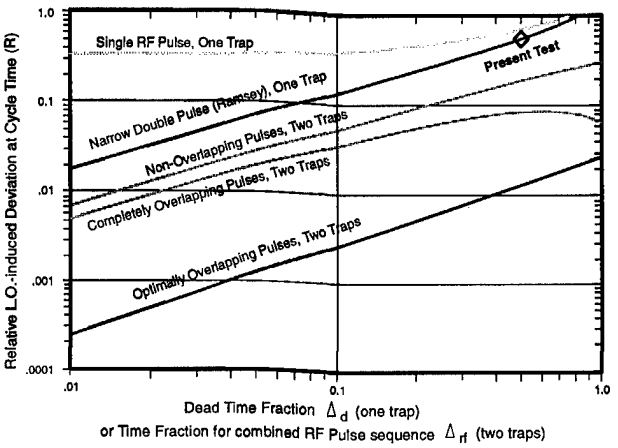


Figure 2: Calculated dependence of dimensionless constant **R** on "dead time" for various interrogation strategies assuming a Flicker–Frequency L.O. **R** describes the L.O.–limited $\sigma \propto \tau^{-}$ deviation for the passive standard (see Fig. 1). A value of $\mathbf{R} = 0.533$ is predicted for the present experiment, corresponding to a dead time fraction of 0.5 and narrow double pulse interrogation.

with an assumed flicker–frequency floor of 3×10^{-13}. The L.O.–induced stability limitation is described in terms of a dimensionless constant **R** which depends on the details of the interrogation strategy, e.g. dead time and rf pulse length. Calculated values for **R** vary from 0.2 to 1.0 for typical interrogation strategies. From Fig. 1 we can see that the performance of this standard will be substantially degraded for these strategies, unless an improved local oscillator can be found.

The top two curves in Figure 2 show the calculated dependence of **R** on "dead time" for single- and double–pulse interrogation for a flicker–frequency L.O.[2] The advantages of a short dead time and of Ramsey interrogation are apparent. For example, if the fractional dead time is only 0.1, **R** is reduced to a value 0.135. This compares to an absolute minimum of $\mathbf{R} = 0.31$ for single–pulse interrogation. Also shown are curves that display the much lower values for **R** which are available if dead time is eliminated by the use of two independently interrogated ion collections.

In this paper we consider the particularly simple case of a repeating interrogation cycle which has a constant sensitivity to L.O. frequency for a period of time, followed by an equal "dead–time" during which the frequency of the L.O. is not observed. This case is approximated in atomic standards by "Ramsey

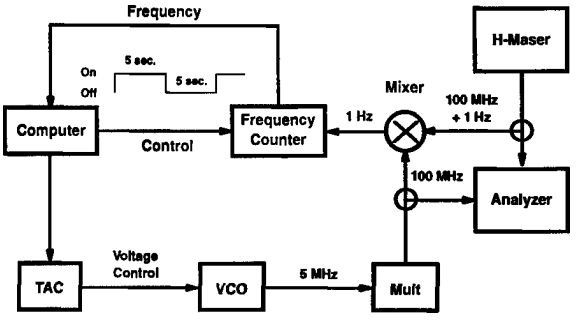


Figure 3: Schematic Diagram showing Feedback Loop and Analyzer

Interrogation" with very short interrogation pulses, and was one of those previously analyzed.

Experimental

Figure 3 shows the feedback loop for the tests described here. A high resolution electronic frequency counter replaces the trapped–ion physics package in order to eliminate the effects of discriminator–induced noise, and thus to isolate those instabilities that are due to the local oscillator and the feedback loop. The tests focused on measurements with a relative large dead time fraction of 0.5. This enables a relatively quick transition from $\sigma(\tau) \propto 1/\tau$ feedback–limited performance to the $\sigma(\tau) \propto 1/\sqrt{\tau}$ L.O.–limited performance. The arrangement has no expected $1/\sqrt{\tau}$ contribution due to the H–maser reference, and so all such effects can be attributed to the L.O. and loop effects.

In addition to experimental tests, computer simulations were performed using a commercial software package with sequences of 16384 to 65536 noise events. Flicker noise was generated using the internal ifft, random noise, and cumulative normal distribution functions with a $1/\sqrt{\omega}$ fourier amplitude multiplying factor. The resulting simulated time sequences gave excellent (nearly constant) Allan Deviation performance for τ values less than one half the total sample size. Another test of this very simple and accessible simulation technique gave fairly good values for the logarithmic correction to an expected t^2 dependence of the time error variance, extrapolated from the beginning interval.[4]

Several different digital feedback loop algorithms

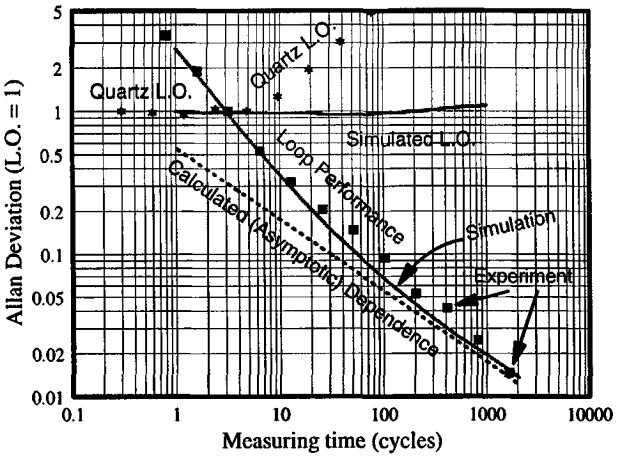


Figure 4: Calculated, simulated and measured loop performance for 50% dead time. Actual quartz oscillator stability was 1.9×10^{-13} at 10 seconds (one cycle).

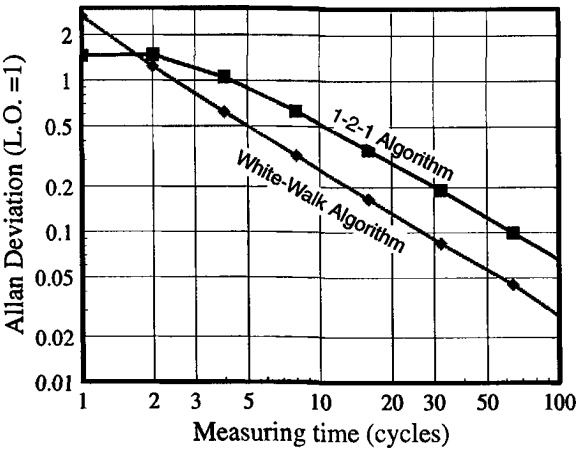


Figure 5: Simulated loop performance for several algorithms without dead time.

were tried, both in simulation and test. These included the lamp drift–rejecting 1–2–1 weighted loop algorithm often used for ion trap feedback systems[3], more conventional digital loops[5] and a new "white-walk" routine based on linear combinations of loops optimized for white and random walk noise.[6]

Results

Theoretical, simulation, and experimental data for Allan Deviation of fractional frequency are shown in Figure 4. The straight line asymptotic dependence corresponds to the value of **R** = 0.533 given by Fig. 2 for the test configuration. All show excellent agreement. Normalization of the experimental and theoretical results was made to the flicker–floor for the TAC–driven quartz oscillator of 1.9×10^{-13}. Loop parameters for both simulation and experimental results were identical, and were optimized for random-walk-of-frequency L.O. noise on account of quartz oscillator frequency wander at the longer measuring times.

The results in Fig. 1 demonstrate a loop attack time of approximately 2.7 times the cycle time. These results are substantially improved over those attainable with the more conventional 1–2–1 feedback scheme. As shown in Figure 5, simulation results without dead time effects show a long–term $1/\tau$ performance improvement by a factor of two. We find that more conventional loops also improve on the 1–2–1 loop, but only by a factor of $\approx \sqrt{2}$.[5] The simulations also show that improper loop characteristics may give rise to $1/\sqrt{\tau}$ contribution from the L.O., even without dead time, or other time variation for the loop gain.

REFERENCES

[1] G.J. Dick, "Calculation of Trapped Ion Local Oscillator Requirements," *Proc. 19th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 133–146 (1988).

[2] G.J. Dick, J.D. Prestage, C.A. Greenhall, and L. Maleki, "Local Oscillator Induced Degradation of Medium–Term Stability in Passive Atomic Frequency Standards," *Proc. 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 487–508 (1991)

[3] J.D. Prestage, G.J. Dick, and L. Maleki, "Linear Ion Trap based Atomic Frequency Standard," *Proc. 44rd Ann. Symp. on Frequency Control*, 82–88 (1990).

[4] C.A. Greenhall "Initializing a Flicker–Noise Generator," *IEEE Transactions on Instrumentation and Measurement*, **IM–35** June 1986.

[5] "Digital Control System Design" G.H. Hostetter, p. 119, Holt Rinehart and Winston, Inc., New York, 1988.

[6] "White–Walk" algorithm, G.J. Dick, to be published.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Reducing the Effect of Interrogation Phase Noise on the Frequency Stability in Passive Frequency Standards : Experimental Results and New Investigation Means

R.Barillet, D. Venot and C. Audoin

LABORATOIRE DE L'HORLOGE ATOMIQUE
Unité Propre de recherche du CNRS associée à l'Université Paris-Sud - Bât. 221 - UPS - 91405 ORSAY - Cedex - France

Abstract - It is known that the local oscillator phase noise limits the frequency stability in passive frequency standards. When the probe signal is permanently applied to the atomic resonator, this limitation is due to an intermodulation effect [1]. We have established that, with a square wave frequency modulation, a proper signal processing using a notch filter in the synthesis chain [2] and a peculiar demodulation waveform [3] [4] significantly reduces the limiting effect. Experimental tests confirm the optimization of the performance by the theoretical demodulation waveforms. The previous State of the Art is improved by a factor of 4.5. At last, in order to determine the optimum signal processing for different modulation conditions, we have realized an automatic system. It compares the frequency stabilities obtained for various signal processing conditions, it increases the number of harmonics involved in the demodulation and changes their rates. This system makes it possible to determine the minimum limiting effect on frequency stability, in passive frequency standards where atoms have a permanent interaction with the probe signal.

I. Introduction

Sophisticated atomic manipulations such as optical pumping and laser cooling make it possible to significantly improve the frequency stability of passive frequency standards. Nevertheless new limitations related to the phase noise of the probe signal have been observed : the expected improvements are presently limited by the local oscillator noise in several high performance experiments [1][5][6]. In passive frequency standards where the probe signal is permanently applied to the atomic resonator, this effect is due to the modulation-demodulation process [1] which is necessary to lock the VCXO to the atomic resonance : when the frequency of the interrogation signal is modulated at frequency f_M, an intermodulation effect translates frequency fluctuations from the even harmonics of f_M (in the Fourier domain) into a low frequency component in the error signal of the VCXO servo loop.

In the probe signal, the frequency noise existing near the second harmonics ($2f_M$) of the modulation frequency provides the largest spurious contribution, as indicated in [1]. This remark led a group at NIST to imagine and implement the first reduction of the limiting effect : it consists in rejecting the phase noise at $\pm 2f_M$ from the carrier by means of a dual notch filter [2]. This first approach is interesting because it improves the ultimate frequency stability that may be obtained, but in fact it improves the probe signal itself and does not change the relation between the frequency stability limit $\sigma_y^l(\tau)$ and the frequency noise $S_y(f)$ of the probe signal [1].

In [7], Kramer suggested that, with square modulation and demodulation waveforms, the limiting effect may disappear if the output signal of the resonator is not filtered. Though actual electronic systems cannot fulfil this condition, Kramer's remark led us to investigate the field of Signal Processing in order to reduce the consequence of the intermodulation effect by changing the initial relation between $\sigma_y^l(\tau)$ and $S_y(f)$ given in [1].

II. Theoretical predictions

When the modulation frequency is much smaller than the linewidth of the atomic resonator, the quasi-static approximation is acceptable and the frequency stability limit is analytically expressed in the usual operating conditions [1], that is when the resonator response is selectively filtered. With the quasi-static assumption and a more general signal processing, a matrix calculation shows that there is no possibility to fully suppress the spurious limiting effect whenever there is any truncation in the spectrum of the resonator signal . Consequently, the only hope to reduce the spurious effect consists in a convenient change in the signal processing.

With the quasi-static assumption and in the general case, that is when the resonator response is not selectively filtered, one obtains relations based on infinite series which describe how the frequency stability is limited, depending on modulation and demodulation waveforms [8]. *When the frequency of the probe signal is square wave modulated*, the analytical expression is relatively simple and the frequency stability limit is proportional to the square root of ξ^c given by (1). ξ^c expresses the limitation of the Allan variance by the L.O.phase noise, λ_{2n} is a power gain applied to the spectral component of the probe signal phase noise at $2nf_M$ and C_{2k+1} is the rate of the $(2k+1)f_M$ component in the demodulation signal.

0-7803-2500-1/95 $4.00 © 1995 IEEE

Relation (1) enables us to determine both a demodulation waveform and a convenient rejection factor that minimize ξ^c.

$$\xi^c = \frac{\sum_{n=1}^{+\infty} \lambda_{2n} \left(\sum_{k=0}^{+\infty} C_{2k+1} \cdot \frac{(4k+2)\,n}{(2k+1)^2 - 4n^2} \right)^2}{4 \cdot \left(\sum_{k=0}^{+\infty} \frac{C_{2k+1}}{2k+1} \right)^2} \qquad (1)$$

When the local oscillator shows a white phase noise in the Fourier frequency domain of interest, without any notch filter at $2nf_M$, the optimizing demodulation function is given by $C_1=1$, $C_3=1/27$, $C_5=1/125$... but the theoretical improvement is only 1%, which is not significant.

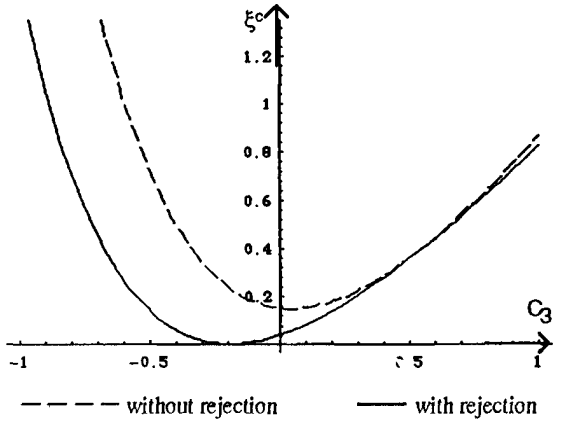


Fig.1 : Allan variance limitation versus C_3

When the phase noise component at $2f_M$ is more or less rejected in the probe signal, one finds interesting optimizing waveforms described by the rates C_{2k+1} of their spectral components. Figure 1 shows the theoretical behavior of ξ^c for no rejection and full rejection at $2f_M$ and a variable rate C_3 of the spectral component at $3f_M$ in the demodulation waveform : obviously, the optimum is obtained for $C_3 \approx -0.2$. Table 1 shows the theoretical improvement ζ of the frequency stability limit expressed in Allan standard deviation, for various optimizing demodulation waveforms. Value 1 (no optim., no rejection in table 1) corresponds to the usual case where the output of the resonator is selectively filtered and the $2f_M$ component is not rejected.

We have observed that the optimum values of the coefficients C_{2k+1} do not change significantly when the number of involved harmonics is increased : for example C_3 remains in the vicinity of -0.19, C_5 in the vicinity of -0.024, etc...

We have noticed that with a square wave frequency modulation, the improvement related to the only rejection of the $2f_M$ component is theoretically less favorable than in the sine wave frequency modulation mentioned in [2]. At last, the expected improvement is obtained with harmonics coefficients quite different from those of a square wave ($C_3=1/3$, $C_5=1/5$...) suggested by Kramer's remark in [7] ; this difference is not surprising because the phase noise spectrum has been modified and the number of involved harmonics is limited.

White phase noise (LO spectrum) Square wave freq. modulation Quasi-static approximation $C_1 = 1$ Demodulation ↓	Improvement factor ζ		
	No rejection	Full rejection at $2f_M$	30dB rejection at $2f_M$
Without any optimization	(A) 1	1.89	(B) 1.89
Optim. with C_3 (~-0.193)	1.00615	5.67	5.54
Optim. with C_3, C_5	1.00695	10.75	(C) 9.89
Optim. with C_3, C_5, C_7	1.00994	17.21	14.17
Optim. with C_3,C_5,... C_{11}	1.01018	32.36	19.73
Optim. with C_3, C_5... C_{21}	1.01037	83.65	23.82

Table I - Theoretical results

The 30dB rejection mentioned in the right column has been realized at NIST [2] by means of a dual notch filter. Consequently, the feasibility of such a circuit -which must fulfil several difficult requirements - is demonstrated and this 30dB rejection is a reasonable basis to estimate the optimum results that can actually be obtained.

Finally, in order to verify our predictions with a different approach, we have numerically simulated the behavior of the frequency standard [4], taking into account the only phase noise of the local oscillator, assumed to be a white phase noise. This simulation is a low level one. It describes step by step what happens in the servo loop of the VCXO and at last calculates the Allan standard deviation of the VCXO signal frequency. The free running behavior of the VCXO and three configurations mentioned in the previous table were considered : (A) without a rejection, (B) with a 30dB rejection at $2f_M$ and no optimization, (C) with a 30dB rejection at $2f_M$ and $\{C_3,C_5\}$ optimization. The results of the simulation are similar to those of Table 1, but always a little bit worse : the obtained improvement is 1.7 (instead of 1.89) with a full rejection at $2f_M$ and approximately 6 instead of 10 with the $\{C_3,C_5\}$ optimization and full rejection at $2f_M$. This discrepancy may be due to a simulation artefact.

III. First experimental apparatus

III.1. Demodulation

The investigation of new signal processing possibilities requires the synthesis of various demodulation waveforms and the proper synchronous demodulation of the resonator response. The waveforms are generated by a simple digital set-up followed by a DAC, whilst the synchronous demodulation is realized by an analog multiplier. A detailed description of these circuits is presented in [9] : here we just mention their general characteristics.

The block diagram of the waveform generator is shown in figure 2. A series of numbers describing the desired waveform is loaded into the memory. The analog demodulation signal is then obtained by permanently reading the memory and converting the digital data. All the frequencies involved in the experimental apparatus are synchronised by the same time base.

This digital synthesizer generates any waveform up to a fundamental frequency of 200 Hz, with 21 harmonics showing rates that can be chosen separately for in phase and quadrature components. The phase of the signal may be adjusted within 1° anywhere inside the fundamental period. The operation of this apparatus has been carefully checked. It has been specially verified that the obtained waveform and the corresponding spectrum where very precisely identical to the expected wave and harmonics coefficients.

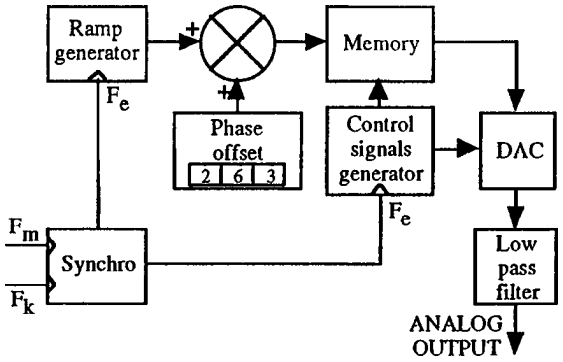


Fig.2 : block diagram of the waveform generator

For the demodulation, commercial synchronous detections are not convenient because they usually realize a square wave demodulation, whilst we need various demodulation waveforms. For that reason, we use an analog multiplier MC1496 operated in satisfactory conditions of linearity so as to properly realize the product of the two input signals. The operating scheme of this classical modulator-demodulator, usually operated in the radio-frequency domain, was modified in order to allow a low frequency operation : 20 Hz → 200 Hz for fundamental frequency and up to 4.2 kHz for harmonic frequencies. The proper noise of the demodulator in linear conditions has been characterized. In the fluctuations of the locked VCXO, it induces a flicker frequency noise which gives a frequency stability floor at 3.10^{-13} expressed in Allan standard deviation. This value is bad for a frequency standard but quite sufficient for our measurements : taking into account the usual performance of the Cs3 resonator used for the experimental verification, it only limits the sampling time to $\tau<45s$, otherwise the measurement result shows the limit related to the demodulator noise.

III.2. Overall experimental apparatus

It is similar to that presented in [10]. This apparatus includes two microwave generators DRO_{1-2} at 9192 MHz that make it possible to determine the spectrum of the probe signal and the corresponding frequency stability of the locked VCXO. An adjustable additive phase noise increases the effect to be characterized. The detail of the operation is described in [10].

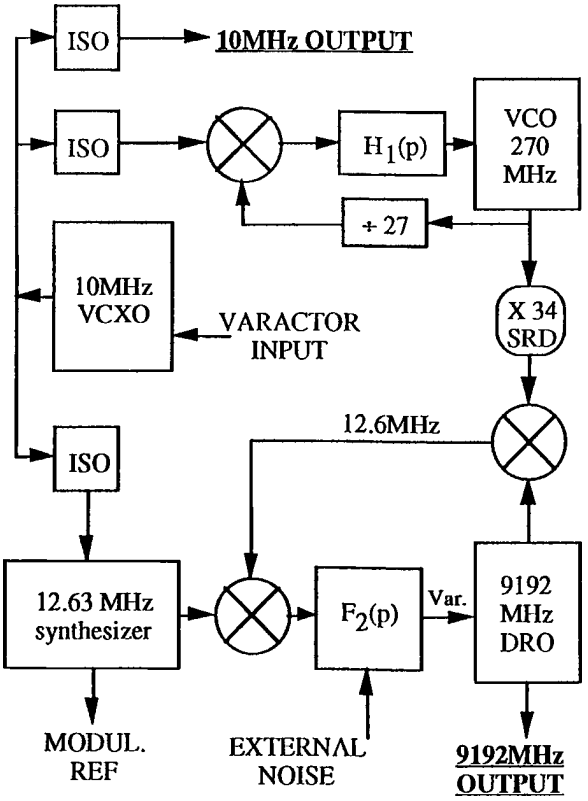


Fig. 3 : microwave generators DRO_{1-2}

Briefly, DRO_1 and DRO_2 are identical low noise microwave generators. The 9192 MHz signal delivered by a Dielectric

Resonator Oscillator (DRO) is phase-locked to a 10 MHz signal (VCXO). The VCXO signal will be frequency-locked to the atomic transition frequency, with a time constant of 1s. The block diagram of $DRO_{1\text{-}2}$ is shown in figure 3. An external noise may be added into the PLL of the DRO. It provides a controlled degradation of the phase noise in the probe signal, and the corresponding spectrum may be as large as the bandwidth of the PLL. However, because of the limited dynamics of the mixer used in this PLL, the injected signal must not exceed the peak value given by the mixer, or else the DRO unlocks and does not come back to lock. In order to easily and cheaply adjust the phase noise spectrum of the probe signal, we do not use a real dual notch filter : the effect of the notch is included in the additional noise generated by an FFT Analyzer Spectral Dynamics 380.With this system, the initial spectrum of the probe signal can be degraded by as much as 35 dB in the Fourier frequency range of interest. In our operating conditions, without additional phase noise, the intrinsic performance of the cesium resonator is in the range $2\text{-}2.5\ 10^{-12}\ \tau^{-1/2}$, depending on ambient conditions.

IV. Experimental results

The following experimental results have been obtained with a square wave frequency modulation at $f_M = 35$ Hz in the probe signal. Various additive noises have been injected into the PLL of the DRO. Taking into account the quasi-static approximation, all the synthesized demodulation waveforms D(t) that we have tested up to now contained only in phase components, that is were described by

$$D(t) = \sin 2\pi f_M t + C_3 \sin 2\pi(3f_M)t + \ldots + C_{2k+1} \sin 2\pi(2k+1)f_M t + \ldots \quad (2)$$

IV.1. Preliminary measurements

We have observed that the transient signals occuring during the frequency switchings were larger than the amplitude required for a linear behavior of the MC1496 circuit. With the low modulation frequency that is used (35 Hz), the corresponding disturbance is a small fraction of the modulation period. Nevertheless we have verified that these transient signals do not significantly disturb the linearity of the analog demodulator.

For that, we measured the frequency stability of the locked VCXO for $\tau = 9.17$s in the usual conditions -with a selective filter at the output of the resonator- by means of a Brookdeal 9503 synchronous detection. Then we measured the frequency stability without a selective filter at the output of the resonator but using the analog demodulator driven by a sine wave demodulation function, which is theoretically equivalent to the Brookdeal measurement. These measurements were made for two conditions of additive noise [9] and no significant difference was observed.

IV.2. Demodulation by various waveforms

We have measured the frequency stability of the locked VCXO in presence of the noise profiles shown in figures 4 and 5. With the $2f_M$ rejection, the phase noise of the probe signal at 9192 MHz is flat in the range 140 Hz - 350 Hz, where its power spectral density is -50 dBc/Hz and the rejection at $2f_M$ is approximately 24 dB. For the measurement reported in the last line of table 2 (**), the phase noise at 9192 MHz is -50 dBc/Hz from f = 70 Hz to 330 Hz (no rejection at $2f_M$ = 70 Hz).

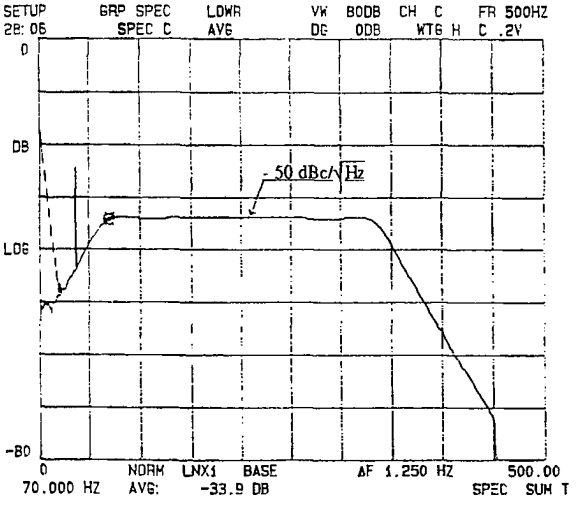


Fig.4 : phase noise profile N°1 (no rejection)

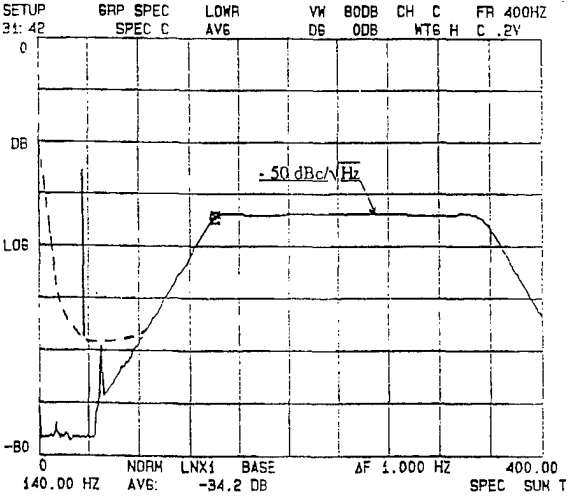


Fig.5 : phase noise profile N°2 ($2f_M$ rejection)

We have changed the spectrum of the demodulation waveform specially by means of the third harmonics rate. Three interesting results are observed in table 2 :

- when the rejection at $2f_M$ is used, the corresponding improvement is only a factor 1.4, instead of the theoretical value 1.87 or the simulation result 1.7. It confirms that with a square wave frequency modulation, it is not possible to significantly reduce the intermodulation effect by the only use of a dual notch filter.
- the 3rd harmonics rate $C_3 = -0.2$ induces a significant improvement probably limited by the natural performance of the Cs3 resonator ($\approx 0.8\ 10^{-12}$ for $\tau = 9.17$s). It is interesting to notice that, in spite of a band limited additive noise spectrum, this first optimization is obtained for a value of C_3 very similar to that of the theoretical results (-0.193) with a quasi infinite white phase noise with rejection at $2f_M$. It is probable that $C_3 = -0.2$ is not exactly the optimum value of C_3 but the corresponding frequency stability will not be very different with the real C_{3opt} because of the flatness of the curve near the optimum, as shown in figure 1.
- the use of a pseudo-square demodulation waveform ($C_{2n+1} = \frac{1}{2n+1}$ up to n=4) is not favorable at all. The same observation has been made with the Brookdeal synchronous detection when the output of the atomic resonator is not filtered, leading to a frequency stability $\sigma_y(\tau=9.17s) = 2.74\ 10^{-12}$.

- Square wave frequency modulation -

Demodulation	$\sigma_y(\tau = 9.17s)$	Number of samples
$C_3 = 0$	$1.70\ 10^{-12}$	90
$C_3 = -0.2$	$1.01\ 10^{-12}$	90
$C_3 = -0.4$	$1.70\ 10^{-12}$	90
$C_3 = +0.4$	$3.74\ 10^{-12}$	90
Pseudo square $C_{2n+1} = \frac{1}{2n+1}$ up to n=4	$5.61\ 10^{-12}$	90
$C_3=0$, no rejection at $2f_M$ **	$2.38\ 10^{-12}$	90

Table 2 - Effect of various demodulation waveforms

IV.3. Effect of the theoretical optimizing demodulation

In order to verify more precisely the theoretical predictions in our operating conditions, we have taken into account the limited spectrum of the additive noise and calculated the corresponding optimized demodulations with (C_3,C_5), then (C_3,C_5,C_7). These values are indicated in table 3. Different C_3 values have also been checked. We have increased the power of the additive noise up to the highest possible level in order to strongly degrade the frequency noise of the probe signal. In those conditions, the phase noise at 9.2 GHz is -46 dBc/Hz from f = 140 Hz to 350 Hz. The rejection at $2f_M$ (70 Hz) is approximately 27 dB. The two shapes of the noise spectrum at 9.2 GHz are similar to those of figures 4 and 5 (dotted line).

- Square wave frequency modulation -

Demodulation	$\sigma_y(\tau = 9.17s)$	Number of samples
$C_3 = 0$	$2.76\ 10^{-12}$	90
$C_3 = -0.1$	$1.75\ 10^{-12}$	90
$C_3 = -0.2$	$1.31\ 10^{-12}$	90
$C_3 = -0.3$	$1.50\ 10^{-12}$	90
$C_3 = -0.18745$ and $C_5 = -0.02310$	$1.07\ 10^{-12}$	270
$C_3 = -0.19037$ and $C_5 = -0.02460$ and $C_7 = -0.00661$	$1.02\ 10^{-12}$	180

Table 3 - Effect of various C_3 and optimizing demodulation waveforms

Experimental results are presented in table 3. It is obvious that the performance of the system is improved when a larger number of optimized harmonics is used. In the last configuration, the frequency stability without an additive noise is 8.10^{-13}. Hence, the corresponding contribution of the additive noise may be estimated about 6.10^{-13}. In this case, the spurious effect related to the phase noise of the local oscillator is reduced by a factor of $2.76/0.6 \approx 4.5$ compared to the case of the only notch filter.

The expected theoretical improvement given by table 1 is $14.17/1.89 \approx 7.5$. It is possible that the quasi-static assumption limits the precision of our determination (waveform and expected improvement). In that case, it means that the real optimum set (C_3, C_5,C_7) may be different from the one mentioned in the last line of Table 3, and the improvement a little better.

In any case, the vicinity of the noise floor of the experiment does not presently enable us to have a precise estimate of the obtained improvement, but these measurements show that it is possible to significantly improve the NIST method [2] by means of a proper signal processing of the resonator output signal, according to the theory developped in [3], [4] and [8].

V. Automatic optimization of the demodulation waveform

The previous theoretical and experimental results concern a square wave frequency modulation. A satisfactory description of the limiting effect and of its improvement is obtained by means of the quasi-static approximation. But in practice the modulation waveform is not necessary square and the modulation frequency is not much smaller than the linewidth of the atomic resonator response. Taking into account the dynamical behavior of the atomic resonator, we need phase and quadrature components for the various harmonics involved in the demodulation waveform.

Furthermore, for a considered phase noise spectrum of the probe signal, a different optimizing demodulation waveform is associated to each modulation waveform. Consequently, for a considered phase noise spectrum of the probe signal, the absolute optimum will associate modulation and demodulation conditions (waveform, frequency, depth).

In order to experimentally determine this absolute optimum, we must first sweep the space of demodulation coefficients C_{2k+1}, with two components for each C_{2k+1}- for various modulation waveforms. In the first experimental apparatus, presented above, we have only verified that the predicted waveforms induce a significant improvement whose order of magnitude is satisfactory. A more precise theoretical determination of the optimum is extremely heavy, while an experimental approach is now easy with a slightly improved system. In the previous experimental apparatus, loading a set of EPROM memories to generate a new demodulation waveform required to plug them in the Shooter, run the program which calculates the new waveform, load the memories and finally plug them in the demodulation waveform generator. All that was very long, needed a permanent human operator and did not allow a continuous operation.

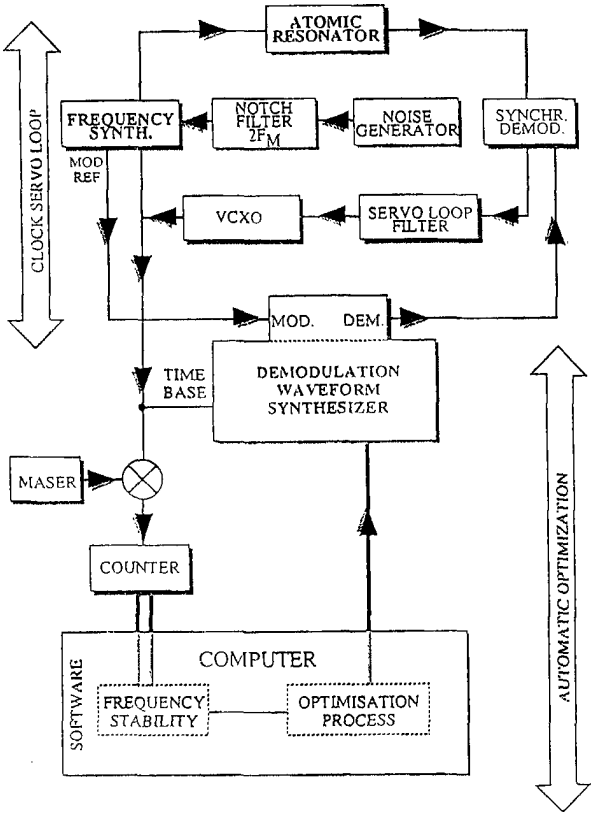


Fig.6 : Experimental Bloc Diagram

Consequently, we have realized an automatic system that makes it possible to continuously generate various demodulation waveforms and to test their efficiency in the experiment. The corresponding block diagram of the overall apparatus and a view of the automatic system are shown in figures 6 and 7 respectively.

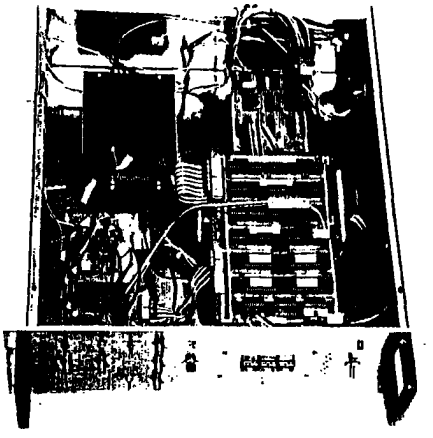


Fig.7 : view of the automatic system

The automatic system is operated with two random access memories. One of the memories is used to synthesize a demodulation waveform (under test) which has been previously loaded, whilst the second one is being loaded by a new waveform. The cycle time is 20mn or more because of our optimization criterion $\sigma_y(\tau = 10s)$. At the end of the cycle, the estimated frequency stability of the disturbed cesium frequency standard is calculated by the computer. When the frequency stability cannot be improved any longer, the optimizing waveform has been determined.

It is the optimizing demodulation waveform related to the considered phase noise spectrum -including a possible rejection- and the considered modulation conditions.

In this optimization process, the system automatically compensates the possible non linearities that may exist, for example in the analog signal processing, such as the optical detector in the atomic resonator or the synchronous demodulator : in fact, this new system will determine the minimum frequency stability limit that may be obtained for a considered phase noise spectrum of the probe signal and a considered frequency modulation.

By scanning a very large range of frequency modulation conditions, it will then be possible to determine the absolute minimum limit of $\sigma_y(\tau)$ related to a peculiar phase noise spectrum of the probe signal. The corresponding optimum demodulation will include the compensation of the possible non linearities in the electronic circuits, and hence it will be significant if these non linearities are lower than -70 dB.

VI. Conclusion

Experimental results show the efficiency of a proper demodulation coupled with the use of a dual notch filter in the case of a square wave frequency modulation. Hence, it is now possible to obtain the intrinsic frequency stability of the atomic resonator even with a relatively noisy local oscillator, in the case of a square wave frequency modulation. A more interesting improvement is expected with frequency modulations using different waveforms, especially the sine waveform. It will be tested by the new automatic system which has been realized and makes it possible to experimentally determine the absolute optimizing demodulation waveform for any phase noise spectrum of the probe signal. More generally, it is interesting to notice that this apparatus may be used in order to optimize systems where non linearities or peculiar processings shift the noise spectral components.

VII. Acknowledgements

We are very grateful to DRET (Direction des Recherches, Etudes et Techniques) who has supported the main part of this work and made it possible to obtain the previous results.

We also are grateful to Omega Technologies deeply involved in our studies.

We are happy to acknowledge Bertrand Boussert for his excellent adjustment of the Cs3 resonator, Benoit Cardoso and Alain Meunier whose technical contributions have been very important in the experimental apparatus.

References

[1] C.Audoin, V. Candelier, and N. Dimarcq "A Limit to the Frequency Stability of Passive Frequency Standards due to an Intermodulation Effect". IEEE Trans. on Instr. and Measur. vol 40, n°2, april 1991, pp 121-125.

[2] C. Szekely, F.L. Walls, L.P. Lowe, R.E. Drullinger and A. Novick. "Reducing Local Oscillator Phase Noise Limitations on the Frequency Stability of Passive Frequency Standards : Tests of a new concept"- IEEE Trans. UFFC, Vol 41 N°4, july 1994, pp 518-521.

[3] D. Venot, R. Barillet and C. Audoin. "On a Possibility of Reducing the Frequency Stability Limitation by the Interrogation Noise in Passive Frequency Standards" Proc. of the 1994 EFTF, Munich, march 94, pp 656-670

[4] D. Venot, R. Barillet and C. Audoin. "A possibility of reducing the effect of interrogation phase noise on the frequency stability in passive frequency standards", presented at CPEM 94, june 1994, Boulder, USA

[5] G. Santarelli, P. Laurent, S.N.Lea, A. Nadir and A. Clairon."Preliminary results in a cesium atomic fountain frequency standard". Proc. of the 1994 EFTF, Munich, march 94, pp 46- 57

[6] G.J.Dick, J.D.Prestage, C.A. Greenhall and L.Maleki. "Local Oscillator Induced Degradation of Medium-term Stability in Passive Atomic Frequency Standards".Proc.22nd Ann.PTTI Applications and Planning Meeting, pp 487-508 (1990)

[7] G. Kramer "Noise in Passive Frequency Standards." CPEM 1974 Digest, IEEE Conference Publication n°113, IEEE Catalog n° CH 0770-8IM (unpublished).

[8] D.Venot. "Démodulation synchrone et effet d'intermodulation"- Rapport interne - LHA, février 1994

[9] R. Barillet, D.Venot and C. Audoin."Reducing the effect of interrogation phase noise on the frequency stability in passive frequency standards : experimental results" to be published in the Proc. of the 1995 EFTF, Besançon, 8-10 march 1995

[10] R. Barillet, V. Giordano, J. Viennet and C. Audoin. "Limitation of the clock frequency stability by the interrogation frequency noise : experimental results". IEEE Trans. on Instr. and Meas., vol 42, N°2, april 1993, pp 276-280.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DIODE LASERS FOR FREQUENCY STANDARDS AND PRECISION SPECTROSCOPY*

L. Hollberg, J. Åman[a], S. Waltman, J.H. Marquardt, M. Stephens,
R.W. Fox, D.A. Van Baak[b], C.S. Weimer, H.G. Robinson,
A.S. Zibrov[c], N. Mackie, T.P. Zibrova[c], and L. Pendrill[d]

National Institute of Standards and Technology, Boulder, CO 80303
[a] Chalmers University of Technology, Göteborg, Sweden
[b] Department of Physics, Calvin College, Grand Rapids, MI 49546
[c] Lebedev Institute of Physics, Moscow, Russia
[d] Swedish Testing and Research Institute, Borås, Sweden

Abstract

As they apply to frequency standards and precision spectroscopy the characteristics and technology of tunable diode lasers are briefly reviewed. It is now possible to use nonlinear optical techniques and high quality diode lasers to extend the useful wavelength coverage of semiconductor lasers into the UV, the IR and even millimeter-wave spectral regions. Progress in developing an all diode-laser system for cooling, trapping and precision spectroscopy of calcium is discussed. New measurements indicated that two-stage optical cooling of calcium may be feasible; this should improve the accuracy of the 657 nm optical wavelength/frequency reference.

Introduction

Diode lasers based on the III-V semiconductor materials are very attractive sources of coherent radiation; they are simple to use, compact, highly efficient (~30 % electrical to optical power conversion), reliable and low cost. In addition they have other beneficial attributes such as room-temperature operation, tunability and high-speed modulation capabilities. Single spatial-mode lasers are available in the red and near infrared regions with power levels that range from 10 to 200 mW . At a few special wavelengths high-power tapered-amplifiers and MOPAs (master oscillator power amplifiers) are now available with power levels approaching 1 watt. Most precision spectroscopy does not require high-power lasers, but the high-power devices greatly simplify the nonlinear mixing that is required to reach other spectral regions. Some of the characteristics of commercially available diode lasers are given in table 1.

There are many compelling reasons to develop diode laser systems for applications in precision spectroscopy and frequency standards. For long-term operation of standards and transportable (non-laboratory) instruments we need practical, robust, efficient, tunable, optical sources with high spectral-purity. For applications that require a tunable source, diode lasers may be the only practical solution. The influence of diode lasers on frequency and length standards is apparent in the research programs at all of the major national standards laboratories. For example, efforts are now underway to develop optical and microwave frequency references using diode lasers with a number of neutral atoms and molecules including: Cs, Rb, Rb 2-photon, Sr, Ba, Ca, I_2, C_2H_2, HCN, and CH_4. Significant work is also progressing to develop diode laser systems for trapped-ion frequency standards based on Hg^+, Ba^+, Sr^+,Ca^+, Yb^+, etc.

Now that there are a number of improved frequency/wavelength references in the visible and infrared, there is a real need to expand the capabilities and simplify the methods that are used to synthesize and measure optical frequencies. Impressive results of past optical frequency measurements were achieved using large, complex, multi-laser frequency multiplication chains. With diode lasers and nonlinear optical techniques these systems can be simplified.

Unfortunately, diode lasers also have well known problems such as: discontinuous and limited tuning range, limited power at most wavelengths, and lack of availability at many wavelengths. Some of these problems can be solved with electronic and optical control systems. The parameters in Table 1 are for solitary single-mode lasers without external control.

0-7803-2500-1/95 $4.00 © 1995 IEEE

TABLE OF SEMICONDUCTOR LASER PARAMETERS

Semiconductor	$(Al_xGa_{1-x})_yIn_{1-y}P$	$Al_xGa_{1-x}As$	$Ga_xIn_{1-x}P_yAs_{1-y}$	MOPA
Wavelength	635-670 nm	750-850 nm	1.3 - 1.5 μm	850, 980 nm
Output power (mW)	3-30	5-200	3-100	500 - 1000
farfield divergence FWHM (degrees)	8 × 40	11 × 33	30 × 35	0.3 x 35
Waveguide mode dimensions, μm	4 x 1	3 × 1	1.25 × 1	1 x 100
Astigmatism	~10 μm	1 to 5 μm	1 to 5 μm	~500 μm
Threshold current	30 - 90 mA	20-60 mA	20 - 50 mA	~0.5 A
Operating Current	50 - 120 mA	50 - 200 mA	40 - 120 mA	~2.5 A
$I_2 = I_1 e^{(T2-T1)/To}$	$T_o \approx 100$ K	$T_o \approx 150$ K	$T_o \approx 60$ K	
slope efficiency, (mW/mA)	0.5 - 0.7	0.7	0.2	1
Refractive index	3.1 - 3.5	3.3 - 3.6	3.2 - 3.5	3.3 -3.6
Freq. vs. Inject. current	~5 GHz/mA	~3 GHz/mA	~ 1 GHz/mA	
Freq. vs. temp. **large scale** **small scale**	 ~0.2 nm/K ~30 GHz/K	 ~0.25 nm/K ~30 GHz/K	 ~0.3 nm/K ~ 10 GHz/K	
gain bandwidth	20 nm	30 nm	50 nm	20 nm
typical linewidth	200 MHz	5 - 20 MHz	100 MHz	
alpha factor -- α		3 to 6	4 to 8	

Table 1. Typical Characteristics of semiconductor diode lasers. (Table is from Fox et al[1]).

The causes of frequency fluctuations in diode lasers include: current and temperature instabilities, optical feedback, spontaneous emission, and aging effects. For spectroscopic applications the most challenging parameters are the change in wavelength with temperature and injection current. The fast-linewidth of a solitary laser is typically about 20 MHz, and in order to optimally use this potential resolution we need to have a precision in tuning that is slightly better than the linewidth, say 10 MHz. This means we require a temperature stability of about 300 microkelvin, and current stability of about 3 microamps (out of a DC current of about 100 mA, or equivalently $\Delta I/I=3x10\text{-}5$). These results are certainly possible, but they are also non-trivial. Obviously, for high resolution applications care must be taken with the precision and stability of the control systems that are used.

Diode laser spectral characteristics

If for the moment, we ignore the fact that these diodes are lasers, and consider them just as tunable oscillators their characteristics don't seem so unusual. Without external optical or electronic feedback the linewidth of a typical diode laser is about 20 MHz, the oscillation frequency is about 3 $x10^{14}$ Hz, and the tuning range is about 10^{13} Hz. In relative terms, this is an oscillator with a center frequency in the visible, with a fractional spectral width of about 10^{-7}, and a tuning range of about 5% of the center frequency. The absolute numbers are large, but the relative numbers are not so different from other tunable semiconductor oscillators.

For applications in the area of frequency and length standards the diode laser's spectral characteristics are of paramount importance. A significant amount of research and development effort has already gone into understanding and controlling these properties. Good reviews of the spectral characteristics of diode lasers as they apply to scientific applications can be found in a number of recent publications.[1,2,3] The details of the spectral characteristics and the dynamics of diode lasers are actually quite complex: there is the fundamental role of spontaneous emission, the cross coupling of amplitude and phase fluctuations, as-well-as a very high sensitivity of the lasers to optical feedback. Typically, for single-mode lasers the amplitude noise is relatively low and under the right conditions can be near the shot-noise limit. On the other hand, when compared with other types of tunable lasers, the frequency noise on diode lasers is relatively large and spectrally broadband.

It is now common practice to use optical and/or electronic feedback to narrow the spectral width of diode lasers. When using optical feedback to control the spectral characteristics there are essentially two stable operating regimes of feedback power. A diode laser with optical feedback can be stable with a level of feedback that is low, or high, compared with the laser's internal facet reflection. We typically use extended-cavity grating-tuned lasers (ECDL) in the strong feedback regime because this configuration usually provides reliable tuning, single-mode operation and narrow linewidths (fast-linewidth ~50 kHz).[4] In this case the wavelength is controlled by the grating, and the tuning with temperature and injection current are typically reduced by a factor of about 100. Unfortunately, the extended cavity lasers are larger, more complicated and more susceptible to mechanical disturbances. An ECDL built in an enclosed box in a laboratory environment will show increased frequency noise at Fourier frequencies less than about 100 kHz. This low frequency noise usually has strong components at the various mechanical and acoustical resonances of the ECDL system.

Extending Wavelength Coverage

Semiconductor lasers can be used in conjunction with nonlinear optical mixing techniques to generate coherent, tunable radiation throughout the electromagnetic spectrum, from the ultraviolet to the microwave region. That is not to imply that all the interesting wavelengths are accessible in a practical system; rather, it now seems feasible (at least in principle) to reach almost any wavelength between 210 nm and 18 μm using diode laser sources. For example, second-harmonic-generation and sum-frequency-mixing can be used with commercial diode lasers to cover the spectral region between 500 nm and 315 nm. Cascading two of these systems in series can extend the coverage into the ultraviolet, as has now been demonstrated in a couple of recent experiments.[5,6] Likewise, difference frequency generation (DFG) can be used to cover the infrared region. In principle, visible and near-visible diode lasers and DFG in $AgGaS_2$ and $AgGaSe_2$ can cover the spectral region from 2 to 18 μm. We are presently using this technology in the 3 μm region. These systems are now being demonstrated for use in a variety of applications.[7] Even lower frequencies are now accessible using special, low-temperature-grown GaAs photomixers (developed by E. Brown et al.[8]) to generate tunable mm-wave radiation as the difference-frequency between two laser sources. Tunable radiation in the frequency range between 40 GHz and 3.7 THz has now been demonstrated. Recent results in our lab and others have shown that these photomixers also work with diode laser sources.

Calcium wavelength/frequency reference

At NIST we have been developing diode laser systems for high resolution spectroscopy and laser cooling of calcium atoms.[9] Similar efforts are underway at other laboratories.[10,11] Ca is of interest as a wavelength/frequency reference, and as a realization of the meter, because of its narrow (400 Hz natural lifetime limited linewidth) 657 nm optical transition in the red.[12] As can be seen from the energy level diagram (Fig. 1) calcium is also a very good match for use with commercial semiconductor lasers. Because of the availability of good quality of diode lasers in the red a high performance, transportable, transfer-standard of wavelength/frequency could be constructed using traditional atomic beam or cell methods.

For the highest accuracy cold atoms are required. As demonstrated by others[13,14,15] calcium can be cooled and trapped using the strong 423 nm line. We have now developed a frequency-doubled, tunable diode laser for this purpose. A grating-tuned extended-cavity diode laser injection locks a 150 mW single-mode laser which is then frequency-doubled using $KNbO_3$ in a resonant ring cavity. This method produces as much as 45 mW of useful output power at 423 nm which has been used for 1-dimensional, transverse cooling of a calcium atomic beam. We have observed 10 kHz wide Ramsey fringes on the 657 nm transition with this system. To improve the resolution, and more importantly, the accuracy of this 657 nm wavelength reference, 3-dimensional cooling is now required.

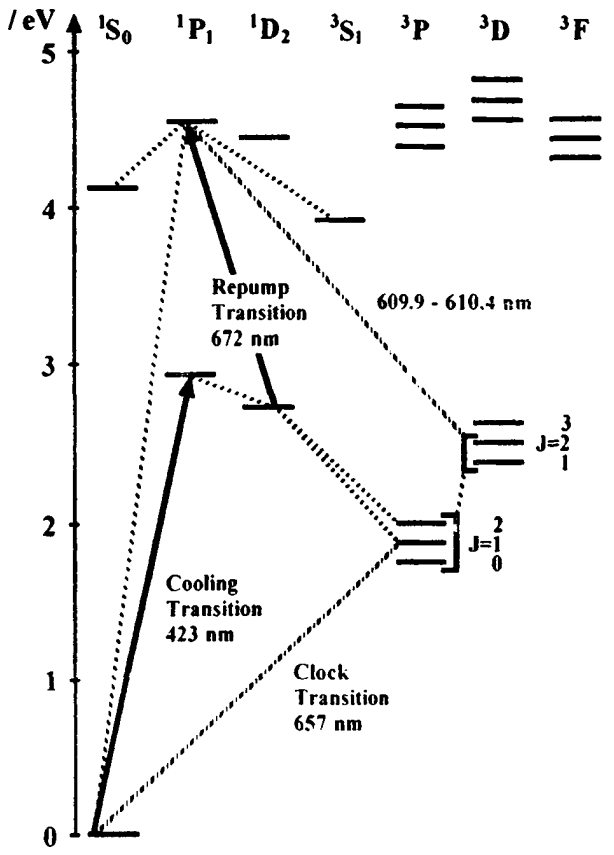


Figure. 1 Calcium energy level diagram showing the narrow 657 nm reference line, the 423 nm cooling/trapping transition and the proposed repumping transition at 672 nm.

It is known that it may be possible to enhance the storage time of the Ca magneto-optic trap by using a laser (at 672 nm) to repump atoms out of the $^1D_2(4s3d)$ state before they are lost out of the laser cooling cycle in the trap.[15] Questions and unresolved discrepancies still remain about the efficiency of this method, the trap lifetime and atomic branching ratios. These issues need to be resolved to determine whether this repumping scheme (or other schemes) could be effective. In particular, the transition-rates out of, and lifetime of the $^1P_1(4s5p)$ state (upper state of the 672 nm transition) are not in good agreement. We have used our transverse-cooled Ca beam to begin to resolve the discrepancies in previous experimental results, and to explore the possibility of using the 672 nm light for repumping. The 609 nm transition is an intercombination line which is a potential loss channel from the trapping and cooling cycle. We have now measured the ratio of the 672 nm decay rate to the 609 nm decay rate. Preliminary results indicate that the ratio of decay rates ($\Gamma_{672}/\Gamma_{609}$) is about 100. This means that the 609 nm transition is not a dominant loss of population out of the $^1P_1(4s5p)$ state, and thus does not preclude the use of the 672 nm line for repumping. However, this measurement does not resolve the discrepancy that remains with the lifetime of the $^1P_1(4s5p)$ state. Since it is possible to cycle numerous photons on the repumping transition it may even be possible to use this transition as a 2nd cooling stage. A high-power tunable 672 nm diode laser is now being developed to explore these possibilities.

* Contribution of NIST not subject to copyright.

References

1. R.W. Fox, A.S. Zibrov, and L. Hollberg, Semiconductor Diode Lasers, in Methods in Experimental Physics, vol. Atomic, Molecular and Optical Physics, to be published Academic Press (1995).

2. M. Ohtsu, *Highly Coherent Semiconductor Lasers*, Artech House, Norwood MA, 1992.

3. H. Telle, Stabilization and Modulation Schemes of Laser Diodes for Applied Spectroscopy, *Spectrochimica Acta Rev.*, vol. 15, no. 5, 301-327, 1993.

4. C. Weiman and L. Hollberg, Using Diode Lasers for Atomic Physics, Rev. Sci. Instrum., 62, 1-20, 1991.

5. C. Zimmerman, V. Vuletic, A. Hemmerich, and T.W. Hänsch, All solid state laser source for tunable blue and ultraviolet radiation, Appl. Phys. Lett. 66, 2318-2320 (1995).

6. Lew Goldberg et al. CLEO/QELS '95, proceed. postdeadline papers, Opt. Soc. Am. Boston, (1995).

7. U. Simon, S. Waltman, I. Loa, F.K. Tittel, and L. Hollberg, External Cavity Difference-Frequency Source in the Mid-Infrared based on $AgGaS_2$ and Diode Lasers, J. Opt. Soc. Am. B, 12, 323-327 (1994).

8. E.R. Brown, F.W. Smith and K.A. McIntosh, Coherent millimeter-wave generation by heterodyne conversion in low-temperature-grown GaAs photoconductors, J. Appl. Phys. 73, 1480 (1993).

9. R. Ellingsen, A.S. Zibrov, R.W. Fox, V.L. Velichansky, C. Weimer, G.M. Tino, and L. Hollberg, High-Resolution Diode Laser Spectroscopy of Calcium, Appl. Phys. B, 59, 327-331 (1994).

10. A. Celikov, F. Riehle, V.L. Velichansky and J. Helmcke, High-resolution spectroscopy in a Ca atomic beam, Opt. Comm., 107, 54-60 (1994).

11. T. Kurosu, J. Ishikawa, N. Ito, and R.W. Fox, High-resolution diode laser spectrometer for the Ca $^1S_0$-$^3P_1$ Transition, proceed. *Laser Frequency Stabilization and Noise Reduction,*SPIE, vol. 2378, 236-244 (1995).

12. R.L. Barger, J.C. Bergquist, T.C. English, and D.J. Glaze, Resolution of the photon recoil structure of the 6573-Å calcium line in an atomic beam with optical Ramsey fringes, Appl. Phys. Lett., 34, 850-852 (1979).

13. N. Beverini, F. Giammanco, E. Maccioni, F. Strumia, and G. Vissani, Measurement of the calcium $^1P_1$-$^1D_2$ transition rate in a laser-cooled atomic beam, J. Opt. Soc. Am. B6, 2188-2193 (1989).

14. Th. Kisters, K. Zeiske, F. Riehle and J. Helmcke, High-resolution spectroscopy with laser-cooled and trapped calcium atoms, Appl. Phys. B, 59, 89-98 (1994).

15. T. Kurosu and F. Shimizu, Laser cooling and trapping of Alkaline earth atoms, Jpn. J. Appl. Phys., 31, 908-912 (1992).

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

W-Band MMIC-Based Component Development and Applications

Huei Wang, Kwo-Wei Chang, Dennis Chung-Wen Lo, and Gee Sam Dow

(Invited Paper)

TRW
Electronic Systems and Technology Division
M5/1041, One Space Park, Redondo Beach, CA 90278

ABSTRACT

This paper summarizes the W-band (75-110 GHz) monolithic components developed at TRW in the past five years. Current MMIC development status is described. The system applications for these W-band MMIC-based components are presented and the trend for future W-band MMIC components, including new material, fabrication, design, and testing technologies are addressed.

I. INTRODUCTION

W-band (75-110 GHz) is an important frequency band in millimeter-wave (MMW) applications due to the fact that in atmosphere the absorption peak of MMW propagation caused by oxygen molecules are around 60 and 120 GHz, and there is an atmospheric window with reasonably low attenuation (0.1 dB/km) centered at 94 GHz with a bandwidth of about 20% [1],[61]. Typical W-band system applications include radar, communication, radiometry, instrumentation, etc. The W-band frequency offers wider RF bandwidth, less congested spectrum, smaller circuit and antenna size when compared with microwave and lower MMW frequencies. In addition, W-band frequency has advantages over infrared and optical frequencies in bad weather conditions, such as in clouds, smoke, fog, and haze.

Monolithic microwave/MMW integrated circuits (MMICs) combine various active and passive circuit functions on a single substrate. In contrast to the conventional hybrid microwave/MMW integrated circuits (MICs), MMICs provide advantages of small size, low weight, high repeatability, and low cost at high volume [2]. Recently, MMICs have demonstrated excellent performance up to W-band. While such MMICs have been applied primarily foward a variety of military systems, W-band MMIC development enables both the commercial and civil applications as well. Table 1 lists the proposed frequency spectrum allocations for some commercial and civil systems [60]. In the past five years, numerous W-band MMIC components including low noise amplifiers, power amplifiers, mixers, frequency multipliers, frequency sources, switches, detectors, phase shifters, and multi-functional chips, have been developed at TRW. These components use high electron mobility transistor (HEMT) and heterojunction bipolar transistor (HBT) technologies fabricated on GaAs or InP substrates. The critical requirements of the W-band MMIC development are (i) adequate device performance at W-band, (ii) a stable MMIC process, which includes the key technologies of material growth via molecular beam epitaxy (MBE) and electronic beam (E-beam) lithography for gate definition of HEMT devices, and (iii) MMIC design/analysis methodology at MMW frequency.

Applications	Freq. Band Allocation (GHz)	Specific Utilization	Comment
Satellite	71-75.5	Earth-to-Space	
	75.5-76	Amateur Satellite	
	81-84	Space-to-Earth	
	84-86	Broadcasting Satellite	
	92-95	Earth-to-Space	
	95-100	Radio-navigation Satellite	
	102-105	Space-to-Earth	
Radio Astronomy and Passive Sensing	86-92		
	100-102		
	105-126		
Vehicular Radar	76-77	Max. ave. transmitter power at antenna = 40mW Max. inband & out-of-band spur. peak power = 0.1mW	
	92-95	Max. ave. transmitter power at antenna = 50mW Max. in-bound & out-of-band spur. peak power = 0.1mW	1-GHz band between 92 & 95 except 93.07 to 93.27 GHz

Table 1. Proposed frequency spectrum allocations of W-band commercial and civil applications.

This paper summarizes the development status of TRW W-band MMIC-based components and their applications. The MMIC-based W-band systems currently under investigation include passive imaging radiometric systems and FMCW radar systems. The structure of this paper is organized as follows: Section II briefly introduces the MMIC device technologies for operation at W-band. Section III presents the developed W-band MMIC components. The system applications are discussed in Section IV. Section V addresses the future trend of MMW monolithic IC development and is followed by a brief summary.

II. MMIC DEVICE TECHNOLOGIES FOR W-BAND APPLICATIONS

HEMTs and HBTs are the devices of a new generation of III-V compound transistors which rely on the concept of

This paper was presented in the 1995 IEEE International Frequency Control Symposium, San Francisco, CA, May 31- June 2, 1995.
This work was supported in part by TRW IR&D Projects and MIMIC Phase 2 Program (Contract No. DAAL01-91-C-0156) from ARPA and Army Research Laboratory.

0-7803-2500-1/95 $4.00 © 1995 IEEE

the heterojunction for their operation [3]. The HEMT has the advantage of low noise and high gain performance at W-band for amplification function due to its high electron mobility and high peak electron drift velocity. On the other hand, the HBT possesses low $1/f$ noise for low phase noise source applications, and high linearity for amplification function. HEMT technology was used to develop most of the W-band MMIC components, while HBT technology was mainly used for low frequency low phase noise oscillators in the W-band source application due to the lower cut off frequency of the HBT device.

A. GaAs-based HEMT

The GaAs-based HEMT technology started with lattice-matched AlGaAs/GaAs material system due to the the ease of material epitaxial growth, and demonstrated device performance up to V-band (50-75 GHz) [62]. Later, the pseudomorphic (PM) AlGaAs/InGaAs/GaAs system was found to offer superior performance with higher unit current frequency (f_T), unit power gain frequency f_{max}, and lower minimum noise figure (F_{min}). In W-band applications, submicron gate lengths shorter than 0.2 μm in PM AlGaAs/InGaAs/GaAs HEMTs are typically required to provide adequate performance.

The 0.1-μm low noise PM AlGaAs/InGaAs/GaAs HEMT [4] has been used to develop our W-band MMIC components at TRW since 1990. The low noise HEMT was optimized for high gain operation at W-band. The W-Band HEMT structure was grown using molecular beam epitaxy (MBE) on three-inch substrates and uses a PM $In_{0.22}Ga_{0.78}As$ channel. The InGaAs PM HEMT used a planar doping layer to achieve high channel aspect ratio as well as higher electron transfer efficiency for higher transconductance. E-beam lithography was used for the 0.1-μm T-shaped gate definition. 0.1-μm T-gate InGaAs PM HEMTs fabricated using this process typically had a G_m of 670 mS/mm with f_T of 130 GHz. Compared to the device with a conventional trapezoidal gate structure, the noise performance of the T-gate PM HEMT was improved due to the lower gate resistance. In 1991, the 0.1-μm PM power AlGaAs/InGaAs HEMT [5] was incorporated in our MMIC technology. The power HEMT differed from the low noise HEMT by the introduction of an additional planar doping in the channel region to increase the device current handling capability. This also improved the transconductance linearity over a wider range of gate voltages. A high Schottky gate-drain breakdown was achieved owing to the Schottky gate being recessed to the undoped AlGaAs region. The typical G_m is 600 mS/mm and the f_T is 110 GHz for the power HEMTs.

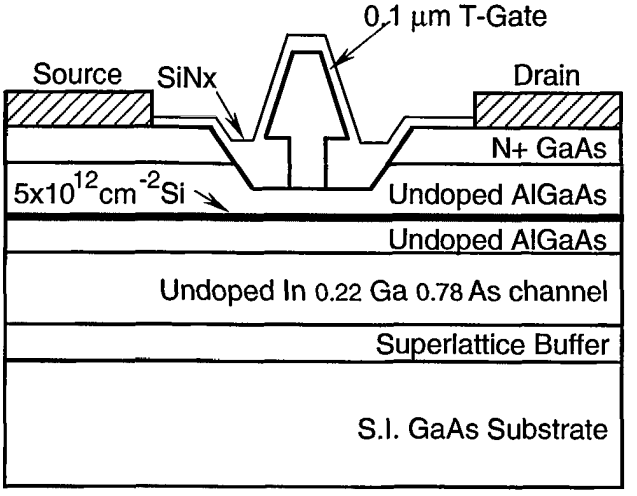
B. GaAs-Based HBT

The HBT devices is suitable for low phase noise source due to its superior $1/f$ noise. In this process, the *npn* HBTs, PIN diodes, and Schottky diodes were integrated on the same substrate. A 2 × 10 μm^2 quad emitter (four-emitter fingers) HBT device had a typical f_T of 22 GHz and f_{max} of 40 GHz. The HBT device process and characteristics were documented in [8]-[10].

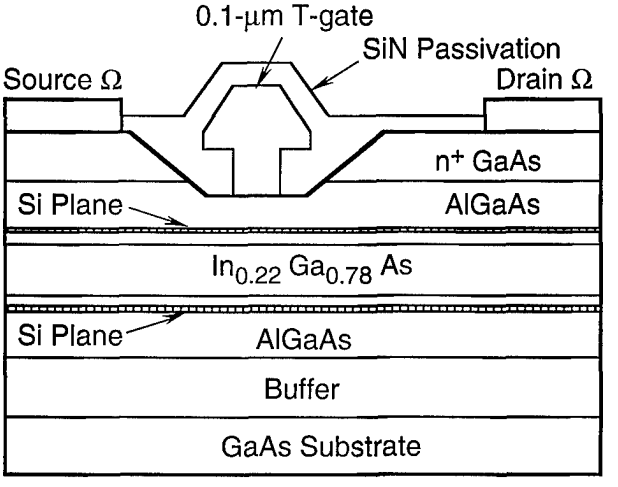
C. Production MMIC Processes

Both the low noise and power GaAs-based 0.1-μm PM HEMT MMIC processes were transferred into production line and passivated the HEMT device with SiN for reliability concern. Both the processes has been validated in our HEMT MMIC production line and shares many of the same process steps as the 0.2-μm gate length low noise and 0.15-μm gate length power HEMT production MMIC processes, ensuring high producibility [55]-[56]. The device performance is only slightly degraded compared with the unpassivated device, but is enough to provide performance for W-band applications (G_m is 600 mS/mm, f_T = 100 GHz, and f_{max} = 250 GHz). The low noise and power HEMT device structures are shown in Fig. 1.

A 2-μm self-aligned base ohmic metal (SABM) HBT process on three-inch diameter GaAs MBE substrate was selected as the baseline HBT process in TRW production line. This process has also been validated, and demonstrated high volume production of microwave, analogue and digital



(a)



(b)

Fig. 1. Device cross section of the 0.1-μm gate planar doped channel PM AlGaAs/InGaAs/GaAs (a) low noise, and (b) power HEMT.

monolithic circuits. Fig. 2 shows the device layer structure of the integrated HBT, PIN, and Schottky diodes. Both HEMT and HBT production processes utilize the reactive ion etching (RIE) process to fabricate back side via holes for grounding. Extensive characterization and statistical process control are employed for material analysis, electron beam lithography, metal-insulator-metal (MIM) dielectric thickness and capacitance, metal thickness, linewidth, resistivity, dc and RF device electrical parameters. The standard thickness of the GaAs substrate is 4 mil (≈ 100 μm).

III. DEVELOPED W-BAND MONOLITHIC COMPONENTS

During past five years, numerous W-band MMIC components have been developed at TRW. These components includes amplification components such as low noise amplifiers, power amplifiers; control components such as mixers, switches, detectors, phase shifters; frequency sources components such as oscillators and frequency multipliers; and multi-functional chips in high-level integration such as single-chip downconverter, pre-amplified detector, and transceiver.

A. Amplification Components

W-band MMIC development at TRW started from the most fundamental active amplification components using HEMT devices. LNAs were developed first and followed by PAs. A number of successful W-band amplifiers were described as follows.

Low Noise Amplifiers

A number of W-band low noise amplifiers were developed since 1990 using the unpassivated 0.1-μm PM AlGaAs/InGaAs/GaAs HEMT MMIC technology with microstrip lines [11]-[13]. A three-stage LNA demonstrating state-of-the-art performance for GaAs-based HEMT LNA of 3.5-dB noise figure with 21-dB gain at 94 GHz was reported in 1992 [13]. Fig. 3 shows a photograph of the monolithic three-stage LNA. Each stage utilizes a 40-μm HEMT with four gate fingers. The circuit was designed for low noise figure based on reactive matching technique. The matching networks were quasi-low pass filter structures and realized by cascade high-low impedance microstrip lines. Edge coupled lines were used for dc block and radial stubs are employed for RF by pass. Shunt resistors are used to ensure bias network stability. In the mean time, a two-stage design using conductor back coplanar waveguide (CBCPW) using the same HEMT devices was also demonstrated with 11-dB gain and 4.6-dB noise figure [14].

The low noise 0.1-μm PM AlGaAs/InGaAs/GaAs HEMT MMIC processes was transferred to production line and ready for mass production in 1992 with only slight degradation in performance. A three-stage LNA has shown an RF yield of more than 50% on 3 wafer lots (total of 16 wafers) with 17-dB gain and 5.5-dB noise figure as screening criteria [16]. The chip photograph and circuit block diagram are shown in Fig. 4.

Power Amplifiers

Power amplifiers are essential for transmitter applications. A two-stage PA with 100-mW output power at 94 GHz was developed in 1992 [15]. Fig. 5. shows the photograph of this PA chip. This PA chip was a balanced two-stage design with 320-μm device at first stage driving a 640 μm second stage. This chip showed a linear gain of 9 dB, an output power of 20 dBm at 3-dB power gain compression with 6% of power added efficiency (PAE).

The power 0.1-μm PM AlGaAs/InGaAs/GaAs HEMT MMIC processes was transferred to production line in 1993. A two-stage PA cascading two single-stage balanced amplifier has a measured output power of 102 mW at 3-dB compression with a small signal gain of 9 dB at 94 GHz [17]. The PA had the same device periphery as the PA developed earlier using unpassivated HEMTs [15] with a similar output power performance. Moreover, a sampling of 37 sites of this PA over six wafers have been tested for gain

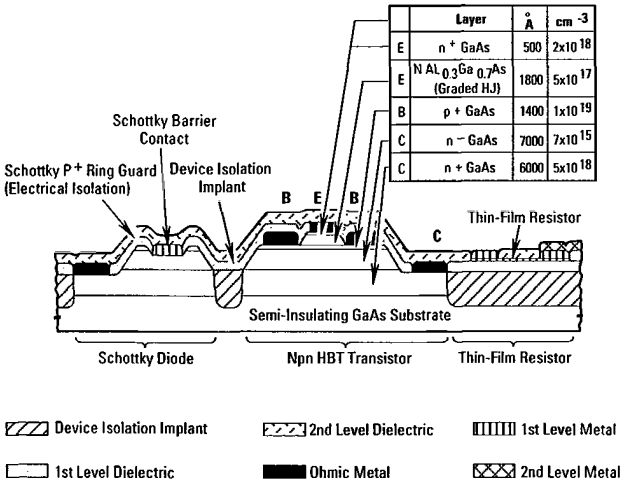


Fig. 2. The device layer structure of the GaAs-based integrated HBT, PIN, and Schottky diodes.

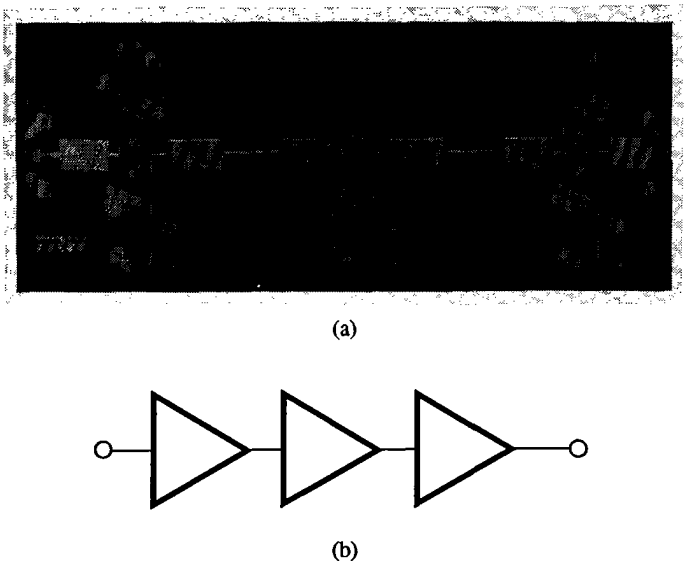
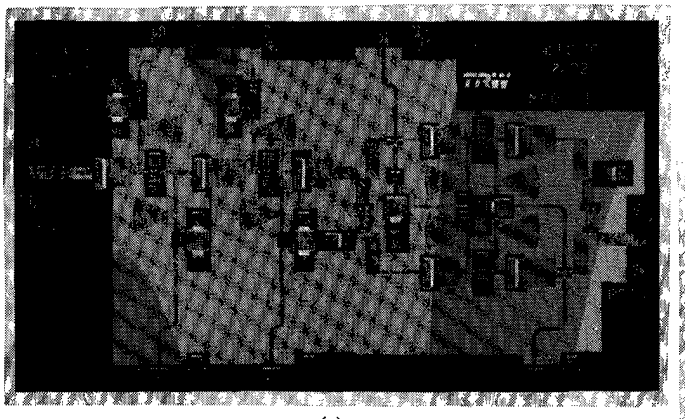


Fig. 3. The photograph of the monolithic three-stage LNA using unpassivated PM low noise HEMT with a chip size of 3.2 × 1.2 mm^2.

performance and showing an RF yield of 37% by using the 7-dB small signal gain criteria from 92 to 96 GHz. Good uniformity from wafer to wafer was observed. Fig. 6 shows the the chip photograph of the monolithic PA.

The other highlight of the PA development is a push-pull PA [18] as shown in Fig. 7, which is the first reported monolithic push-pull amplifier at millimeter-wave frequencies. The novel design approach utilized a push-pull topology to take advantage of a virtual ground between the device pair, eliminating the series feedback of the via hole inductance, and thus improving the performance of power amplifier at millimeter-wave frequencies. The measurement results showed a small signal gain of 13 dB, a saturated output power of 19.4 dBm at 90 GHz. The best power added efficiency of 13.3% was achieved at an output power of 18.8 dBm under a lower bias condition. The gain and efficiency results represent state-of-the-art performance.



(a)

(b)

Fig. 4. The (a) photograph, and (b) block diagram of the monolithic three-stage LNA using passivated production process PM power HEMT with a chip size of 3.5×2 mm^2.

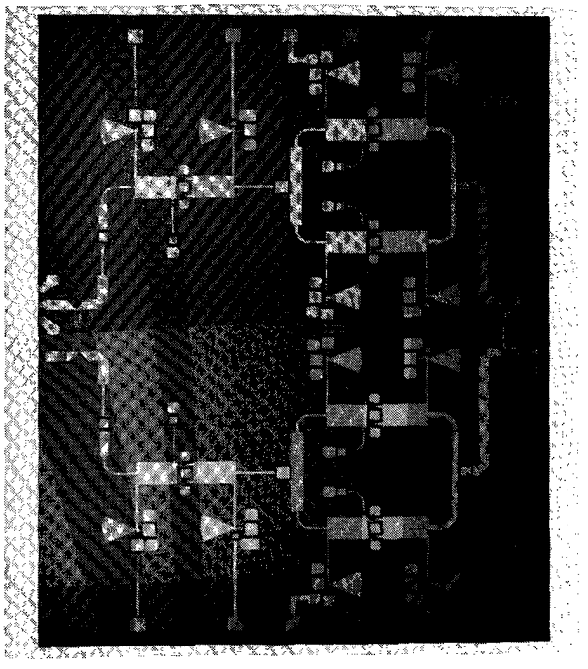


Fig. 5. The photograph of the monolithic two-stage balanced PA using unpassivated PM power HEMT with a chip size of 3.45×4.1 mm^2.

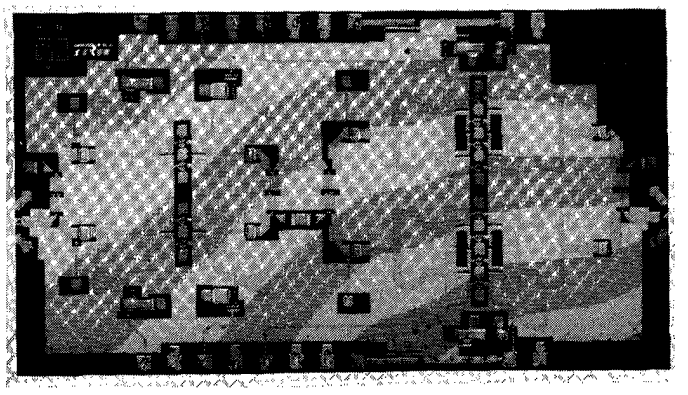
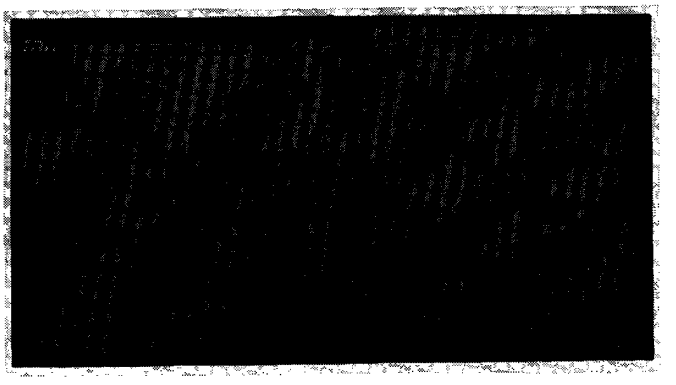
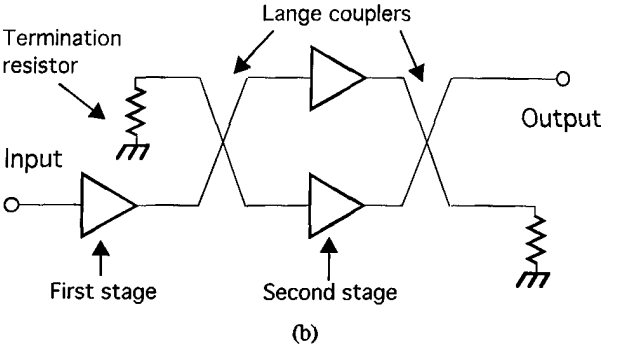


Fig. 6. The photograph of the monolithic two-stage balanced PA using passivated production process PM power HEMT with a chip size of 4.4×2.4 mm^2.



(a)



(b)

Fig. 7. The (a) photograph, and (b) block diagram of the monolithic two-stage push-pull PA using passivated production process PM power HEMT with a chip size of 4.4×2.4 mm^2.

A larger output device periphery (640-μm driving 1280-μm devices) monolithic W-band PA has been developed to increase the output power performance [19]. This monolithic two-stage balanced power amplifier has demonstrated a small signal gain of 7 dB and an output power of 175 mW at 90 GHz, which is the highest output power for a monolithic amplifier at this frequency.

B. Control Components

Control components are typically built using passive devices such as diodes and passive HEMTs. In our W-band MMICs, the gate-HEMT Schottky diodes were used to construct mixers and detectors due to its compatibility of the GaAs-based HEMT process, and HEMT transistors operating in the passive mode were used to realize the switching function.

Mixers

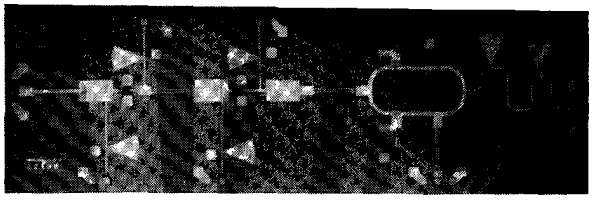
Schottky diode mixers have been developed for W-band frequency converters. The Schottky diode was constructed by connecting the source and drain metal of a HEMT device as the cathode of the diode and the gate pad was used as the anode. Thus the diode fabrication process remains compatible with our W-band 0.1-μm PM AlGaAs/InGaAs/GaAs HEMT process. The series resistance, reverse leakage current and the ideality factor of the diode were calculated from the diode dc-IV characteristics, while the junction capacitor and other parasitics were obtained from the measured S-parameters at various bias conditions. A cut-off frequency of 450 GHz for a two-finger 16-μm diode was estimated from the model. The diode modeling procedure was presented in [20].

A monolithic downconverter which integrated a single-balanced diode mixer and a two-stage front-end LNA has been reported [20], as shown in Fig. 8. The diode mixer had a conversion loss of 8-dB for a 94-GHz RF input with a 1-GHz IF output at an LO drive of 10 dBm. The complete downconverter had a conversion gain of 5 dB with a noise figure of 6 dB. In order to eliminate the image frequency, an image rejection mixer (IRM) was developed [21], as shown in Fig. 9. It was constructed using two single-balanced diode mixers, a Lange coupler, and a Wilkinson power divider. The RF and LO signals were in quadrature and in phase, respectively. The 90° Lange coupler was used for the RF port to achieve good input return loss. The IRM has shown more than 15 dB image rejection.

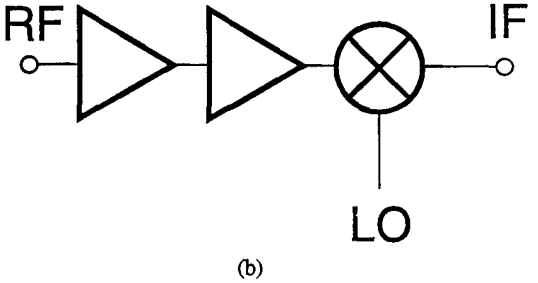
A monolithic image rejection downconverter (IRD) was developed using the HEMT MMIC production process [7]. This IRD consists of the three stage LNA [16] and an IRM modified from [21]. The complete monolithic IRD has a measured conversion gain of 7-9 dB with a single side band (SSB) noise figure of 6 dB when downconverting 93-95 GHz RF signal to 50-500 MHz. The complete downconversion requires an LO power of 9 dBm. The image rejection level is better than 16 dB.

Detector

The Schottky diodes may also used to build detectors.

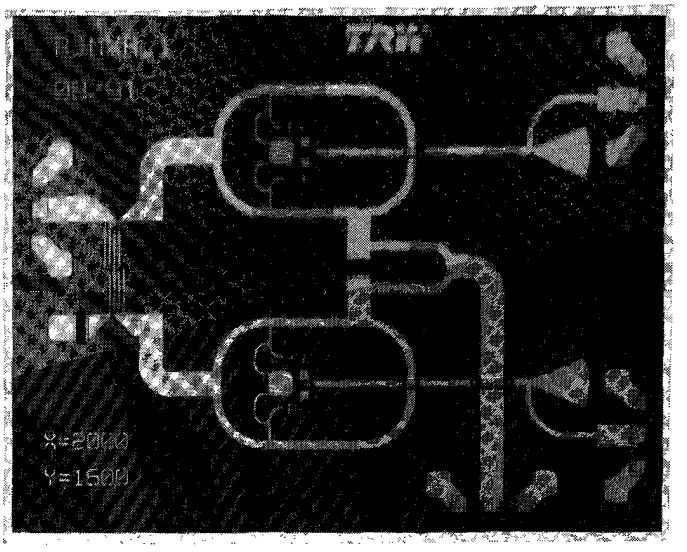


(a)

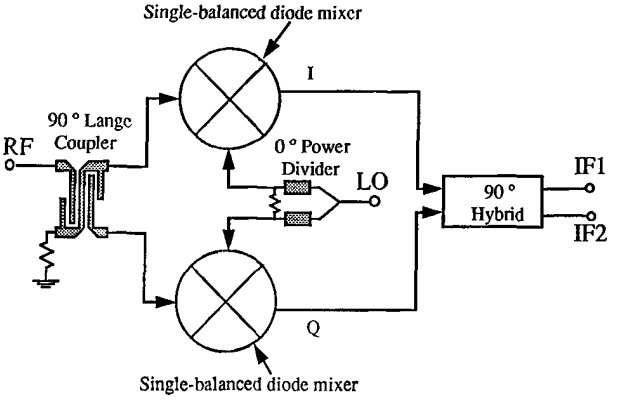


(b)

Fig. 8. The (a) photograph, and (b) block diagram of the monolithic downconverter with a chip size of 4 × 1.2 mm^2.



(a)



(b)

Fig. 9. The (a) photograph, and (b) block diagram of the image rejection mixer with a chip size of 2 × 1.6 mm^2.

A monolithic W-band preamplified diode detector was developed based on 0.1-μm pseudomorphic AlGaAs/InGaAs/GaAs HEMT technology [22]. This chip consists of a Schottky diode detector with a two-stage W-band low noise amplifier as shown in Fig. 10.

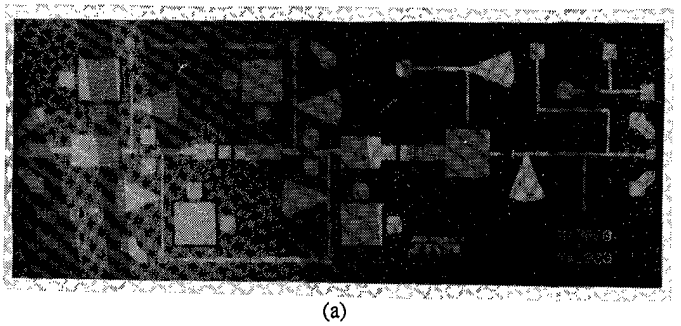
The detector employed a two-finger 16-μm diode. The diode matching network was constructed with cascaded high-low impedance microstrip lines. A quarter-wave high impedance line together with a radial stub was used to apply the diode bias voltage. A low-pass-filter with a shunt microstrip line is placed at the diode output port for low frequency signal and dc return. The Schottky-diode detector chip alone was tested with a measured responsivity of a diode detector was 7.2 V/mW at 94 GHz and the tangential sensitivity was -45 dBm for a 1-MHz video bandwidth. With the two-stage preamplified LNA in the front-end, the responsivity increased to 300 V/mW at 94 GHz and tangential sensitivity became -62 dBm for the complete MMIC preamplified detector.

A higher sensitivity preamplified detector was built using cascade of two monolithic three-stage W-band LNAs with the preamplified detector chip [22]. Edge coupled lines fabricated on a 5-mil thick quartz were placed between the monolithic chips for dc blocking. The measured tangential sensitivity is about -85 dBm, which is suitable for passive imaging application.

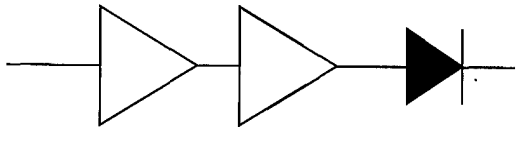
Phase Shifter and Switch

Conventional FET or PIN switches have the disadvantages such as high current, high insertion loss, parasitic effect, and etc. at high frequency [23], [58]-[59]. In order to avoid these disadvantages, a novel multi-functional balanced switching low noise amplifier (BSLNA) was proposed to implement a low noise W-band switch and a broad band 180° phase shifter, where the switching element loss does not increase the LNA noise figure significantly [24].

The monolithic BSLNA at W-band frequencies was developed using the 0.1-μm PM HEMT technology, as shown in Fig. 11. The W-band BSLNA, had a noise figure of 6 dB with associated gain of 7 dB from 92 to 96 GHz, an isolation of higher than 20 dB for a 5-GHz bandwidth, and a 180 ± 5° phase shift from 80 to 110 GHz. Potential applications of these monolithic MMW BSLNAs include on-off keying (OFK) or binary phase shift keying (BPSK) in communication systems. Further modifications of this BSLNA structure enable the circuit to have several different functions, such as a single-pole double throw switch and a cross-bar switch to interchange two signal paths.

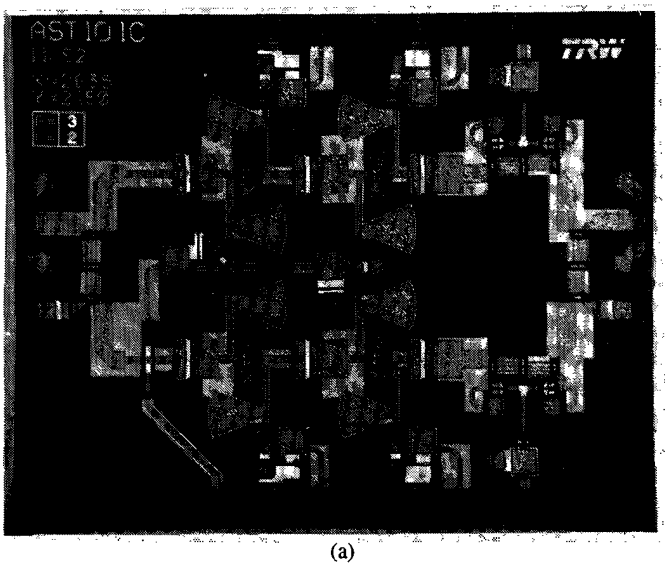


(a)

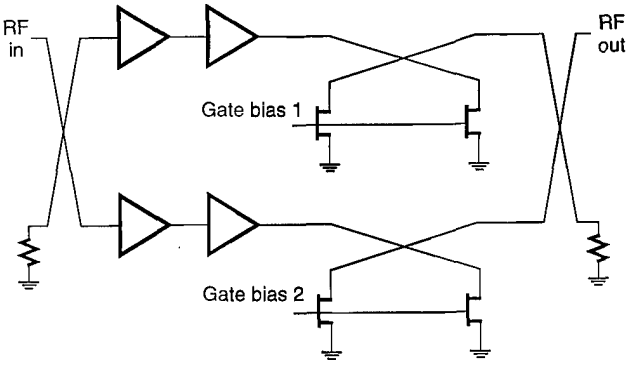


(b)

Fig. 10. The (a) photograph, and (b) block diagram of the preamplified detector with a chip size of 3.2 × 1.2 mm^2.



(a)



Gate bias 1	Gate bias 2	Switch	Phase Shifter
0.2 V	0.2 V	Off	
0.2 V	-1 V	On	ϕ
-1 V	0.2 V	On	ϕ + 180°
-1 V	-1 V	Off	

(b)

Fig. 11. The (a) photograph, and (b) block diagram of the balanced switching LNA (can be used as a 180° phase shifter, an SPDT switch or QPSK modulator) with a chip size of 2.6 × 2.2 mm^2.

C. Frequency Sources

Low phase noise and stable frequency sources are required in W-band systems. These frequency sources can be generated by either fundamental frequency oscillators or using lower frequency oscillators in conjunction with frequency multipliers to obtain desired frequencies. There are advantages for the later approach. It is easier to achieve low phase noise for an oscillator operating at lower frequency. Also, a wider bandwidth can be achieved owing to frequency multiplication and reduced parasitics at lower frequency. On the other hand, the fundamental frequency oscillator approach provides a simple means to achieve a W-band source.

W-band Fundamental Frequency VCOs

In the past, most of the high frequency oscillators have been based on two terminal devices such as IMPATT and Gunn diodes. Although they have demonstrated good power performance, their processes are not compatible with other MMIC components. GaAs HEMT based VCOs using both low noise and power HEMTs have been developed [25]. Fig. 12 shows the photograph of the monolithic W-band VCOs using a four-finger 80 μm HEMT device. The VCO is a common gate design and the oscillation frequency is adjusted by the gate voltage. The output power was coupled out of the drain terminal via microstrip edge coupled lines. An output power of 7.6 dBm at 92 GHz was obtained by using the low noise HEMT and 8.8 dBm at 90.5 GHz by using the power HEMT. The tuning range was about 800 MHz with a phase noise of -70 dBc/Hz at 1-MHz offset.

K-band VCO

A K-band (23.5 GHz) GaAs-based HBT VCO was developed for W-band source application. The 23.5-GHz VCO utilized a common-collector bias-tuned VCO topology and 2×10 μm^2 quad emitter HBT device [25], as shown in Fig 13. The desired tuning bandwidth is 1 GHz. A self-biased feedback buffer amplifier using the same size device as that of the VCO was cascaded with the VCO to desensitize the loading effect of the output port and provide higher output power. The complete chip was designed for a single +5V bias operation. This VCO exhibited an average tuning bandwidth of 1 GHz (23 to 24 GHz) for the tuning voltage of 1.8-3.6V. The output power varied from 3 to 5 dBm with a typical phase noise about -104 dBc/Hz at 1 MHz offset.

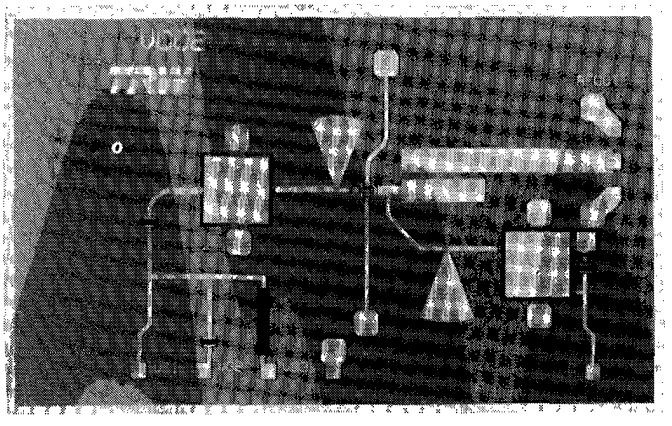


Fig. 12. The photograph of the W-band fundamental mode HEMT VCO with a chip size of 2.1×1.2 mm^2.

K- to W-band Frequency Quadrupler

The frequency quadrupler consists of a 23.5 to 47 GHz HEMT doubler, a 47 to 94 GHz HEMT doubler, and a 47 GHz two-stage buffer amplifier between the two doublers [26] as shown in Fig. 14. The frequency doublers utilized four-finger 80-μm HEMT devices. The design followed a conventional single-ended common source configuration design procedure described in [57]. An open stub providing an RF short at the fundamental frequency was employed in output matching network to suppress the fundamental signal. The HEMT device operates near its pinch-off condition in order to obtain optimal non-linearity for maximum output power of the second harmonic signal, thus good dc to RF conversion efficiency was achieved. The 47 GHz buffer amplifier was a single-ended two-stage design with resistive feed back at first stage for stability consideration. A conversion loss of 5 dB was obtained at an input power of 2 dBm at 24.5 GHz. During the test, the HEMT devices in the doublers were biased at 2 V drain voltages with gate biased near pinch-off region.

W-band Source Module

A W-band source module providing 4-GHz tuning bandwidth (92.5-96.5 GHz) was developed [27] using three chips: a 23.5 GHz HBT VCO [27], a 23.5 to 94 GHz HEMT frequency quadrupler [26] and the W-band three-stage HEMT output amplifier [16], all fabricated in TRW production lines. The source module exhibited a measured output power of 3 dBm at 94-95 GHz and a 3-dB tuning bandwidth greater than 3 GHz, with a phase noise of -92 dBc/Hz at 1 MHz offset. This module demonstrated a new and efficient

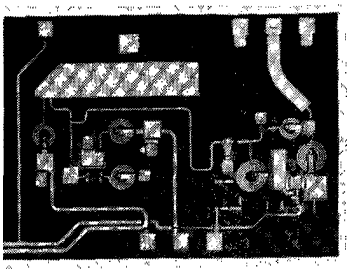


Fig. 13. The photograph of the K-band fundamental mode HBT VCO with a chip size of 2×1.5 mm^2.

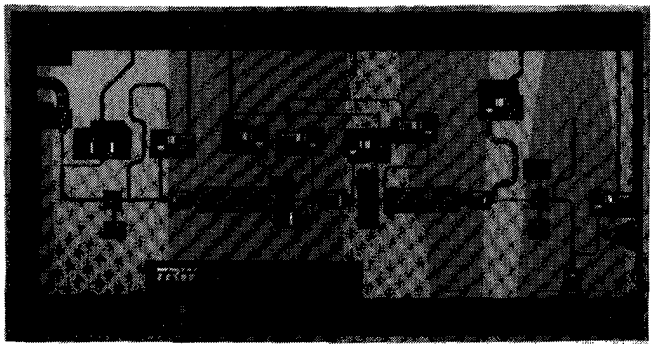


Fig. 14. The photograph of the K- to W-band frequency quadrupler with a chip size of 4×2 mm^2.

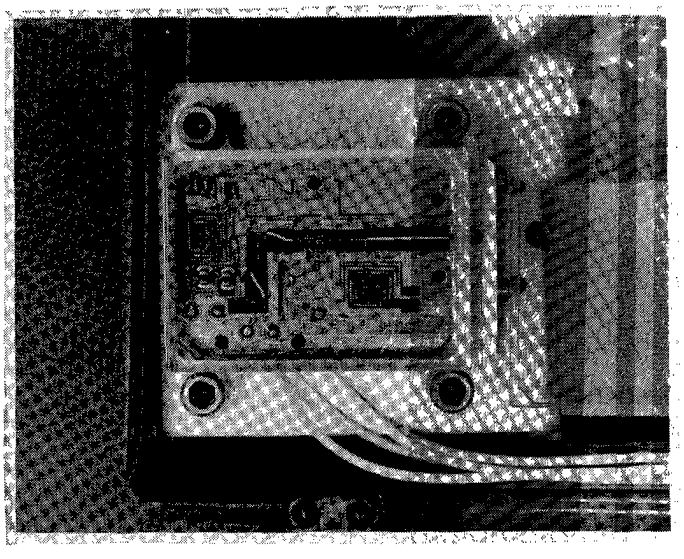
way to implement a high performance W-band source. Its wide tuning bandwidth with good phase noise performance, as well as design simplicity, makes this approach attractive for many W-band system applications. The photograph and block diagram of module hardware are shown in Fig. 15.

IV. MULTI-FUNCTIONAL MMICS AND SYSTEM APPLICATIONS

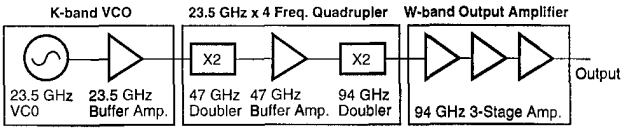
Advance of W-band MMIC technology not only improves the existing system performance but also enables several new system designs. Systems utilizing the developed W-band monolithic LNAs by TRW have been constructed and have demonstrated excellent results. The LNA chip [11] was used in a 94 GHz FM-CW imaging radar for an autonomous aircraft landing system [28]. The ultra low noise W-band LNA [13] has been used in the front end of an image-rejection downconverter for smart munition systems [29]. We further developed several multi-functional W-band MMIC chips for new system applications. One example is the direct-detection radiometric imaging system [32] and another example is the FMCW radar for automotive application [34].

A. Passive Imaging Application

The interest of MMW imagining is primarily due to its capability of imaging through fog, cloud, smoke, sandstorms, and at dark [1]. Owing to lack of MMW components, existing MMW imaging arrays were typically based on the heterodyne approach and relied on a front-end mixer to downconvert an incoming signal to low frequency for signal processing. Not only does this approach suffer from high conversion loss, high noise figure, but also it requires complicated LO signal generation and distribution. The LO problem is aggravated as the imaging array size increases. An alternative approach is to employ a direct-detection architecture which requires a high gain, low noise amplifier (LNA) and a detector circuit [30]. This approach has the advantages of no LO requirement, low dc power consumption, fewer parts, and low cost. These advantages becomes even more pronounced in focal plane staring array systems. However, despite the advantages described, the direct detection imaging receiver has yet to be developed due to the lack of enabling circuit components, in particular, the high gain LNA. Owing to the recent advance in MMW MMIC LNAs, the simple direct detection approach has become feasible.



(a)

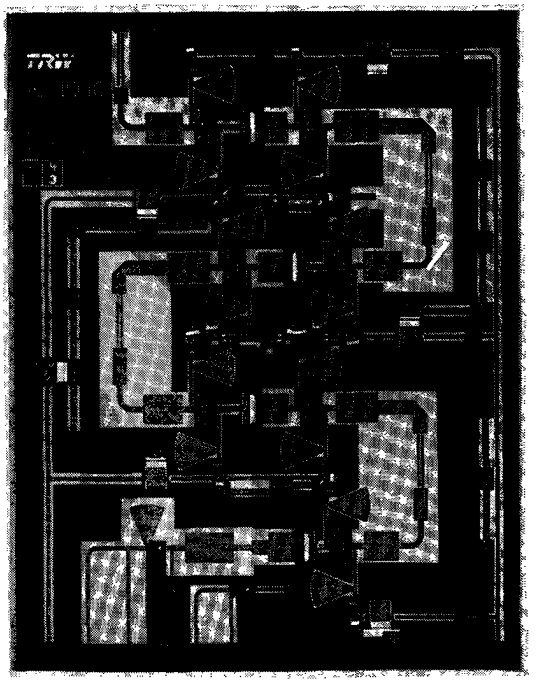


(b)

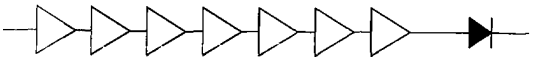
Fig. 15. The (a) photograph, and (b) block diagram of W-band source module hardware.

Monolithic Integrated Seven-Stage LNA and Detector

A W-band monolithic integrated seven-stage LNA and detector on a single chip was developed [31] as shown in Fig. 16. All the gate lines and drain lines of the seven stages were connected together, respectively, for ease of operation. The LNA has a linear gain of 34 dB from 92 to 96 GHz with a noise figure of 4 dB. The complete monolithic chip was tested on-wafer with a W-band noise source as the input signal and showed a net output of 10 mV between the noise source on and off.



(a)



(b)

Fig. 16. The (a) photograph, and (b) block diagram of the W-band seven-stage LNA and detector with a chip size of 2.7×3.5 mm^2.

W-band Radiometer -- Passive Imaging Focal Plane Array

Fig. 17 shows the assembled direct-detection imaging array in 1 × 8 configuration and a single pixel in the array [32]. Each pixel consists of a linear taper slot antenna, a monolithic W-band seven-stage LNA/detector chip [31], and a post video amplifier. Incident radiometric signals are first received by the slot antenna, then slot line coupled to the LNA via a transition, and amplified to a power level at least 10-dB higher than the tangential sensitivity of the following W-band detector. The post amplifier enhances the detector output signal to a level suitable for signal processing.

The imaging array was installed at the focal plane of a 2-foot diameter, f/1.1 Rexolite lens. The whole imaging system was then mounted on an azimuth-elevation drive system for vertical and horizontal scanning. Fig. 18(a) shows a radiometric image of two vehicles in a parking lot taken by the imaging array. The image had a field view of 25° in azimuth and 13° in elevation, containing 252 horizontal and 261 vertical pixels. The average scene distance was 80 feet. The lighter color in Fig. 18(a) indicates colder area, while the darker color indicates the hotter area. The corresponding photo of the same scene is shown in Fig. 18(b).

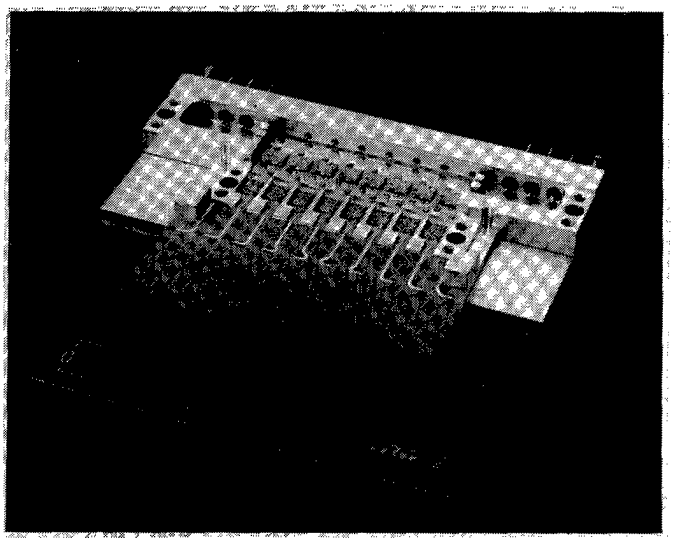
B. Single-Chip Radar Transceiver for Automotive Radar Application

Over the past few years, automotive sensors/radars have become an important and interesting area for microwave and millimeter-wave (MMW) applications. They not only have the huge market potential of the automotive industry but also will play an important safety role in future intelligent vehicle highway systems (IVHS). The forward collision warning (FCW) and autonomous intelligent cruise control (AICC) radars are forward looking automotive radars (FLARs) that require a sensor range of about 100 meters. The FCW radar will detect potential forward collision situations and provide a warning to the driver, while the AICC will operate in an autonomous control loop to maintain a continuous safe following distance.

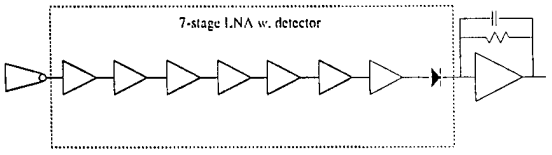
There are two dominant factors, in addition to the performance requirement for all-weather conditions, that drive the technology for automotive radars: cost and hardware size. Low cost is the key factor for consumers to accept the radar as an affordable safety component of their vehicles. The size constraint is essential for easy integration of the radar onto the vehicle without major impact on the vehicle design and performance. These problems may be overcome by W-band MMIC components. Due to the low cost concern, a W-band single-chip transceiver using GaAs-based HEMT technology was developed for the FLAR application.

Single-Chip Radar Transceiver

The block diagram and the photograph of the single-

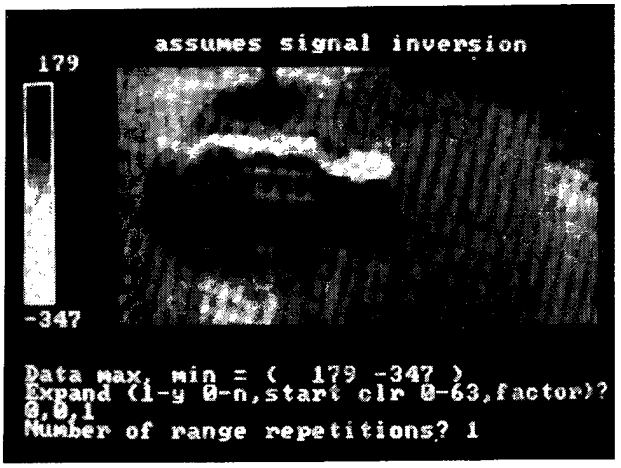


(a)

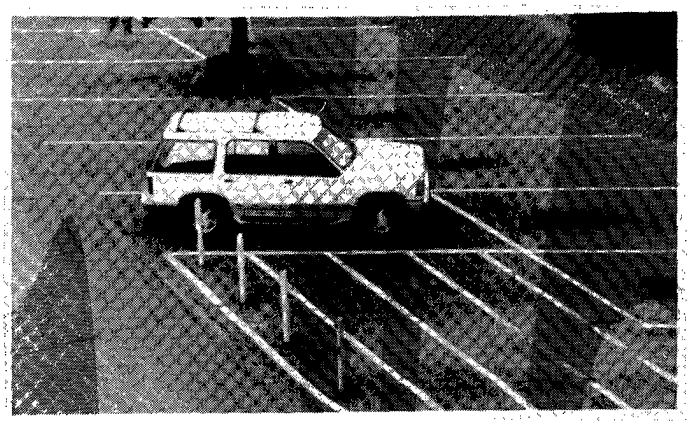


(b)

Fig. 17. (a) The assembled direct-detection radiometric imaging array, (b) the block diagram of a single pixel in the array.



(a)



(b)

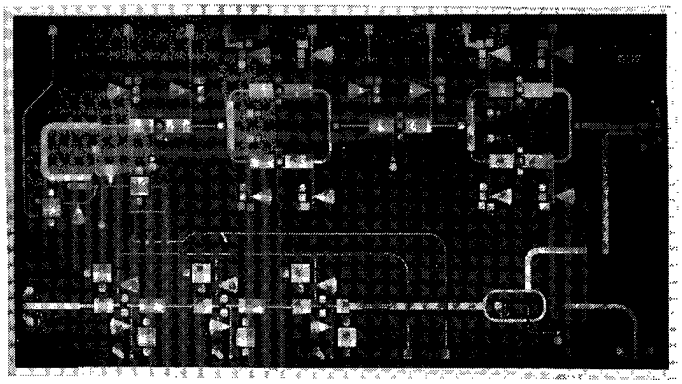
Fig. 18. (a) Radiometric image of a parking lot scene taken by the direct-detection radiometric imaging array, (b) photograph of the same scene for comparison.

chip W-band FMCW transceiver [33] is shown in Fig. 19. This transceiver chip consists of a VCO [25], a PA [15], a mixer [20], and an LNA [13] which were all developed in the past. The transmit and receive channels used separate antennas for better isolation. An FMCW signal was generated from the VCO and fed to the transmit amplifiers. A portion of the transmit power was coupled back to the receiving channel and used as an LO source for the mixer. The transmitter had more than 10 dBm output power with a tuning range of 500 MHz around 92.5 GHz. The receiving chain had a conversion gain of 9 dB. However, the receiving noise figure was higher than 20 dB for IFs less than 10 MHz because of the $1/f$ noise of the gate-HEMT diode mixer. This might impact the system sensitivity for the homodyne FMCW operation for an object at a long range.

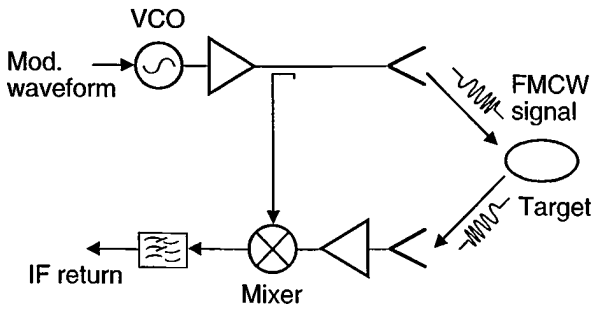
Another radar structure, namely, the superheterodyne FMCW radar using a single-chip single sideband transceiver [55] was also investigated to improve the receiving noise figure and thus the system sensitivity. This approach utilized an off-chip W-band LO source [27] and an extra LO source of 300 MHz for the second downconversion to avoid the high $1/f$ noise of the mixer [55]. The SNR ratio achieved more than 10 dB improvement over the single chip homodyne FMCW radar [33]. However, the system was more complicated with a higher associated cost.

Forward Looking Automotive Radar

The transceiver MMIC [33] and voltage-regulated dc power supply were all packaged in a housing. The top portion included the MMIC and two finline transitions for interface of waveguide and transceiver chip. Two SMA connectors were used for IF signal output and VCO tuning voltage input, respectively. Eight voltage regulators were assembled in the bottom housing. The complete unit required a +5V power supply and consumed less than 250 mA total current.



(a)



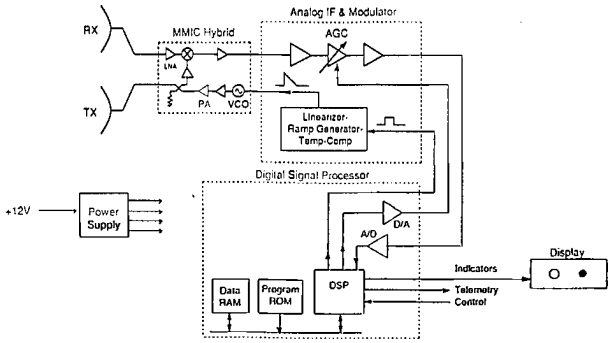
(b)

Fig. 19. The (a) photograph, and (b) block diagram of the single-chip W-band homodyne FMCW transceiver with a chip size of 6.9 × 3.6 mm^2.

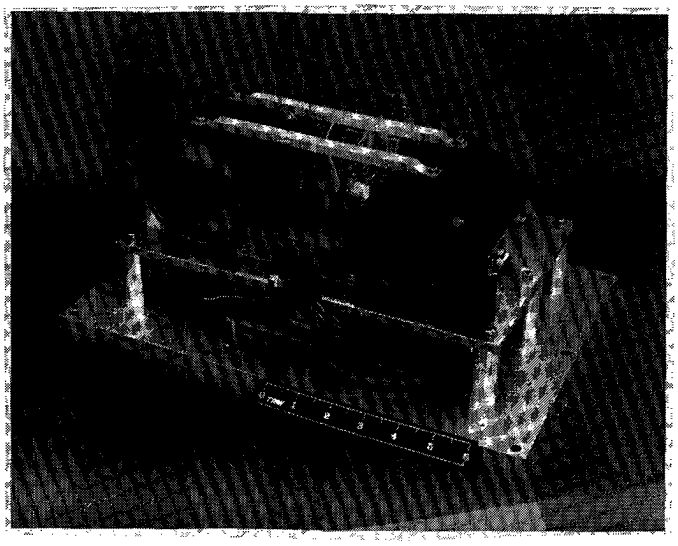
Fig. 20(a) shows the block diagram of the prototype radar. While the requirements for RF front-end and IF receiver may be generic for different types of radars, the antenna and DSP requirements are different for different applications. The photograph of this prototype radar unit is shown in Fig. 20(b) [34].

Since FLAR is intended for looking out along the road surface at close ranges, there be many cars in the field of view. Also, large amounts of clutter will be observed in the form of road signs, guard rails, overpasses, cars parked along side the roadway, etc. These factors require sophisticated signal processing to ignore clutter and maintain tracking on the target of interest.

An AICC requires high precision on the range and velocity measurements because they are part of a closed-loop system servoing around zero relative velocity. A colli-



(a)



(b)

Fig. 20. The (a) function block diagram, and (b) photograph, of the FLAR prototype unit.

sion warning radar does not require high precision, but requires the capability to track high closing velocities and acquire new targets rapidly. It also requires a detection probability as close to one as possible and a false alarm probability as close to zero as possible.

As shown in Fig. 21, the prototype FLAR was mounted on a platform in front of the grill of the vehicle. The platform could be slightly adjusted to align the antenna beam in a proper direction. The test result on a highway showed that the radar acquired a target at 35 meters range and the range varied as the driver changed the vehicle speed to vary the distance between the target and driver. The radar tracked the target up to more than 100 meters although the target range resolution decreased at longer distance, indicating that the VCO linearity needs to be further improved for better results. It also featured an audio signal to warn the driver of a potential obstacle. In a closed-loop test with the cruise control, this prototype FLAR automatically adjusted the vehicle speed to maintain a constant safe following distance from the lead vehicle.

V. FUTURE TREND OF W-BAND MMIC DEVELOPMENT

New technologies will certainly direct new MMIC development. We will discuss the following issues regarding future W-band MMIC trend: new MMIC material and fabrication technologies, MMW monolithic circuit design/analysis methodology, and on-wafer MMIC testing.

A. New MMIC Technologies

The new MMIC technologies include InP-based HEMT and HBT devices for better performance, selective epitaxy for monolithic integration of HEMT and HBT devices, and thin substrate (less than 3 mil) for improved thermal conductivity in high power device application.

InP-Based HEMT and HBT

The InAlAs/InGaAs/InP-based HEMTs offer advantages over GaAs-based HEMTs as amplification devices due to the higher electron peak drift velocity and higher mobility of the InGaAs channel in the former devices. Lattice matched and PM InAlAs/InGaAs/InP HEMTs exhibit a high channel electron mobility and peak velocity and a large bandgap discontinuity between the InAlAs and InGaAs layers. These lead to higher device gain, higher cutoff frequency, lower noise figure and lower dc power consumption compared to their GaAs-based counterparts.

A number of InP-based HEMT LNAs have developed using the ultra high performance devices [35] in W-band and higher frequencies. The performance highlights of W-band MMIC LNAs include a three-stage design which has demonstrated a noise figure of 3.3 dB and an associated gain of 20 dB with a dc power consumption of less than 20 mW. This is three times less than the GaAs-based counterpart [36]. In addition, a balanced four-stage LNA exhibited a gain of 23 ± 3 dB from 75 to 110 GHz [37]. The highest frequency MMIC LNA, a 142-GHz two-stage LNA showing a 9-dB small signal gain in Fig. 22 [38], was also developed based on this 0.1-μm InP HEMT device. However, the InP-based HEMT MMIC technology is relatively new compared with the GaAs-based HEMT MMIC technology. Further improvement of InP HEMT MMIC process techniques and circuit performance may be expected.

The InP HBT not only offers higher frequency performance than that of GaAs-base HBT but also lower $1/f$ noise because of the lower surface recombination velocity in InP-based materials and the absence of DX centers in the emitter [39]-[40]. Moreover, the Schottky diode using InP HBT process also has a lower threshold voltage (less than +0.2V, compared with +0.7V of GaAs HBT process), which allows for a low LO power requirement for mixers. A 39-GHz diode single-balanced mixer using the InP HBT process has shown a 7-dB conversion loss with only 1-dBm LO drive [41]. The low phase noise capability of InP HBT frequency sources has been demonstrated via a K-band monolithic VCO [41]. This VCO showed a phase noise 10 dB lower than the GaAs HBT VCO of a similar design. The highest fundamental frequency VCO to date is a 62-GHz VCO using the InP HBT [42]. Further improvement of the frequency performance for the InP HBT device will lead to a W-band



Fig. 21. Photograph showing that the prototype FLAR was mounted on a platform in front of the grill of the vehicle.

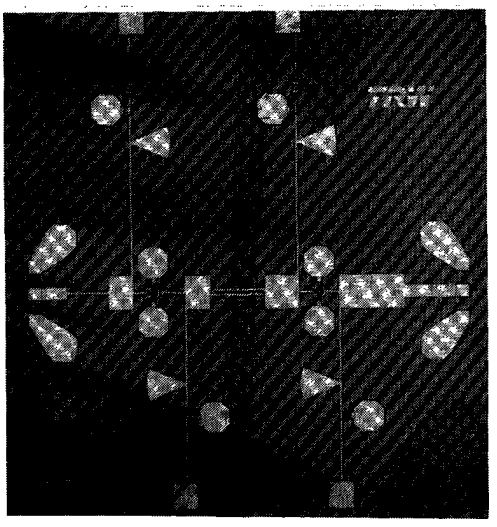


Fig. 22. The photograph of the 140-GHz InP HEMT two-stage LNA with a chip size of 1.8×2 mm^2.

fundamental mode oscillator with good phase noise.

Selective Epitaxial Regrowth Material

The technique for the selective epitaxial regrowth of material enables fabricating the HEMT and HBT/PIN devices on the same substrate and further expands the level of integration of MMICs. A number of GaAs-based MMICs using this technology have been successfully developed at lower frequencies [43]-[44]. For W-band application, further improvement of process technique will be required to maintain the performance of the devices.

Thin Substrate

For high power and high efficiency application, the conventional 4-mil substrate has a limited thermal conductivity. In order to improve the heat sink capability, the substrate thickness of MMIC power amplifier should be thinner. Moreover, on a thinner substrate, the width of microstrip lines are narrower and the backside via holes are more compact. These will lead to a more compact design. A number of 2-mil or thiner substrate PAs have been reported at lower frequencies [45]-[47]. We expect W-band PAs based on 2-mil substrates to be developed in the near future.

B. MMW MMIC Design and Analysis Methodology

A good design/analysis methodology is crucial for a successful W-band monolithic circuit design since simplified assumptions used at microwave frequencies, such as microwave circuit theory and quasi-static models may not be adequate at this frequency. Analysis procedures have been developed and incorporated into TRW's monolithic chips design. These include accurate active device modeling, full-wave electromagnetic (EM) analysis of passive structures.

The device modeling and design/analysis procedure was documented in [48], and the design procedure is presented as shown in Fig. 23. After careful device modeling has been complete, conventional circuit synthesis and simulation using existing models of passive elements should be performed for the initial design. Next, the critical components need to be identified and characterized by EM analysis. Some first-pass-success W-band monolithic LNAs have been demonstrated to have good match between measured and simulated results [48]. This demonstrates the effectiveness of the design/analysis methodology.

Regarding chip cost concern, the compact MMIC design is always favorable. In contrast to designs at lower frequency, the coupling effect between elements will be much more severe at W-band, particularly for the compact designs. To analyze (or even to tweak) complete matching structures, high-speed EM analysis tools will be required. Further implementation of the "tuning" or "optimization" capability in EM analysis tools, analogous to those existing in the circuit-based simulation tools, will be very helpful for the compact MMIC design at high frequencies.

C. W-band MMIC On-wafer Testing

The on-wafer testing capability of W-band MMIC has recently been established due to the development W-band on-wafer probes [49]-[50], enhancing testing throughput as well as greatly reducing testing costs. Various on-wafer characterization capabilities have been developed at TRW including testing of amplifier small signal gain and noise figure [51], mixer conversion gain/loss and noise figure [52], VCO tuning range and output power, and detector sensitivity. To remove the thermal effect due to dc-bias, a pulsed-dc and pulsed RF W-band amplifier on-wafer testing has been demonstrated to emulate an isothermal environment [53]. The on-wafer testing capability has been extremely beneficial to our W-band MMIC development. We can foresee that the role of W-band on-wafer testing will be even more important for future MMIC development.

VI. SUMMARY

The development status of W-band MMIC in the past five years at TRW has been summarized in this paper. Superior device performance, repeatable MMIC process, and rigorous MMIC design/analysis methodology are essential for W-band monolithic MMIC development. The system applications for these W-band MMICs have presented. New technologies and the future trend for W-band MMICs have also been addressed. With succeeding new technologies, both higher performance and level of integration for W-band MMICs are expected. Moreover, reducing the cost and enhancing the reliability of current W-band MMICs are also important to facilitate insertion into alternative system applications.

ACKNOWLEDGMENT

The authors would like to acknowledge the following TRW colleagues for their help in this work: Drs. J. Berenz, B. R. Allen, P. D. Chow, for their discussions and sugges-

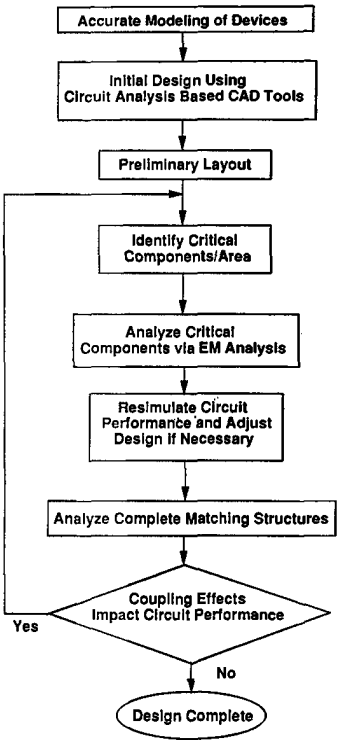


Fig. 23. The flow chart showing the MMW monolithic design procedure using EM analysis tool.

tions; our design team members M. Aust, T. N. Ton, Dr. Y. Hwang for their chip design effort; K. L. Tan, Drs. R. Lai, M. Biedenbender and their HEMT process team for the contribution on HEMT MMIC fabrication, L. Tran, A. Oki, Dr. J. Cowles and the HBT process team for their contribution on the HBT MMIC fabrication; Drs. D. C. Streit, T. Block and their MBE team for their contribution on the material growth; Dr. P. H. Liu and his team for the contribution on E-beam lithography; S. Esparza, P. Yocom, E. Barnachea, S. Chen, D. Garske, D. Cisneros, S. Back, the Hayashibaras, Drs. D. C. Yang, T. W. Huang, E. W. Lin, for their help on the MMIC testing; T. Trinh and his group for the layout support; Drs. P. Lee, L. Yujiri, M. Mussetto and their team for the passive imaging system definition and test; R. Dixit, A. Lawrence, J. Yonaki, M. Yu, and the automotive radar team for the construction of FLAR system and testing effort. Thanks also go to the members and managers of the Space and Electronics Group and Transportation Electronics Division of TRW for technical support and encouragement. Special thanks go to Dr. T. H. Chen for his contribution on the consulting of MMIC design and modeling.

REFERENCES

[1] E. S. Roseblum, "Atmosphere absorption of 10-400 KMQS radiation: summary and biography up to 1961," *Microwave Journal*, vol. 4, pp. 91-96, March, 1961.

[2] Ravender Goyal, *Monolithic Microwave Integrated Circuits: Technology & Design*, ch. 1, Artech House, Norwood, MA, 1989.

[3] Fazal Ali and Aditya Gupta, *HEMTs & HBTs: Devices Fabrication, and Circuits*, Artech House, Norwood, MA, 1991.

[4] K. L. Tan, R. M. Dia, D. C. Streit, T. Lin, T. Q. Trinh, A. C. Han, P. H. Liu, P. D. Chow, and H. C. Yen, "94 GHz 0.1 μm T-gate low noise pseudomorphic InGaAs HEMTs," *IEEE Electronic Device Letters*, vol. 11, no. 12., pp. 585-587, Dec., 1990.

[5] D. C. Streit, K. L. Tan, R. M. Dia, J. K. Liu, A. C. Han, J. R. Velebir, S. K. Wang, T. Q. Trinh, P. -M. D. Chow, and H. C. Yen, "High gain W-band pseudomorphic InGaAs power HEMT's," *IEEE Electronic Device Letters*, vol. 12, no. 4, pp. 149-150, April, 1991.

[6] K. L. Tan, P. H. Liu, D. C. Streit, R. Dia, A. C. Han, A. Freudenthal, J. Velebir, K. Stolt, J. Lee, M. Biedenbender, R. Lai, H. Wang, and B. Allen, "A high performance 0.1 μm pseudomorphic AlGaAs/InGaAs HEMT process for W-band MMICs," *14th Annual IEEE GaAs IC Symposium Digest*, pp. 251-254, Miami, FL, Oct., 1992.

[7] M. Biedenbender, R. Lai, J. Lee, S. Chen, K. L. Tan, P. H. Liu, D. C. Streit, B. Allen, and H. Wang, "A 0.1 μm W-band HEMT production process for high yield and high performance low noise and power MMICs," *16th Annual IEEE GaAs IC Symposium Digest*, pp. 323-327, Philadelphia, PA, Oct., 1994.

[8] K. W. Kobayashi et al., "HBT low power consumption 2-4.5 GHz variable gain feedback amplifier," *14th Annual IEEE GaAs IC Symposium Digest*, pp. 309-312, Miami, FL, Oct., 1992.

[9] M. E. Kim et al., "GaAs heterojunction bipolar transistor device and IC technology for high performance analog and microwave applications," *IEEE Trans. on Microwave Theory and Tech.*, vol. MTT-37, no. 9, pp. 1286-1303, Sept., 1989.

[10] F. M. Yamada et al., "Reliability analysis of microwave GaAs/AlGaAs HBTs with beryllium and carbon doped base," *1992 IEEE MTT-S International Microwave Symposium Digest*, vol. 2, pp. 739-742, June, 1992.

[11] H. Wang, G. S. Dow, K. Tan, J. Berenz, T. N. Ton, T. S. Lin, P. Liu, D. Streit, P. D. Chow, and B. Allen, "A high performance W-band monolithic InGaAs pseudomorphic HEMT LNA," *1991 IEEE MTT-S International Microwave Symposium Digest*, vol. 3, pp. 943-946, Boston, MA, June, 1991.

[12] H. Wang, K. Tan, G. S. Dow, J. Berenz, D. Garske, P. Rodgers, and G. Hayashibara, "State-of-the-art low noise performance of 94 GHz monolithic amplifiers using 0.1 μm InGaAs/GaAs pseudomorphic HEMT technology," *IEEE International Electronic Device Meeting Digest Technical Digest*, pp. 939-942, Washington D. C., Dec., 1991.

[13] H. Wang, T. N. Ton, K. L. Tan, G. S. Dow, T. H. Chen, K. W. Chang, J. Berenz, B. Allen, P. Liu, D. Streit, G. Hayashibara, and L. C. T. Liu, "An ultra low noise monolithic three-stage amplifier using 0.1 μm InGaAs/GaAs pseudomorphic HEMT technology," *1992 IEEE MTT-S International Microwave Symposium Digest*, vol. 2, pp. 803-806, New Mexico, June, 1992.

[14] T. N. Ton, H. Wang, S. Chen, K. L. Tan, G. S. Dow, B. R. Allen, and J. Berenz, "W-band monolithic pseudomorphic AlGaAs/InGaAs/GaAs HEMT CBCPW LNA," *Electronics Letters*, vol. 29, no. 20, pp. 1804-1805, 30th Sept., 1993.

[15] T. H. Chen, K. L. Tan, G. S. Dow, and H. Wang, K. W. Chang, T. N. Ton, B. Allen, J. Berenz, P. H. Liu, D. Streit, and G. Hayashibara, "A 0.1-W W-band pseudomorphic HEMT MMIC power amplifier," *14th Annual IEEE GaAs IC Symposium Digest*, pp. 71-74, Miami, FL, Oct., 1992.

[16] Huei Wang, Kwo Wei Chang, Thuy Nhung Ton, Michael Biedenbender, Sian Tek Chen, Jane Lee, Gee S. Dow, Kin L. Tan, and Barry R. Allen, "High yield W-Band monolithic HEMT low noise amplifier and image rejection downconverter chips," *IEEE Microwave and Guided Wave Letters*, vol. 3, no. 8, pp. 281-283, Aug., 1993.

[17] M. Aust, H. Wang, M. Biedenbender, R. Lai, D. C. Streit, P. H. Liu, G. S. Dow, and B. R. Allen, "A 94 GHz monolithic balanced power amplifier using 0.1-μm gate GaAs-based HEMT MMIC production process technology," *IEEE Microwave and Guided Wave Letters*, vol. 5, no. 1, pp. 12-15, Jan., 1995.

[18] H. Wang, G. S. Dow, M. Aust, K. W. Chang, R. Lai, M. Biedenbender, D. Streit, and B. Allen, "A novel W-band monolithic push-pull power amplifier," *16th Annual IEEE GaAs IC Symposium Digest*, pp. 77-80, Philadelphia, PA, Oct., 1994.

[19] H. Wang, Y. Hwang, T. H. Chen, M. Biedenbender, D. C. Streit, D. C. W. Lo, G. S. Dow, and B. R. Allen, "A W-band monolithic 175-mW power amplifier," *1995 IEEE MTT-S International Microwave Symposium Digest*, vol. 2, pp. 419-422, Orlando, FL, May, 1995.

[20] K. W. Chang, H. Wang, K. L. Tan, S. B. Bui, T. H. Chen, G. S. Dow, J. Berenz, T. N. Ton, D. C. Garske, T. S. Lin, and L. C. T. Liu, "A W-band monolithic downconverter," *IEEE Trans. on Microwave Theory and Tech.*, vol. 39, no. 12, pp. 1972-1979, Dec., 1991.

[21] T. N. Ton, T. H. Chen, K. W. Chang, H. Wang, K. L. Tan, G. S. Dow, G. Hayashibara, B. Allen, and J. Berenz, "A W-band monolithic InGaAs/GaAs HEMT Schottky diode image reject mixer, *14th Annual IEEE GaAs IC Symposium Digest*, pp. 63-66, Miami, FL, Oct., 1992.

[22] H. Wang, W. Lam, T. N. Ton, C. W. D. Lo, K. L. Tan, G. S. Dow, B. Allen, and J. Berenz, "A monolithic W-band preamplified diode detector," *1993 IEEE MTT-S International Microwave Symposium Digest,* vol. 1, pp. 365-368, Atlanta, GA, June, 1993.

[23] J. Bellantoni et al.," A monolithic high power Ka-band PIN switch," *IEEE Microwave and Millimeter-wave Monolithic Circuits Symposium Digest,* pp. 47-50, 1988.

[24] D. C. W. Lo, H. Wang, B. R. Allen, G. S. Dow, K. W. Chang, M. Biedenbender, and R. Lai, "Novel monolithic multi-functional balanced switching low noise amplifiers," *IEEE Trans. on Microwave Theory and Tech.,* vol. 42, no. 12, part 2, pp. 2629-2634, Dec., 1994. .IP [24]

[25] H. Wang, K. W. Chang, T. H. Chen, K. L. Tan, G. S. Dow, B. Allen, and J. Berenz, "Monolithic W-band VCOs using pseudomorphic AlGaAs/InGaAs/GaAs HEMTs," *14th Annual IEEE GaAs IC Symposium Digest,* pp. 47-50, Miami, FL, Oct., 1992.

[26] H. Wang, K. W. Chang, D. C. W. Lo, K. L. Tan, D. Streit, G. S. Dow, and B. R. Allen "A Monolithic 23.5 to 94 GHz frequency quadrupler using 0.1 μm pseudomorphic AlGaAs/InGaAs/GaAs HEMT technology," *IEEE Microwave and Guided Wave Letters,* vol. 4, no. 3, pp. 77-79, March, 1994.

[27] H. Wang, K. W. Chang, D. Smith, G. S. Dow, K. L. Tan, A. Oki, and B. R. Allen, "A W-band source module using MMICs," *IEEE Trans. on Microwave Theory and Tech.,* vol. 43, no. 5, pp. 1010-1016, May, 1995.

[28] L. Q. Bui et al., "Test results of an experimental autonomous aircraft landing system utilizing a 94 GHz FM-CW imaging radar," *1993 IEEE MTT-S International Microwave Symposium Digest,* vol. 2, pp. 857-860, Atlanta, GA, June, 1993.

[29] G. Hayashibara and H. Wang, T. N. Ton, K. L. Tan, G. S. Dow, T. H. Chen, K. W. Chang, J. Berenz, B. Allen, P. Liu, D. Streit, and L. C. T. Liu, "An ultra low noise W-band monolithic downconverter for smart munitions," *1992 Government Microcircuit Applications Conference (GOMAC) Digest,* pp. 601-604, Las Vegas, NV, Nov., 1992.

[30] S. Weinreb, "Monolithic intcgrated circuit imaging radiometers," *1991 IEEE MTT-S International Microwave Symposium Digest,* vol. 1, pp. 405-408, Boston, MA, June, 1991.

[31] D. C. W. Lo, G. S. Dow, L. Yujiri, S. Chen, M. Biedenbender, H. Wang, M. Mussetto, and B. R. Allen, "A monolithic W-band seven-stage LNA/detector for radiometric imaging application," *Electronics Letters,* vol. 30, no. 13, pp. 1075-1077, 23rd June, 1994.

[32] D. C. W. Lo, L. Yujiri, G. S. Dow, T. N. Ton, M. Mussetto, and B. R. Allen, "A W-band direct-detection radiometric imaging array," *IEEE 1994 Microwave and Millimeter-wave Monolithic Circuits Symposium Digest,* pp. 41-44, San Diego, CA, May, 1994.

[33] K. W. Chang, S. Dow, H. Wang, T. H. Chen, K. Tan, B. Allen, J. Berenz, J. Wheling, and R. Lin, "W-band single-chip transceiver for FMCW radar," *IEEE 1993 Microwave and Millimeter-wave Monolithic Circuits Symposium Digest,* pp. 41-44, Atlanta, GA, June, 1993.

[34] K. W. Chang, H. Wang, G. Shreve, J. Harrison, M. Core, A. Paxton, M. Yu, C. H. Chen, and G. S. Dow, "Forward Looking automotive radar using a W-band single-chip transceiver," *IEEE Trans. on Microwave Theory and Tech.,* vol. 43, no. 7, July, 1995.

[35] K. L. Tan et al., "140 GHz 0.1 μm gate-length pseudomorphic $In_{0.52}Al_{0.48}As/In_{0.60}Ga_{0.40}As/InP$ HEMT," *IEEE International Electronic Device Meeting Digest,* pp. 239-242, Washington D. C., Dec., 1991.

[36] G. I. Ng, R. Lai, Y. Hwang, H. Wang, D. C. W. Lo, T. Block, K. L. Tan, D. C. Streit, R. M. Dia, P. H. Liu, P. D. Chow, and J. Berenz, "A fully passivated ultra low noise W-band monolithic InGaAs/InAlAs/InP HEMT amplifier," *IEEE 1995 Microwave and Millimeter-wave Monolithic Circuits Symposium Digest,* pp. 63-66, Orlando, FL, May, 1995.

[37] H. Wang, R. Lai, S. Chen, and J. Berenz, "A monolithic 75-110 GHz balanced InP-based HEMT amplifier," *IEEE Microwave and Guided Wave Letters,* vol. 3, no. 10, pp. 381-383, Oct., 1993.

[38] H. Wang, R. Lai, D. C. W. Lo, D. C. Streit, P. H. Liu R. M. Dia, M. W. Pospieszalski, and J. Berenz, "A 140-GHz monolithic low noise amplifier," *IEEE Microwave and Guided Wave Letters,* vol. 5, no. 5, pp. 150-152, May, 1995.

[39] S. Tanaka et al, "Low frequency noise performance of self-aligned InAlAs/InGaAs heterojunction transistor," *Electron. Lett.,* vol. 26, no. 18, p. 1439, 1990.

[40] D. Costa et al, "Low frequency noise properties of n-p-n AlGaAs/GaAs heterojunction bipolar transistors," *IEEE Trans. Electron Devices,* vol. 39, no. 10, pp. 2383-2394, Oct. 1992.

[41] L. Tran, J. Cowles, T. Block, H. Wang, J. Yonaki, D. Lo, S. Dow, B. Allen, D. Streit, A. Oki, and S. Loughran, "Monolithic VCO and mixer for Q-band transceiver using InP-based HBT technology," *IEEE 1995 Microwave and Millimeter-wave Monolithic Circuits Symposium Digest,* pp. 101-104, Orlando, FL, May, 1995.

[42] H. Wang, K. W. Chang, D. C. W. Lo, G. S. Dow, L. Tran, J. Cowles, D. Streit, A. Oki, and B. Allen, "A 62-GHz monolithic InP-based HBT VCO," to be in *IEEE Microwave and Guided Wave Letters.*

[43] D. C. Streit, D. K. Umemoto, K. W. Kobayashi, and A. K. Oki, "Monolithic HEMT-HBT integration for novel microwave circuit applications," *16th Annual IEEE GaAs IC Symposium Digest,* pp. 329-332, Philadelphia, PA, Oct., 1994.

[44] P. Huang, W. L. Jones, A. Oki, D. Streit, W. Yamasaki, P. Liu, S. Bui, and B. Nelson, "A 9-16 GHz monolithic low noise amplifier with embedded limiters," *IEEE 1995 Microwave and Millimeter-wave Monolithic Circuits Symposium Digest,* pp. 185-186, Orlando, FL, May, 1995.

[45] J. C. Huang, P. Saledas, J. Wendler, A. Platzker, W. Boulais, S. Shanfield, W. Hoke, P. Lyman, L. Aucoin, A. Miquelarena, C. Bedard, and D. Atwood, "A double-recessed $Al_{0.24}GaAs/In_{0.16}GaAs$ pseudomorphic HEMT for Ka and Q-band power applications," *IEEE Electronic Device Letters,* vol. 14, no. 9, pp. 456-458, Sept., 1993.

[46] P. M. Smith, C. T. Creamer, W. F. Kopp, D. W. Ferguson, P. Ho, and J. R. Willhite, "A high-power Q-band PHEMT for communication terminal applications," *1994 IEEE International Microwave Symposium Digest,* vol. 2, pp. 809-812, San Diego, CA, May, 1994.

[47] W. Lam, M Matloubian, A. Igawa, C. Chou, A. Kurdoghlian, C. Ngo, L. Jelloian, A. Brown, M. Thompson, and L. Larson, "44-GHz high efficiency InP HEMT MMIC power amplifier, *IEEE Microwave and Guided Wave Letters,* vol. 4, no. 8, pp. 277-278, Aug., 1994.

[48] H. Wang, G. S. Dow, B. Allen, T. N. Ton, K. Tan, K. W. Chang, T. H. Chen, J. Berenz, T. S. Lin, P. Liu, D. Streit, S. Bui, J. J. Raggio, and P. D. Chow, "High performance W-band monolithic InGaAs pseudomorphic HEMT LNAs and design/analysis methodology," *IEEE Trans. on Microwave Theory and Tech.,* vol. 40, no. 3, pp. 417-428, March, 1992.

[49] Edward Godshalk, "A W-band Wafer probe," *1993 IEEE MTT-S International Microwave Symposium Digest,* vol. 1, pp. 171-174, Atlanta, GA, June, 1993.

[50] S. M. J. Liu and G. G. Boll, "A new probe for W-band on-wafer measurements," *1993 IEEE MTT-S International Microwave Symposium Digest,* vol. 3, pp. 1335-1338, Atlanta, GA, June, 1993.

[51] S. Chen, D. C. Yang, H. Wang, K. Hayashibara, E. M. Godshalk and B. R. Allen, "A W-band automated on-wafer probing noise figure measurement system," *41st Automatic RF Techniques Group (ARFTG) Conference Digest,* pp. 48-56, Atlanta, GA, June, 1993.

[52] E. W. Lin, D. C. W. Lo, H. Wang, T. W. Huang, M. Biedenbender, G. S. Dow, and B. R. Allen, "On-wafer testing of a W-band HEMT image-rejection downconverter MMIC," *1995 IEEE MTT-S International Microwave Symposium Digest,* vol. 3, pp. 1479-1482, Orlando, FL, May, 1995.

[53] Huei Wang and Mohamed Sayed, "W-band MMIC Characterization in an Isothermal Environment," to be published.

[54] K. W. Chang, H. Wang, G. S. Dow, T. H. Chen, D. C. W. Lo, M. Biedenbender, and B. Allen, "W-band monolithic single sideband transceiver for automotive radar applications," *16th Annual IEEE GaAs IC Symposium Digest,* pp. 85-89, Philadelphia, PA, Oct., 1994.

[55] R. Lai et al, "High power 0.15 μm V-band pseudomorphic InGaAs/AlGaAs/GaAs HEMT," *IEEE Microwave and Guided Wave Letters,* vol. 3, no. 10, pp. 363-365, Oct., 1993.

[56] M. D. Biedenbender et al., "A power HEMT production process for high-efficiency Ka-band MMIC power amplifiers," *15th Annual IEEE GaAs IC Symposium Digest,* pp. 341-344, San Jose, CA, Oct., 1993.

[57] S. A. Maas, *Nonlinear Microwave Circuits,* Artech House, Norwood, MA, 1988.

[58] M. Schindler and A. Morris, " DC-40 GHz and 20-40 GHz MMIC SPDT switches," *IEEE Trans. on Microwave Theory and Techniques,* vol. 35, p. 1486, 1987.

[59] P. Bernkopf, M. Schindler and A. Bertrand," A high power K/Ka-band monolithic T/R switch," *IEEE Microwave and Millimeter-wave Monolithic Circuits Symposium Digest,* pp. 15-18, 1991.

[60] M. J. Marcus, "Spectrum management implications of millimeter-wave technology," *1994 IEEE International Microwave Symposium Digest,* vol. 2, pp. 631-634, San Diego, CA, May, 1994.

[61] Bharathi Bhat and Shiban K. Koul, *Analysis, Design and Applications of Fin Lines,* ch. 1, "Introduction to millimeter-wave circuits," 1987.

[62] M. Sholley and A. Nichols, "60 and 70 GHz amplifiers," *1986 IEEE MTT-S International Microwave Symposium Digest,* pp. 463-465, Baltimore, MD, June, 1986.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

HARMONIC GENERATION AND MIXING IN HIGH-T_C JOSEPHSON JUNCTIONS WITH TERAHERTZ BANDWIDTH

E. N. GROSSMAN, L. R. VALE, L. R. ZINK, AND K. M. EVENSON,
National Institute of Standards and Technology
325 Broadway
Boulder, CO 80303

Recently, we have developed step-edge YBCO Josephson junctions with characteristic frequencies $f_c = (2e/h)(I_c R_n)$ of up to 5 THz (where I_C and R_n are the junction's critical current and normal-state resistance) [1,2]. Our project's overall aim is to explore the capabilities of these junctions for several types of frequency translation, utilizing the junction's ability to support very high frequency currents. These include high bandwidth mixing at CO_2 laser (30 THz) and visible frequencies, low noise mixing at THz frequencies, and harmonic mixing from mm-wave up to THz frequencies.

The high frequency behavior of a Josephson junction is most directly probed through phase-locking of the internal Josephson currents to an externally applied AC signal and its harmonics (as manifested by Shapiro steps in the current-voltage characteristic). We have measured the power, frequency, and temperature dependence of the Shapiro step heights under THz frequency illumination. Our best device has shown Shapiro steps (phase-locked Josephson currents) up to the 18th harmonic of an applied 404 GHz signal. (See figure 1 below.) The Shapiro step data qualitatively fits the resistively and capacitively shunted junction (RCSJ) model.

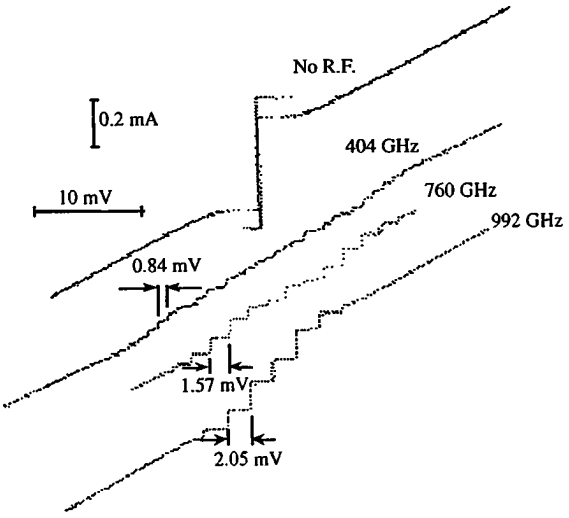


Figure 1. Current-voltage characteristics of a YBCO step-edge Josephson junction with f_C = 4.8 THz, showing Shapiro steps under illumination by far-infrared laser. The junction was coupled to the laser via a broadband, log-periodic, lithographic antenna. Operating temperature was 9 K.

We have also performed direct mixing experiments between two CO_2 lasers with frequency differences from 10 MHz to 12 GHz, and higher order mixing experiments between two CO_2 lasers and an applied microwave signal and its harmonics, with a separation of the CO_2 lasers of up to 27 GHz [3]. The CO_2 mixing data shows a clear

U.S. Government work not protected by U.S. copyright

separation between two different physical mechanisms, hot-electron mixing in the thin-film YBCO of the junction's banks, and Josephson mixing in the junction itself (see figure 2). The hot-electron mixing in the YBCO banks of the junction is directly comparable to the direct modulation experiments performed by the Moscow/Chalmers group [4]. The mixing in the Josephson junction itself is of interest because (a) it may

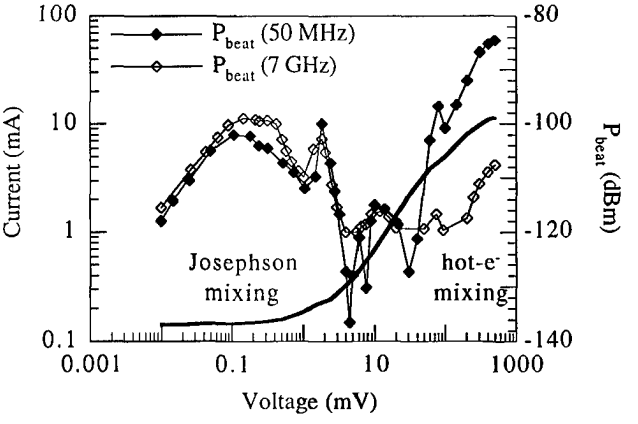


Figure 2. Bias dependence of the power in the beat note between two 30 THz CO_2 lasers, separated by 50 Mz and 7 GHz. The DC current-voltage characteristic is the solid line. The two mixing mechanisms occur in two distinct bias regimes: Josephson mixing in the junction at low biases, with positive curvature in the I-V curve, and hot electron mixing in the banks at high biases, with negative curvature in the I-V curve.

have a much higher bandwidth than hot-electron mixing in the film, and (b) it is compatible with very high order harmonic mixing. The most applicable theory for the mixing in the junction [5] was developed to describe similar high frequency mixing experiments done on point-contact junctions made from low-T_C superconductors. That theory predicts that the bandwidth of such a mixer is limited by the characteristic frequency of the junction. Hot-electron mixing in the film, on the other hand, is limited in bandwidth by the electron-phonon relaxation rate, thought to be of order 100 GHz in YBCO. Likewise the capabilities of Josephson junctions for very high-order harmonic mixing have been demonstrated with low-T_C junctions [6]. Combining the functions of harmonic generation and mixing in a single device, a high-T_C Josephson junction offers the ability to measure multi-THz frequency differences, without the neccesity of dealing with THz frequency signals directly.

[1] P. A. Rosenthal, E. N. Grossman, R. H. Ono, and L. R. Vale, "High Temperature Superconductor-Normal Metal-Superconductor Josephson Junctions with High Characteristic Voltages", Appl. Phys. Lett., **63**, pp. 1984-1986, (1993)

[2] P. A. Rosenthal and E. N. Grossman, "Terahertz Shapiro Steps in High Temperature SNS Josephson Junctions", IEEE Trans. on Microwave Theory and Techniques, **42**, pp.707-714 (1994)

[3] E. N. Grossman, L. R. Vale, D. A. Rudman, K. M. Evenson, and L. R. Zink, "30 THz Mixing Experiments on High Temperature Superconducting Josephson Junctions", IEEE Trans. on Appl. Superconductivity, **5**, pp. 3061-3064, (1995)

[4] M. Lindgren, V. Trifonov, M. Zorin, M. Danerud, D. Winkler, B. S. Karasik, G. N. Gol'tsmann, and E. M. Gershonzon, "Transient resistive photoresponse of $YBa_2Cu_3O_{7-\delta}$ films using low power 0.8 and 10.6 μm laser radiation, Appl. Phys. Lett., **64**, pp. 3036-3038, (1994)

[5] M. Tinkham, M. Octavio, and W. J. Skocpol, "Heating effects in high-frequency metallic Josephson devices : Voltage limit, bolometric mixing, and noise," J. Appl. Phys., **48**, pp. 1311-1320, (1977)

[6] D. G. McDonald, A. S. Risley, J. D. Cupp, K. M. Evenson, and J. R. Ashley, "Four-hundredth order harmonic mixing of microwave and infrared laser radiation using a Josephson junction and a maser," Appl. Phys. Lett., vol. 20, pp. 296-299, April 1972

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A High Purity, High Speed Direct Digital Synthesizer

Gary W. Kent
Rockwell International Corporation
Collins Avionics & Communications Division
855 35th. St. NE, 137-153
Cedar Rapids, IA 52498
(319) 395-5346
FAX: 319-395-5248

Neng-Haung Sheng
Rockwell International Corporation
Science Center
1049 Camino Dos Rios, 083-A16
Thousand Oaks, CA 91360
(805) 373-4110
FAX: 805-373-4775

Abstract

A Direct Digital Synthesizer (DDS), that clocks at 500 MHz, has been constructed in a 70 pin hybrid circuit package that is 1.14 x 2.33 x 0.2 inches. Spectral purity is better than -55 dBc worst case spur, up to 245 MHz output frequency. TTL compatible, parallel lines are provided for 28 bits of frequency and 8 bits of phase control. The hybrid is based on two GaAs chips, a HBT digital to analog converter and a MESFET accumulator/ROM combination. Two silicon ECL chips are used for clock amplification and distribution. This allows an AC coupled sinusoidal input clock of 0 dBm nominal amplitude. Power dissipation is typically 5 watts using +5.0, -5.2, and -2.2 volt supplies.

This paper describes the DDS architecture, design of the GaAs chips, and special problems encountered during development of the hybrid.

Introduction

Specifications for new communications equipment increasingly require high speed Direct Digital Synthesizers (DDS). At the highest level, a DDS consists of three parts: a phase accumulator, a sine function read-only memory (ROM) look-up table, and a digital-to-analog converter (DAC), as shown in fig. 1. Since frequency is the rate of phase change, the command word which is added to the accumulator each clock period determines the output frequency. Because the fixed-length accumulator regularly overflows, the phase is periodic, with a frequency proportional to both the increment rate and the incremental amount. The largest value that can be stored in the accumulator represents 2π radians, so that overflow corresponds to a module 2π operation, and is not seen in the output. The accumulator addresses the ROM table to determine a digitized sine amplitude value. The DAC is then employed to convert the digital data to an analog sine-wave. For output frequencies close to one-half the clock frequency, a lowpass filter may be used to eliminate the alias frequency component created by the sampling process [1][5].

A DDS provides many significant advantages over phase-locked-loop (PLL) approaches. Fast settling time, sub-Hertz frequency resolution, continuous-phase frequency switching and low phase noise are features easily obtainable in DDS systems. However, this approach has

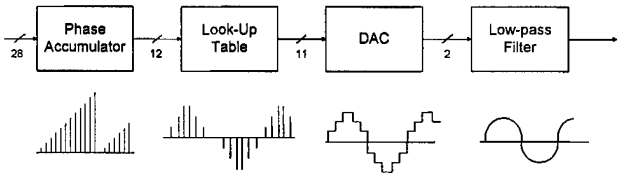


Figure 1. Direct Digital Synthesis

been limited to low frequency applications, since the component circuits, such as the ROM and DAC, have to support clock rates of about two times the synthesized frequency. Semiconductor technologies previously would not support the high clock rates required. Also, high spurious output made the concept unusable for many applications. High speed DDS technology has been dramatically improving. There are now several companies that offer products that clock at speeds in excess of 500 MHz and provide reasonable spurious performance. However, when available chips are reviewed for new radio applications, they have one or more of the following deficiencies:

1) Spurious is still too high.
2) Size is too large.
3) Power dissipation is too high.

HBT Technology

High bandwidth DDS systems can only be realized by the use of ultra-high speed transistors, such as Hetrojunction

0-7803-2500-1/95 $4.00 © 1995 IEEE

Bipolar Transistors (HBTs). The AlGaAs/GaAs HBT technology has been developing at Rockwell for the past several years. A baseline process has recently been demonstrated for manufacturing LSI circuits of up to 8000 transistors at Rockwell's III-V facility. We use MOCVD-grown 4 inch GaAs substrates, with a Carbon-doped p+ base. Minimum transistor geometry is 1.4 x 2.0 μm^2. The base is very thin (700 angstroms) and doped with carbon to reduce base resistance. D.C. current gain for these transistors is measured to be in excess of 100. The f_T and f_{max} is about 60 GHz. Three years ago, Rockwell recognized that these HBTs would be advantageous in high speed DAC design and a research program was started to develop DAC prototypes that would clock at speeds in excess of 1 GHz.

DAC Architecture

The 12 bit digital-to-analog converter is a current steering type with complimentary outputs. Output levels are 500 mV into a 50 ohm load. The 4 most significant data bits are converted from binary codes into 15 thermometer codes. The remaining eight bits are sent to current sources that drive a R-2R ladder. Currents through all 23 sources are identical. This is called a segmented architecture (figure 2). It was chosen over an "all ladder" approach to minimize glitch energy at the code where the MSB switches from zero to one, and all the other bits switch from one to zero. This is traditionally the most difficult switching point to realize monotonic behavior. Data registers are used for time alignment at the DAC input and at the input to the current switches. This results in a latency of two clocks periods.

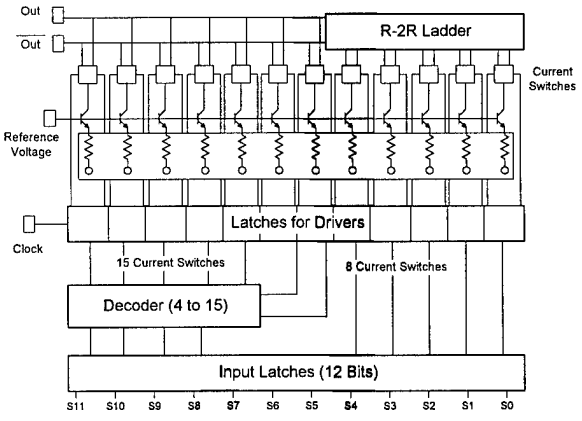


Figure 2. 12-bit, 4-bit segmented, 8-bit R-2R, current steering DAC.

The DAC circuit has been fabricated using the Rockwell HBT process described above. The resulting chip measures 88 x 88 mils. Power dissipation is 1.2 watts.

NCO Architecture

The phase accumulator and ROM sine wave mapping circuits are combined into a single MESFET chip called a numerical controlled oscillator (NCO). The 28-bit frequency select word is strobed into the input register and synchronized with the main clock, so that frequency changes are phase continuous. The accumulator uses seven four-bit adders in a pipeline configuration (figure 3). The output is truncated to the upper 12 bits. Phase modulation is provided by two four bit adders that sum the uppermost eight bits of the accumulator with a phase control word before they address the ROM.

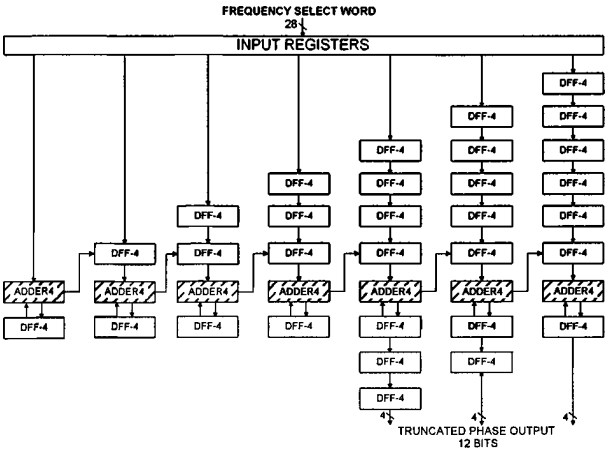


Figure 3. Fully pipelined Phase Accumulator.

It is impractical to map all 28 bits from the accumulator to a sine wave amplitude value. Reducing the number of ROM input bits introduces phase truncation error, and reducing the number of ROM output bits introduces amplitude truncation error. Both phase and amplitude truncation errors will generate unwanted spurs. Simulations show that 12 bits of phase and 11 bits of amplitude should result in -68 dBc worst case spurs.

Still, a look-up table with 12 bits of phase and 11 bits of amplitude resolution would correspond to a ROM size of 45K bits. To ensure adequate yield in fabrication and minimal power consumption, a compact architecture that minimizes storage requirements has been presented by Nicholas and Samueli [2]. The ROM used in the Rockwell DDS is based on a similar concept, although the architecture was designed in 1985 using a less sophisticated method. Great storage compression is

achieved by partitioning the look-up table into three small ROMs. Phase arguments into the ROMs are segmented as a low phase resolution coarse ROM, with additional phase resolution provided by interpolation between coarse ROM samples using the slope of the sine wave at each coarse sample. The third ROM provides correction factors when needed. This approach reduces storage to about 272 bits, achieving a memory compression ratio of 1:165. The tradeoff is extra multipliers, adders and logic circuits required to decode the complete sine function [3].

The most elementary method of sine storage compression is to exploit the symmetry of the sine function about π and $\pi/2$. By properly inverting the phase and amplitude of the sine function, look-up table samples need only be stored for phase values between 0 and $\pi/2$ (figure 4).

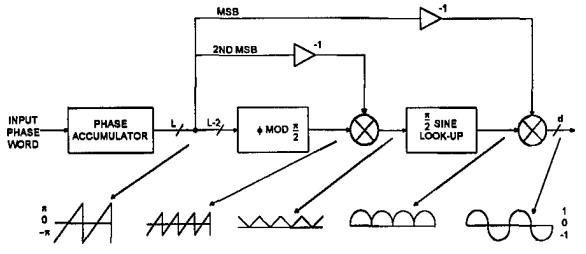


Figure 4. NCO architecture, with logic to exploit quarter-wave symmetry.

The complete sine wave can be constructed using hardware to truncate the phase MSB, and then using the second MSB to full wave rectify the magnitude of the phase. The remaining bits are used to address the $\pi/2$ look-up table that is partitioned into a coarse ROM, a slope ROM and a carry ROM. Several computer programs were used to optimize the partitioning for optimum ROM size and spur performance. Based on the simulation results, the quarter sine is divided into 16 segments, and a coarse ROM value of ten amplitude bits determined for the beginning of each segment. The remaining 64 values in each segment are calculated by multiplying the relative position in the segment by a segment slope, and adding it to the coarse value. The slope values are 5 bits. The amplitude error resulting from this approach can be as much as 2 LSB's, so in portions of some segments an extra bit is added in to correct the values to within 1 LSB. The information for this correction is stored in a 16x2 carry ROM that is combined with the slope ROM (figure 5).

Now that the ROMs are so small, (16x10 and 16x7) it is convenient to realize them using linear addressing and hard logic encoding instead of modern ROM architecture. The four address bits are decoded, and then multi-input OR gates determine an output value for a particular address (figure 6). A simple mapping procedure is used to determine the logic necessary. If the circuitry is too slow, it is readily pipelined to increase the speed as necessary. The requirements of this design result in a pipeline of 22 stages for the entire NCO.

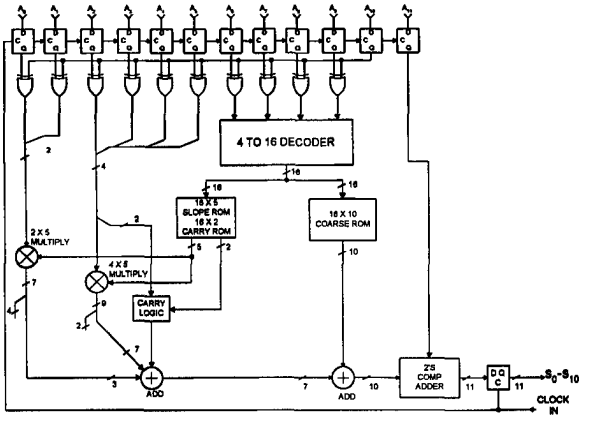


Figure 5. Sine-Converter logic.

The NCO circuit has been fabricated by Vitesse Semiconductor using their 0.6 micron H-GaAsIII MESFET process. The design uses the VGFX20K gate array that has a size of 218 x 135 mils. Power dissipation is 3 watts.

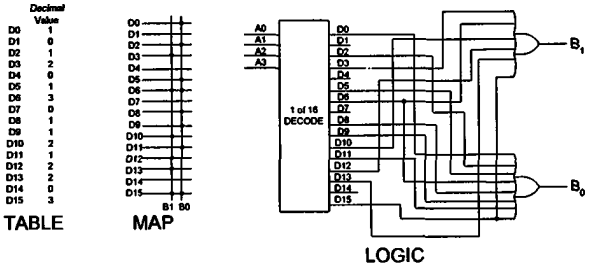


Figure 6. Carry ROM, with mapping procedure.

Hybrid Packaging

To combine the two chips into a complete DDS, a 70 pin hybrid, from National Hybrid Inc., has been selected. The package dimensions are 2.33 x 1.14 x 0.2 inches. It has a screw mount molybdenum flange for excellent heat sinking. The DAC chip is attached directly to the flange through a hole in the ceramic substrate. The NCO is attached to a beryllium oxide (BeO) insulator because the GaAs substrate must be connected to the Vee (-2.2V) supply and isolated from ground. Thermal analysis shows that the temperature rise from the hybrid flange to the

DAC junction is only 14 degrees Celsius. This is important because many DAC linearity effects are temperature sensitive.

Parallel lines provide the 28 frequency and 8 phase lines. A frequency or phase strobe is used to load the data after it has settled. Some DDS circuits load serially, or maybe several bits at a time, to minimize the number of I/O lines. The all-parallel approach allows high speed frequency-or-phase modulation.

Two power planes are used on the hybrid substrate to provide -5.2V and -2.2V to the chips. Even so, there is over 0.1V drop from the pins to the -2.0V Vee pad on the NCO chip, due to the 1.5A current needed. Care must be taken to prevent NCO switching currents from coupling onto the DAC clock and creating spurious output frequencies.

High Speed Digital Interface

The interface of a NCO and DAC is not trivial at clock speeds approaching 1 GHz. The set-up and hold times of the Rockwell DAC are approximately 180 picoseconds. Even though the hybrid circuit is small, line length differences can have a big impact on the DAC's ability to latch the correct NCO data. Ideally, the DDS clock should be fed to the NCO first and then passed through to the DAC with the proper clock to data timing relationships determined by NCO gate delays. Then, as process variations change the propagation delays of the NCO circuits, both the output data and output clock timing will be affected in the same manner. Their relationship to each other will remain somewhat constant. With proper design of the timing relationship, the DAC will clock correct data with the extremes of fast and slow NCO chips.

This approach was planned for the Rockwell DDS, but excessive jitter prevented the use of the NCO clock output as the DAC clock input. Since the DAC clock controls the register that drives the output current switches, it controls the digital-to-analog conversion process and must be considered an analog signal. It's purity has a direct effect on the output spurious. The NCO clock output circuit was automatically routed during the layout of the gate array. Little attention was paid to possible interaction with the other digital signals, especially the high current outputs. As a consequence, the clock edges are pulled due to parasitic coupling with the output data lines. This has no effect on proper digital operation of the NCO, but corruption of the clock output makes it unsuitable for driving the DAC.

To keep the DAC clock clean, the DDS system clock path is split by an ECL clock driver chip. One path goes directly to the NCO chip. The other must include a delay before reaching the DAC. The amount of delay is approximately equal to the clock-to-data delay of the NCO. Unfortunately, the delay of the NCO will vary from device to device, so the delay in the DAC clock path must be adjustable, to provide the proper timing relationships at the input of the DAC. The DDS uses an ECL programmable delay chip to provide the flexibility needed to prevent violations of the DAC set-up and hold times (figure 7).

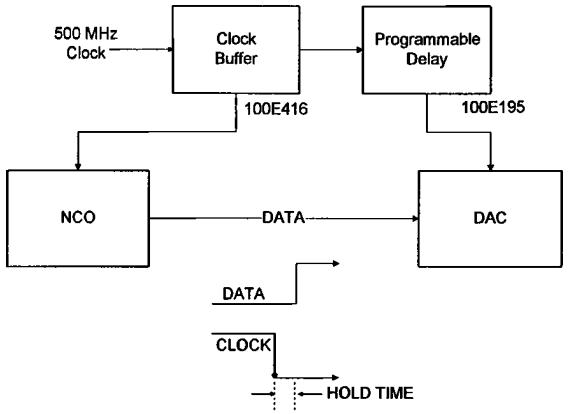


Figure 7. Hybrid DDS clock path. Programmable delay adjusts timing so that hold time is not violated.

Performance Results

Spurious performance and bandwidth are the most critical parameters of any DDS, since other features such as phase noise, frequency switching speed and modulation accuracy are almost ideal. The spurs are usually put in the best light possible by presenting spectrum results of only a few output frequencies, and letting the observer extrapolate to predict spurs at other frequencies/bit-patterns. A commercial manufacturer of synthesizers has presented a more definitive automated spur characterization technique [4]. The one spur with the highest amplitude and the spur with the second-highest amplitude are plotted on the y-axis and the output frequency on the x-axis. This permits evaluation of the unit's performance at multiple outputs up to the maximum bandwidth. A lowpass filter follows the DDS to attenuate out-of-band spurs. When the Rockwell Hybrid DDS is measured in this manner, the worst spur is -55 dBc (figure 8).

Looking at the two worst spurs is certainly a meaningful characterization technique. But some applications are interested in the spectral power of the spurs, where many

small spurs can add up to be significant. Figure 9 shows the Total Integrated Spurious of the Rockwell Hybrid DDS. Here the output frequency was plotted from 100 to 245 MHz, and all of the spurs, within a bandwidth of 50 to 250 MHz, were added up. Each dot on the graph represents the total sum for one output frequency. The following formula calculates the value for each dot, in dBc:

$$\text{Total Integrated Spurious} = 10\log(10^{x_1/10} + 10^{x_2/10} + \ldots 10^{x_n/10})$$

$$x^n = \text{spur in dBc}$$

When the Rockwell Hybrid DDS is measured in this manner, the worst total sum is -50 dBc.

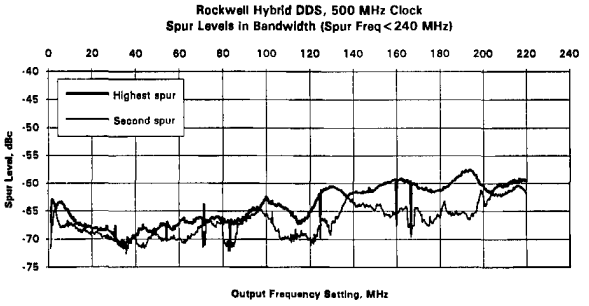


Figure 8. Worst Case Spurious

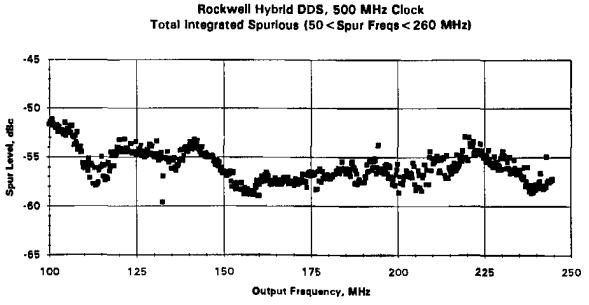


Figure 9. Total Integrated Spurious

Conclusion

A high speed DDS has been described that has significant improvements in spurious performance, size, and power dissipation. The NCO look-up table is compressed sufficiently to use logic gates instead of an actual ROM. The path of the DAC clock requires special attention to prevent timing jitter and the resulting degradation of spurious performance. Also presented is the concept of Total Integrated Spurious performance as a valuable measure of system performance.

AlGaAs/GaAs HBT technology has advanced to the point where a "next generation" DDS will be realized as a single large chip. Clock speeds in excess of 2 GHz should be possible.

Acknowledgments

The authors would like to thank Bob Marston, Glenn Doughty, and the rest of Rockwell Communications Systems Division management for their support of DDS research. We also acknowledge the contributions of the Rockwell Microelectronics Technology Center for fabricating and enhancing the performance of the HBT DAC. Special thanks goes to Scott Voyek for technical support of the test hardware.

References

[1] J. Tierney, C.M. Rader, B. Gold, "A Digital Frequency Synthesizer", IEEE Trans. Audio Electroacoustic., vol. AU-19, pp. 48-57, 1971.

[2] H.T. Nicholas, H. Samueli, "A 150-MHz Direct Digital Frequency Synthesizer in 1.25 μm CMOS with -90 dBc Spurious Performance", IEEE Journal of Solid-State Circuits, vol. 26, pp. 1959-1969, Dec. 1991.

[3] R.A. Freeman, "Digital Sine Conversion Circuit For Use In Direct Digital Synthesizers", U.S. Patent # 4,809,205, Feb. 28, 1989.

[4] "Frequency Synthesis & RF Subsystems Catalog", SCITEQ Electronics, Inc., San Diego, CA.

[5] D.A. Sunderland, R.A. Strauch, S.S. Wharfield, H.T. Peterson, and C.R. Cole, "CMOS/SOS Frequency Synthesizer LSI Circuit for Spread Spectrum Communications" IEEE Journal of Solid State Circuits, vol. SC-19, pp. 497-505, Aug. 1984.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Digital Frequency-Lock Loop Using High-Resolution TAC and Advanced Control Algorithm for Trapped-Ion Frequency Standards

D.A. Stowers, R.L. Sydnor, R.L. Hamell, M.J. Grimm & G.J. Dick

California Institute of Technology
Jet Propulsion Laboratory
Pasadena, California

Abstract

A local oscillator for the Linear Ion Trap Standard (LITS) has been designed and tested. It uses a Time-to-Analogue (TAC) converter to achieve the required resolution and low noise. The TAC has an instantaneous resolution of 23 bits and an accumulative resolution of approximately 34 bits. The noise analysis, circuit details and block diagram are given, as well as the performance of the loop.

Introduction

In order to use a VCXO as the local oscillator for the LITS very stringent requirements must be met. The *LITS* has a cycle time between oscillator updates between 5 and 16 seconds. This allows the filling of the trap, pumping of the ions to the upper state, Ramsey interrogation, and calculation of the required control for the VCXO. The diagram of the system is shown in Figure 1. One desires that the smallest increment of frequency produced by the VCXO be much less than the expected frequency stability of the trap. For the JPL LITS, this leads to a smallest frequency increment of 1×10^{-17}. Similarly, the largest required increment depends on the acquisition and warm-up characteristics of the VCXO and the loop. For this system, this amounts to a fractional frequency increment of 1×10^{-10}. The actual implementation uses a pulse which varies in length from 1 microsecond to 5 seconds. This gives a total of 1,000,001 steps, or 23.25 bits of control. Since the unit incorporates an integrator, the maximum control voltage reached is achieved by integration of a number of these bits, up the maximum output of the VCXO (and the maximum control voltage of the VCXO) of 10 volts.

System Block Diagram

Figure 2 is a block diagram of the TAC implementation in the LITS. The LITS controller drives a digital pulse generator that puts out a TTL word that sets the length of the pulses to the TAC. The TAC output then steers the VCXO to the proper frequency. A variable width "enabling" pulse determines the integration time and a "sign" pulse sets the polarity of the steering voltage to the VCXO.

The logic gates translate the input TTL pulse commands into appropriate drive voltages to feed a pair of high speed operational amplifiers. These amplifiers bias the current-steering diodes in order to gate a highly stable steering current of the required polarity into the integrator summing junction. The series-connected Schottky diode current switches and shunt diode voltage limiters provide the necessary high switching speed and precise control of current to the summing junction.

This steering current in established by a 2 megohm resistor which is fed from a band-gap referenced voltage regulator, producing a 7.5 μa current into the summing junction of the integrator/operational amplifier. The integrator consists of a low-leakage

0-7803-2500-1/95 $4.00 © 1995 IEEE

500 microfarad polycarbonate capacitor as feed-back around a low-noise OP-27E operational amplifier and current driver IC.

For the time duration that the pulse is activated the integrator output will ramp at the I/C volts per second, or 15 millivolts/second for a 7.5 μa charging current and a 500 μf capacitor. The ramp polarity is set by the sign pulse from the pulse generator.

The integrator performance is dependent on the OP-27 input characteristics which include an input current as high as ± 40 na, and an offset voltage-temperature coefficient as high as 0.6 μV/°C. The static input current is nulled out by adding an adjustable compensating current into the summing junction. Temperature induced instability is reduced to a negligible value by using a thermoelectric control system that holds the temperature stable to within ±0.01°C.

In addition to the above instabilities, the TAC output exhibits a random drift rate of as much as 5 millivolts per 24 hour period. For a VCXO frequency sensitivity of 4×10^{-9} df/f per volt and a 10 second sampling interval, an error of 2.3×10^{-15} can accumulate between frequency corrections. At 40.5 GHz where frequency comparison takes place, this will move the detected frequency less than 0.1 millihertz from the fringe center.

Operational Amplifier Selection

To insure that the frequency noise at the output of the VCXO which is caused by the driving network (integrator circuit) is less than the intrinsic noise of the VCXO, and analysis of the noise contribution due to the integrator operational amplifier was performed. Because of the high impedance at the summing junction, it was assumed at first that a FET input operational amplifier (LMC6001AIH) would be the best choice. By analyzing the circuit with the equivalent noise currents and voltages inherent to the operational amplifier it was determined that, while the current noise contribution was negligible, the voltage noise contribution was much too high. The next choice was an a low-noise bipolar amplifier (OP-27E). The same procedure was used and, by optimizing the impedance levels, the resultant frequency noise was much less than that of the VCXO itself. The final circuit utilized this operational amplifier.

Performance Results

Figure 3 shows the performance of the VCXO, the basic stability of the trapped ion, and the combined performance of the VCXO and trapped ion. This demonstrates that the system performs as expected. The description of the algorithm which was used in the controller is given in the paper "Experimental Verification of Local Oscillator Induced Performance and Limitation in Pulse-Mode Atomic Frequency Standards" in these Proceedings.

Acknowledgment

This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration.

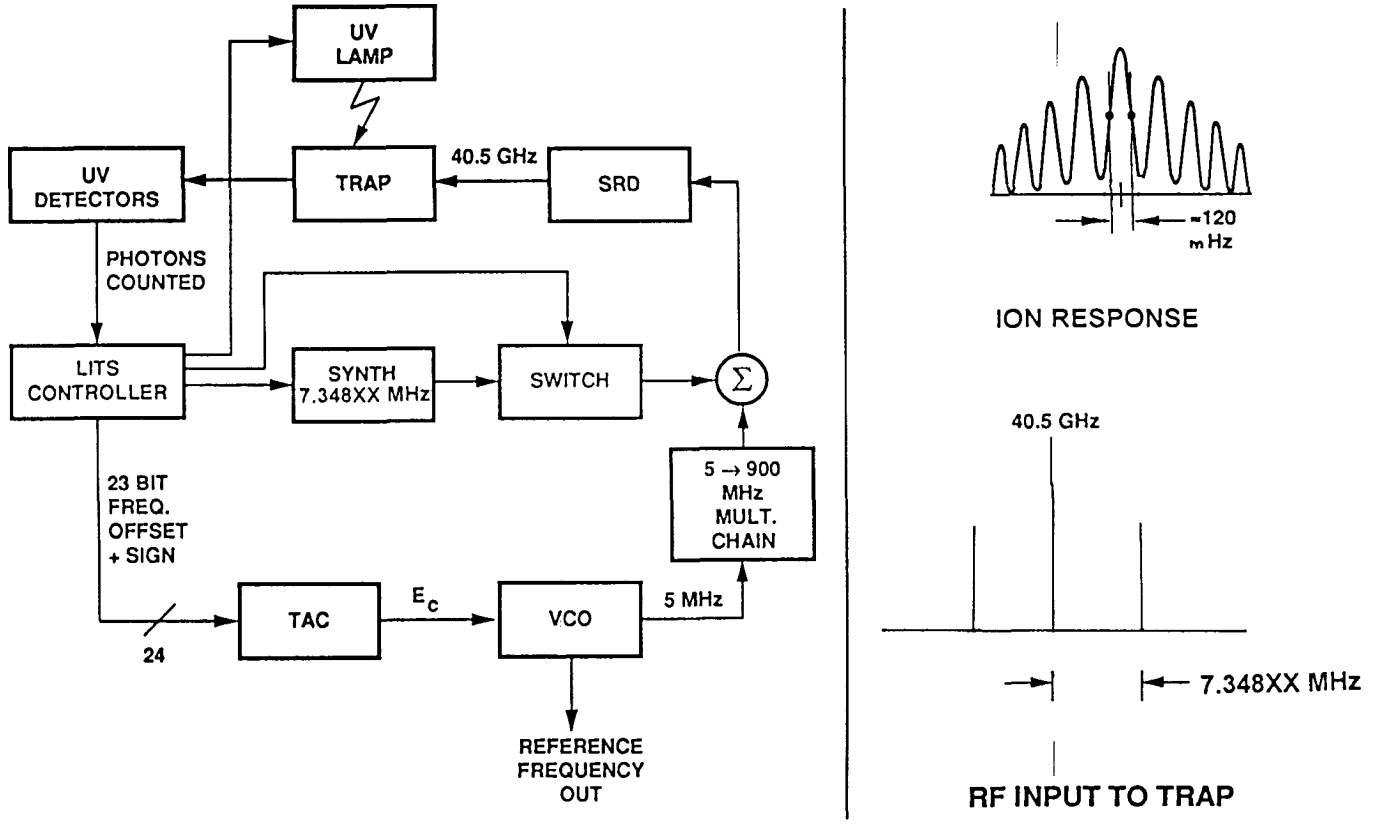


Figure 1: LITS Frequency Locked Loop

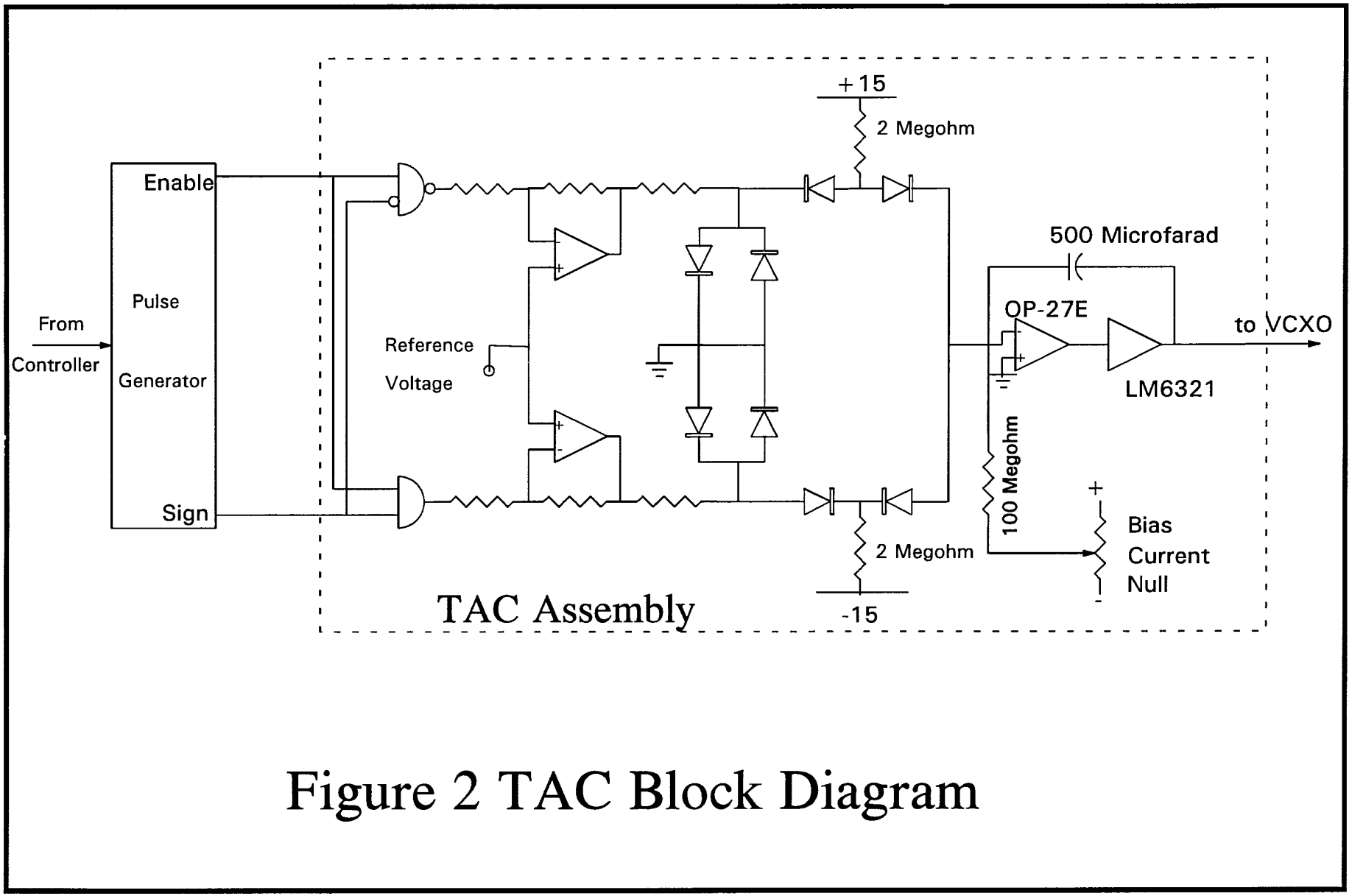


Figure 2 TAC Block Diagram

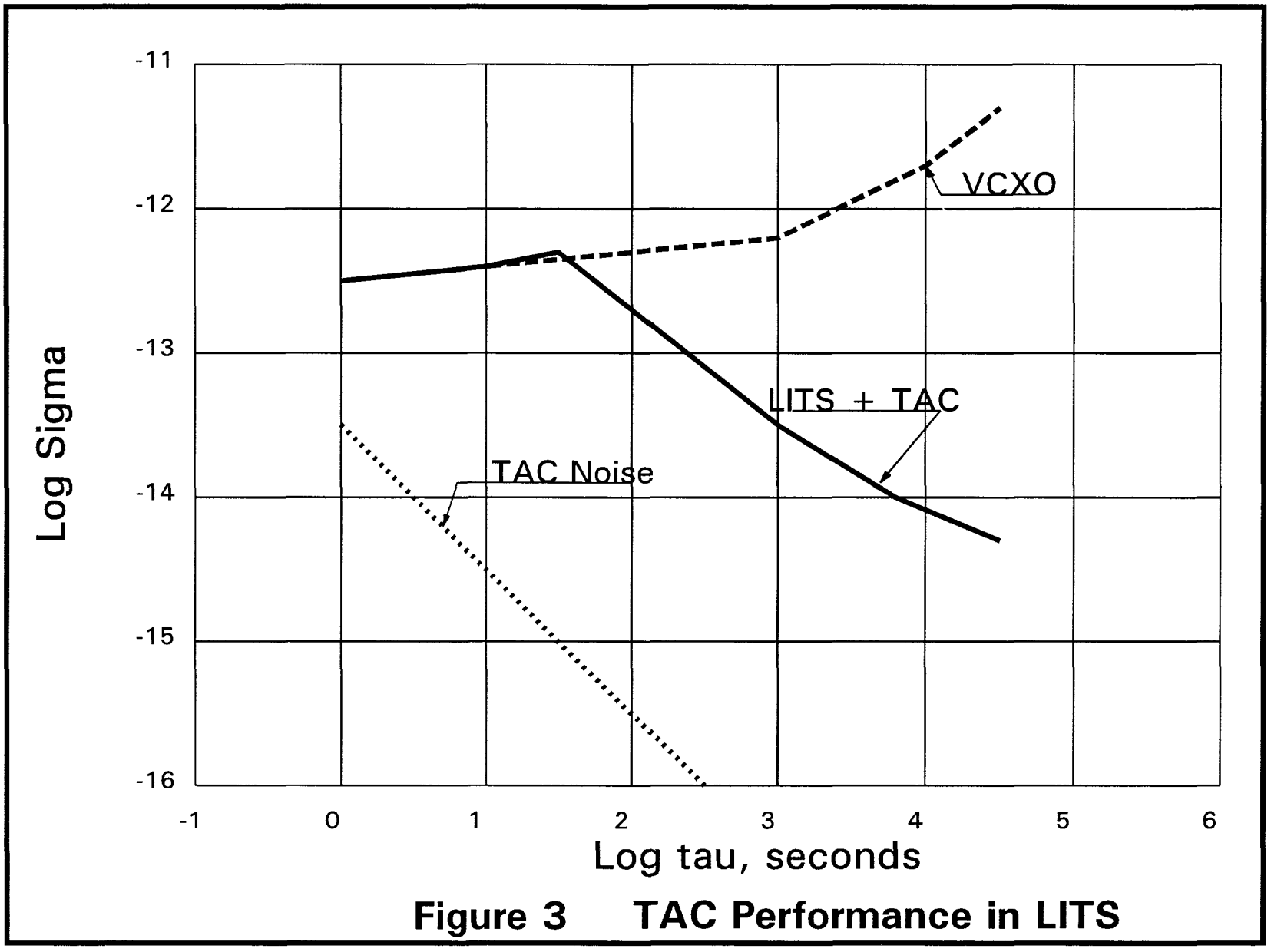


Figure 3 TAC Performance in LITS

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A NARROW BAND HIGH-RESOLUTION SYNTHESIZER USING A DIRECT DIGITAL SYNTHESIZER FOLLOWED BY REPEATED DIVIDING AND MIXING

Richard Karlquist, Hewlett-Packard Company, Santa Clara Division, 5301 Stevens Creek Blvd., MS 52U/7, Santa Clara, CA 95052 email: rkarlqu@scd.hp.com

Abstract

A synthesizer is described that generates a narrow band of frequencies around 10 Mhz with extremely small step size and high spectral purity. A DDS running at 667 kHz is upconverted to 10.667 Mhz and divided back to 667 kHz several times to improve the resolution and spurious suppression of the DDS. The final 10.667 Mhz offset frequency is then mixed with 667 kHz resulting in a 10 Mhz output. The post mixer filtering is done with inexpensive 10.7 Mhz ceramic filters of the type used in FM broadcast receivers.

Applications

Metrology

It is frequently necessary to add a small, precisely controllable offset to a standard frequency signal. In some cases, the purpose is to "steer" a free running clock to agree with a known time reference such as UTC without losing the information about the error between the clock's own time and the reference. In other cases, the purpose is to get a beat note between the clock being tested and a reference clock, in order to apply resolution enhancement techniques to the frequency difference measurement [1]. A third related use would be as the "vernier" section of a synthesizer used in an atomic frequency standard.

In these applications, it is usually unsatisfactory to directly steer the clock in question. It may not be steerable at all, or may use a suboptimal method such as variations in the C field of an atomic standard. Even if none of these problems is present, it is desirable to have both a corrected and uncorrected version of the clock available simultaneously so that the uncorrected version can be compared to UTC. Hence what is called for is an external two-port offset frequency synthesizer.

Communications

A non-metrology use of an offset synthesizer would be to facilitate a technique known in the paging industry as "simulcasting" where a paging receiver may receive pages from multiple base station transmitters at the same time. If two stations are in "zero beat", destructive interference can result. To get a controlled beat note, a small offset is added to the base station's reference frequency. Since this reference is frequency multiplied to 800 MHz or more, the offset synthesizer needs to have very good spectral purity. Another communications use might occur when building modulators. With some modifications to the architecture, it is possible to build frequency stable analog modulators. In many cases, analog modulation is still simpler and more accurate than digital modulation if there is a solution to the usual carrier frequency drift problem.

Prior art

A popular method of generating frequency offsets with moderate performance is the phase microstepper. It makes use of the fact that a small frequency offset is equivalent to a slow phase ramp. The microstepper has a controller which increments or decrements a digital phase shifter at appropriate intervals. When the phase reaches 360 degrees, it is reset to 0 degrees, which is transparent if and only if done perfectly. Phase microsteppers are subject to phase jumps, phase drift and spectral impurities. To improve on this performance, sophisticated multiloop phase locked VCXO synthesizers have been built [2]. While the performance of these systems is excellent, it depends on the use of high quality VCXO's. Also, this architecture doesn't lend itself to tunable offsets because of tuning range limitations of the VCXO's. In some cases, general purpose synthesized signal generators have been pressed into service for frequency offset work, but even the best of these is only marginally useful in metrology work and in any case not very cost effective for this type of task. The most appropriate general purpose synthesizers for this

0-7803-2500-1/95 $4.00 © 1995 IEEE

application are ones fashioned after the repeated mix and divide architecture popularized by the Hewlett-Packard 5100A designed in the 1960's [3]. Even these architectures tend to have a problem with temperature coefficient of phase. They are also very complex and expensive to implement.

A synthesizer will be described below that is essentially a simplified 5100 type architecture with just enough functionality to implement small frequency offsets of a 10 Mhz. reference. Later, other reference frequencies will be discussed. The important issue of minimizing temperature induced phase drift will be covered in some detail.

Requirements

Frequency tuning range

A tuning range of +/- 1 Hz is quite adequate for many metrology applications and will work for simulcasting. For generating beat notes to measure frequency differences, up to +/- 1 kHz may be needed. In modulator work, a range up to +/- 5 kHz is sufficient for most FM voice and data channels. As shown below, there is a tradeoff: more resolution and spectral purity come at the price of reduced tuning range.

Spectral purity

In metrology applications, it is customary to require non harmonically related spurious sidebands to be better than -80 dBc (minimum requirement), with -100 to -120 dBc being desirable. When observed with phase noise test instrumentation, spurs may be visible down to -140 dBc or even less. In communications applications, a good rule of thumb is to keep all spurs below -40 dBc *after multiplication.* In many cases, regulations require -60 to -80 dBc performance. With multiplication factors in the range of 100 to 1000, the original spurs are enhanced 40 to 60 dB. Hence the synthesizer needs to hold spurious sidebands to -80 dBc to -140 dBc as measured at the 10 MHz reference.

Resolution (step size)

For metrology, resolutions of at least 1 part in 10^{15} are desirable, which means steps measured in nanoHertz at 10 MHz. Less critical applications may only need to have steps in the milliHertz range. The synthesizer to be described utilizes a direct digital synthesizer (DDS) at the front end of the synthesizer to generate steps, so step size is to some extent a non issue. By extending the DDS accumulator size (a relatively easy thing to do) it is possible to achieve any desired step size. While the divide and mix sections to be described later do add 4 bits of resolution apiece, this is not their major function, a least in the case where a DDS is used. In modulator applications where the DDS is replaced by an analog frequency/phase modulator, this 4 bits of resolution is more properly stated as 24 dB of additional dynamic range, which is a non trivial improvement in the analog domain.

Phase stability vs temperature

By definition, frequency is the time rate of change of phase. If phase is temperature sensitive, and the ambient temperature changes with time, there will effectively be a frequency error [4]. The required phase stability depends on the assumed rate of change of temperature and the frequency error tolerance, but a good rule of thumb is to keep the overall phase drift expressed as a change in delay time to under 1 nanosecond over the full expected temperature range, such as 0 to 70 degrees C.

The classical mix and divide architecture

We begin with a brief review of the HP 5100 type of synthesizer, which represents the mix and divide family of architectures. Figure 1 shows the 5100 block diagram A synthesizer driver system on a separate chassis (not shown) generates reference frequencies at 24 Mhz and 3.0, 3.1, 3.2, ..., 3.8, 3.9 MHz. from the 5 MHz reference frequency input. The 24 MHz signal goes to all modules. A crossbar switch routes the 3 to 3.9 MHz signals to the appropriate divide and mix modules according to the frequency selected on the front panel. Each module is connected to a digit switch on the front panel that has a possible setting of 0 through 9. If we call this setting "N", then the switch will feed 3.N MHz to the corresponding module. In a two-stage upconversion process, each module converts an incoming signal in the 3 to 3.1 MHz band to a signal in the 30 to 31 MHz band, then divides the resulting frequency by 10 to get an output that again falls in the 3 to 3.1 MHz band. Although the input and output frequencies fall in the same band, they are different because a decade of resolution has effectively been added. For example, if the input frequency to a module is 3.89 MHz, the output frequency will be 3.N89 MHz, where N is the setting of the digit for that module (if N=5, then the frequency

would be 3.589 MHz). The modules are cascaded so that any amount of resolution can be realized by using a sufficient number of modules. In the 5100, the last module omits the divide-by-10 to get a 30 to 31 MHz output band. The first module's input that would normally come from the previous module is either hardwired to 3 MHz or can be connected to a "search oscillator" which tunes continuously from 3 to 3.1 MHz. This allows continous resolution with good stability because any frequency drift in this oscillator is reduced by a factor of 10^M, where M is the number of mix and divide modules. Each module improves the spectral purity of the signal it receives from the previous module by 20 dB due to its divide-by-10 frequency divider.

The 5100 architecture as an offset synthesizer

Suppose a tuning range of only 1 Hz were needed for a typical metrology application. Then a search oscillator tunable over 100 kHz could be followed by 5 mix and divide modules, resulting in the required coverage. This would give an output in the 3 to 3.1 MHz band, assuming the final module contained a divide by 10 block. In this example, the 5 modules could be hardwired to a particular digit setting eliminating the crossbar switch. If, furthermore, the output frequency range could be chosen as 3.000000 to 3.000001 MHz, then all modules could be hardwired to 3 MHz thereby eliminating the 3.1 through 3.9 MHz generators. Having done this, the effect would be that each module's second mixer would always add 3 MHz to the output frequency of the first mixer, which itself always adds 24 MHz to the module input frequency. Hence the two mixers could be replaced by a single mixer that adds 27 MHz to the input frequency (i.e., an LO frequency of 27 MHz.) The result is the reduction of the 5100 architecture to little more than a fixed 27 MHz synthesizer and the decade modules, and the modules are simplified to have only one stage of mixing and filtering instead of two. Finally, the search oscillator can be replaced by its modern equivalent, the direct digital synthesizer, to make the system a true synthesizer. The DDS can have a tuning range of 3 to 4 MHz in the modified architecture as compared to 3 to 3.1 MHz for the 5100 search oscillator. The advantage of this arrangement compared to using the DDS output directly is that each module reduces DDS spurs by 20 dB when the signal is frequency divided and adds 4 bits of resolution to the DDS. The resulting configuration is shown in Figure 2.

Filtering

The bandpass filter after each mixer is essential to getting high spectral purity. In the original 5100 and in the initial stages of the simplified architecture of Figure 2, the bandpass filter must pass 30 to 31 MHz, a fractional bandwidth of 3%. In the later stages of the simplified architecture, the passband can be much narrower, if one is willing to have non-identical modules. In all cases, the filter must be capable of rejecting the 24 MHz image and the 3 and 27 MHz input frequencies. The 9 to 1 spread of the mixer input frequencies results in a situation where the only mixer products that fall within the 3% passband are high order ones that decrease rapidly as the mixer drive level is backed off. It is very important that the out-of-band mixer products not be allowed to reach the divider. If they do, the non-linear nature of the divider will alias them back into the passband and compromise spectral purity. It is also essential to low pass filter the output of the divide-by-10 frequency divider to eliminate harmonics that would fall in the output band. Hence the main factor in maintaining spectral purity is proper filtering.

The tradeoffs in filtering that apply here are bandwidth limitations, temperature drift, microphonics, phase noise, and cost. With the 3% bandwidth required in the 5100, crystal filters are ruled out because they are incapable of achieving that much bandwidth. On the other hand, 3% is near the minimum limit of feasibility for practical LC filters and there isn't much prospect for improvements in this technology. Besides the manufacturing problems of producing such narrow filters, LC filters have a substantial problem with temperature stability. This could cause a temperature dependent phase shift which would be most damaging in the last stage. One possible approach to this problem would be to use a crystal filter for the last filter only, since the required passband at that stage would be narrow enough to implement with a crystal filter. This would not necessarily alleviate temperature stability problems because of the long delay time that goes hand in hand with the narrow bandwidth of a crystal filter. A long delay time represents many radians of phase shift and hence a small change in center frequency would cause a large phase shift. Other problems with crystal filters are phase noise and possible microphonics.

The architecture of figure 2 has reduced susceptibility to phase drift compared to figure 1

because the drift of the 3.X MHz generators is not present.

Ceramic filters

What is needed in the figure 2 architecture is a moderate bandwidth filter with good stability. A so-called "ceramic" piezoelectric filter is a good match for this requirement. It is optimal for bandwidths too narrow for LC filters and too wide for crystal filters, which describes the requirements at hand. Ceramic filters are more stable than LC filters but are wide enough to avoid the close in phase noise and microphonic problems of crystal filters. Additionally, ceramic filters can be much cheaper than either of the alternatives. They are used worldwide in FM broadcast receivers for IF filtering and cost a fraction of a dollar in production quantities. These filters are generally available only with a center frequency of 10.7 MHz, and with various bandwidths from less than 100 kHz to 500 kHz. Although they have been in use for decades, their performance has improved in recent years. When receivers were changed from analog tuning to synthesizer tuning, it became necessary to improve the tolerance and stability of the center frequency. This is because it was no longer possible to compensate for an error in the IF filter center frequency by merely offsetting the analog tuning slightly. It is now easy to buy filters that are centered very precisely at 10.7 MHz and have low temperature drift. Recently, larger bandwidths such as 400 or 500 kHz have been put into production for use in DBS receivers where the deviation is wider than with FM broadcast.

10.7 MHz architecture

The key to using 10.7 MHz filters is to determine a technique for scaling the simplified 5100 type architecture such that the filter frequency becomes 10.7 MHz. If all the frequencies in the architecture in figure 2 are merely scaled by a factor of 10.7/30, the filters would of course change to 10.7 MHz but the 27 MHz reference frequency would scale to 9.63 MHz, which is quite inconvenient to synthesize from 10 MHz (figure 3a.) If instead the frequencies are scaled by a factor of 10/27, then the 10 MHz may be used directly, but the filters have now changed to 11.111 MHz (figure 3b.) The problem is that the reference and filter frequencies are too far apart (one would like them to be spaced 700 kHz.) This constraint can be alleviated by changing the divide ratio to something other than 10. The mix and divide process can be generalized as shown in figure 3c. The frequencies obey the equation:

$$F_{REF} = \frac{N-1}{N} \cdot F_{IF}$$

where N is an integer equal to the frequency divider ratio. Figure 4a shows the results of this equation for values of N of interest, assuming F_{IF} = 10.7 MHz. The values of N that give the closest reference frequency to 10 MHz are 15 and 16, which result in 9.987 and 10.031 MHz respectively. It is now possible to take advantage of the fact that it is acceptable to operate the filter slightly off center, if a small decrease in tuning range can be tolerated. If the N=15 and N=16 configurations are scaled in frequency to get the refererence frequency to exactly 10 MHz, then the resulting IF frequencies are 10.714286 MHz (10 5/7) and 10.666667 MHz (10 2/3) respectively (figures 4a and 4b). With several hundred kHz of bandwidth available, sacrificing 14 or 33 kHz is a reasonable compromise. Although the IF is closer to 10.7 MHz with N=15, it is usually preferrable to use N=16 for three reasons:

1. The divider is easier to implement in digital hardware and gives an output with very low even harmonics.

2. It is usually easier to deal with fractional frequencies involving 1/3's than 1/7's (the lower the divisor the better) assuming it is necessary to convert the repeating decimal frequency to a more useful number. For example, 10 2/3 can be used to drive a frequency tripler to produce 32 MHz. With 10 5/7, a difficult to build frequency septupler would be required to produce 75 MHz.

3. A lower DDS frequency is possible, as will be explained below.

Front and back ends for the mix and divide chain

A chain constructed of the modules shown in figure 4c requires an input at 667 kHz. (i.e. 2/3 MHz) for the first module. In order to have the capability of acheiving zero offset for one of the steps, the input must be at exactly 2/3 MHz, not merely a good approximation. Most DDS's are binary based, and the remainder are decimal based. Neither of these can generate exactly 2/3 MHz starting from 10 MHz., no matter how much resolution is available. That would

take a ternary (base 3) DDS, which is theoretically possible, but highly unlikely to be available. Instead, the DDS is operated in a band centered at 2 MHz and is followed by a divide-by-3 frequency divider (figure 5.) Since exactly 2 MHz is available from virtually any DDS, it is usually possible to include a zero offset step. If the N=15 architecture were chosen, then the DDS would need to generate 5 MHz instead of 2 MHz and a divide-by-7 circuit would need to be used to obtain 714.286 kHz. (5/7 MHz).

At the end of the chain, 10.667 MHz is available with programmable offset, but of course it needs to be converted back to 10 MHz to be useful. This is accomplished with one additional mixer and filter stage. However in this case the LO for the mixer is the 10.667 MHz from the previous stage. and the filter is tuned to 10 MHz (figure 5.) An alternate output arrangement is shown in figure 12. It is slightly more difficult to implement but reduces the influence of the stage before the output stage by a factor 15, leaving only the final 10 MHz filter as a significant factor in phase drift. Compared to the original 5100 architecture (figure 1), this is a big improvement because the 3.X MHz and 24 MHz drift is eliminated. The technique of figure 12 can also be applied in some cases to the architectures for reference frequencies other than 10 MHz that are described below.

The output stage raises the issue of filtering, since the standard 10.7 Mhz filters are no longer applicable. At the end of the chain, the filter bandwidth required is negligible so it is possible to use a crystal filter. This filter should be as wide as possible to minimize phase noise and microphonics effects. (There is generally no reason for it to be narrow since its purpose is to reject 10.7 and 9.3 MHz.) Another possibility is to construct a 10 MHz filter from individual 10 MHz ceramic resonators, which are similar to the resonators incorporated into the 10.7 MHz filters. They are normally used as low priced replacements for microprocessor clock crystals, and 10 MHz is a readily available frequency. Their moderate Q is a good match for this requirement since the nearest frequency to be rejected is 7% away from the passband center. It is also possible, of course, to use a phase locked VCXO as a cleanup loop to get pure 10 MHz out. Figure 5 shows the complete synthesizer system design [7].

The blocks in figure 5 have been regrouped into iterated divide/mix/filter modules, which will be henceforth be referred to as "**frequency interpolators**". This organization is more logical for this design than the 5100 partitioning into mix/filter/divide modules.

Design constraints

The architecture of figure 5 raises the question of how many interpolators are appropriate. Each iteration decreases step size by 16 (which may or may not be very significant) and it also improves spectral purity (i.e. phase noise, spurious sidebands, jitter, etc). by 24 dB = 20 log 16 , which is usually extremely relevant. After a sufficient number of stages, the total tuning range may become too small to be usable or the 24 db per stage improvement in spectral purity may cease to occur because the noise floor of the interpolators or the reference frequency input becomes the limiting factor. The limit on tuning range depends on the number of interpolators and the bandwidth of the IF filters used. With the DBS type filters, about 400 kHz. of range is available. This results in final tuning ranges of 25 kHz., 1.5 kHz., 100 Hz, 6 Hz, and 0.4 Hz. for 1, 2, 3, 4, and 5 interpolations respectively. Since 5 interpolations gives a theoretical spectral improvement of 120 dB (in addition to the 10 dB from the divide by 3 after the DDS), it is unlikely that more than 5 interpolators would ever be used, even if less tuning range were acceptable.

Critical system design details

To get maximum performance for the overall design, it is important that the 10 MHz. distribution amplifier (not shown in figure 5) have high isolation between outputs to keep the slightly different IF frequencies of each interpolator from intermixing. Also, it is helpful for it to have low distortion, especially even order, in order to minimize mixer distortion. Of course it should also have low phase noise and good short term stability so as not to degrade the reference source. The various interpolators should be well isolated from each other by proper ground plane management and PC board layout or installed in individual shielded boxes if necessary. Finally, adequate power supply filtering should be used to preclude coupling through the power distribution system.

Critical module design details

At the front end of the interpolator, a high quality sine wave to logic level converter should be used. A excellent treatment of this subject is given in reference [5]. After converting to logic levels, a low jitter logic family such as 74ACXXX should be used for the divider, which is preferably of the synchronous type, not the ripple type. The divider must be followed by a low pass filter to suppress harmonics; an elliptic function filter is probably the most appropriate type for this purpose, with 5, 7, or 9 poles depending on the required performance. Alternately, a stepped sine wave can be generated by a shift register followed by a non-uniformly weighted summing amplifier to ease the filtering requirements [4]. A pad should be used between the filter and the mixer to properly terminate both the filter and the IF port of the mixer which is driven with the filtered 667 kHz. The amount of attenuation in the pad can be adjusted to give the desired mixer drive level. A drive level of -30 to -10 dBm is reasonable for a +7dBm. LO mixer depending on the required level of performance. For extremely high performance systems, it may be appropriate to use high LO drive level mixers. The LO for the mixers is supplied by the distribution amplifier.

The output of the mixer typically undergoes 2 to 4 stages of filtering. It is convenient to insert amplifiers between the filters to prevent interaction between the ceramic filters since they are not designed to be cascaded directly. The amplifiers not only must make up for the filter loss of 3 to 6 dB. but also provide a net gain to help raise the low level signal coming out of the mixer to a higher amplitude to drive the clock of the next divider. The filters, when used as recommended have a rounded ("stoop shouldered") amplitude response in order to get constant group delay, which is critical in FM demodulation. When several of them are cascaded, the resulting response has an even more pronounced rounding. Fortunately, it is possible to fix this problem by deliberately mismatching the filters. It is also possible to change the center frequency somewhat by adding capacitors (or inductors) in series with the input and output terminals. The effect of temperature on phase for a 2 pole ceramic filter is typically +/- 200 milliradians over the range of 0 to 70 degrees C. The phase does not accumulate through the interpolators. Rather, it is only critical in the last interpolator (unless the configuration of figure 12 is used) and the final frequency translation stage filtering.

Input and output frequencies other than 10 MHz

It is possible to change the final translation stage to produce frequencies of interest other than 10 MHz. For example, it might be desired to synthesize the telecom frequency of 2.048 MHz. using a 10 MHz reference. Figure 6 shows the last divider changed to divide-by-25 to convert 10.667 Mhz to 10.24 MHz. This can then be divided by 5 to get offset 2.048 MHz. In figure 7, the last divider has been returned to divide-by-16 but the previous ones have been changed to divide-by-25. This makes it possible to multiply a 2.048 MHz input by 5 to get 10.24 MHz which is used as the LO for the mixers. The IF frequency is still 10.667 MHz and the output stage of figure 5 is used to produce offset 10 MHz. Figure 8 combines the techniques of the two previous figures to get a synthesizer with 2.048 MHz input and offset 2.048 Mhz output. The architecture of figure 8 can also be used for a 10.23 MHz input, offset 10.23 MHz output synthesizer for GPS by omitting the multiply-and divide-by-5 functions. Figure 13 shows a scheme for generating 10.23 MHz starting from 10 MHz. In figure 9, the North American telecom frequency of 1.544 Mhz is multiplied by 4/3 to get 2.058667 MHz This is close enough to 2.048 MHz to use the scheme of figure 8, but it is more optimum to increase the divide ratio from 25 to 26 to get the signal closer to the center of the passband. The result in figure 9 is the North American equivalent of figure 8, with 1.544 Mhz in and offset 1.544 MHz out. In figure 10, output sections are shown for generating the Sonet/SDH frequency reference of 51.84 MHz or 1.544 MHz starting with either the 10 MHz input synthesizer shown in figure 5 or the 2.048 MHz input synthesizer shown in figure 7, which both produce 10.667 MHz as an intermediate frequency. Finally, figure 11 shows a synthesizer for exciting cesium beam tubes (or absorption cells) with a direct multiply frequency scheme. The advantage of a direct multiply scheme compared to the traditional sideband excitation is that it is easier to generate the required microwave power and there are a lot fewer spurious sidebands to contend with. The synthesizer produces 10.714 MHz which can be multiplied by 78 then multiplied by 11 resulting in the cesium transition frequency of 9.192 GHz. There is sufficient tuning range to move 50 kHz off the main line to the Zeeman line for measuring the effective C field magnitude [6]. The output of the times 78 multiplier in figure 11 is conveniently filtered by an off-the-shelf 836.5 MHz dielectric resonator filter of the type used for cellular telephone

transceivers. Modulation for operating the atomic servo loop can be applied via the DDS. The main caveat when using this scheme is that the times-78 and times-11 multipliers must be well shielded so that harmonics do not leak into the beam region of the CBT and cause extraneous quantum transitions. This is more likely to happen than with conventional architectures since the harmonics fall directly on the main Ramsey pattern.

Frequency modulator applications

If the DDS is replaced by a frequency or phase modulator, then the synthesizer can be used to reduce the modulation index. This allows generating the signal with much wider deviation than what is used on the air. The advantage is that any center frequency drift of the carrier is reduced correspondingly. This drift reduction permits the use of a true frequency modulator utilizing a VCO with DC coupled tuning voltage rather than an AC coupled technique such as an Armstrong modulator. DC coupled modulation is important in asynchronous data transmission such as paging. Current workarounds for this problem involve questionable schemes such as putting a very narrow phase locked loop on the VCO so that the data is "almost" DC coupled or using a DDS to generate mark and space frequencies. The problem with using a DDS is that the data pulse shaping filters have to be digital filters. Some systems try to get around this by using a one-size-fits-all look up table that describes the transition from mark to space or vice versa. This method suffers from intersymbol interference and spectral splatter problems.

Experimental verification

An experimental model was built using two interpolators with 3 filters each and a final 10 Mhz filter with 5 discrete ceramic resonators. A synthesized signal generator was used in place of the DDS. With a moderate amount of attention to the details enumerated above, it was easy to keep spurious sidebands down at least 120 dB. The ceramic filters did not significantly increase phase noise.

References

[1] D. W. Allan and H. Daams, "Picosecond time difference measurement system," in Proceedings of the 29th Annual Symposium on Frequency Control, 1975, pp. 404-411.

[2] R. L. Hamell, P. F. Kuhnle and R. L. Sydnor, "An Improved Offset Generator Developed for Allan Deviation Measurement of Ultra Stable Frequency Standards," in Proceedings of the 23rd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting (NASA confereence publication 3159), 3-5 Dec. 1991, pp. 209-218.

[3] V. E. Van Duzer, "A 0-50 Mc Frequency Synthesizer with Excellent Stability, Fast Switching, and Fine Resolution," Hewlett-Packard Journal, vol. 15, no. 9, May 1964, pp. 1-6.

[4] R. K. Karlquist, "A New RF Architecture for Cesium Frequency Standards," in Proceedings of the 1992 IEEE Frequency Control Symposium, 27-29 May 1992, pp. 134-142.

[5] G. J. Dick, P. F. Kuhnle and R. L. Sydnor, "Zero-Crossing Detector with Sub-Microsecond Jitter and Crosstalk," in Proceedings of the 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting (NASA Conference Publication 3116), 4-6 Dec. 1990, pp. 269-282.

[6] L. S. Cutler and R. B. Giffard, "Architecture and Algorithms for New Cesium Beam Frequency Standard Electronics," in Proceedings of the 1992 IEEE Frequency Control Symposium, 27-29 May 1992, pp. 127-133.

[7] U. S. Patent Pending.

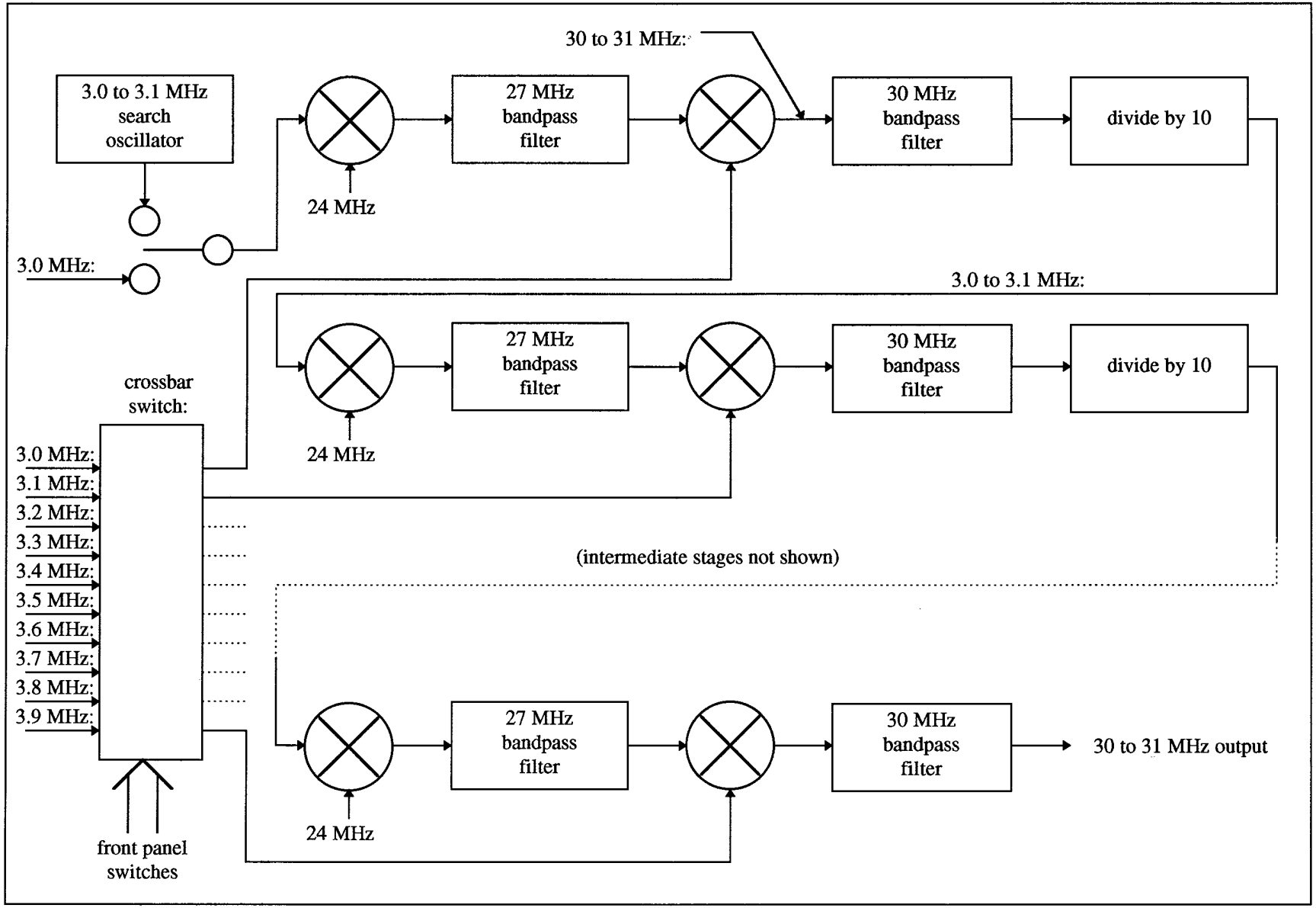


Figure 1. HP 5100 type architecture

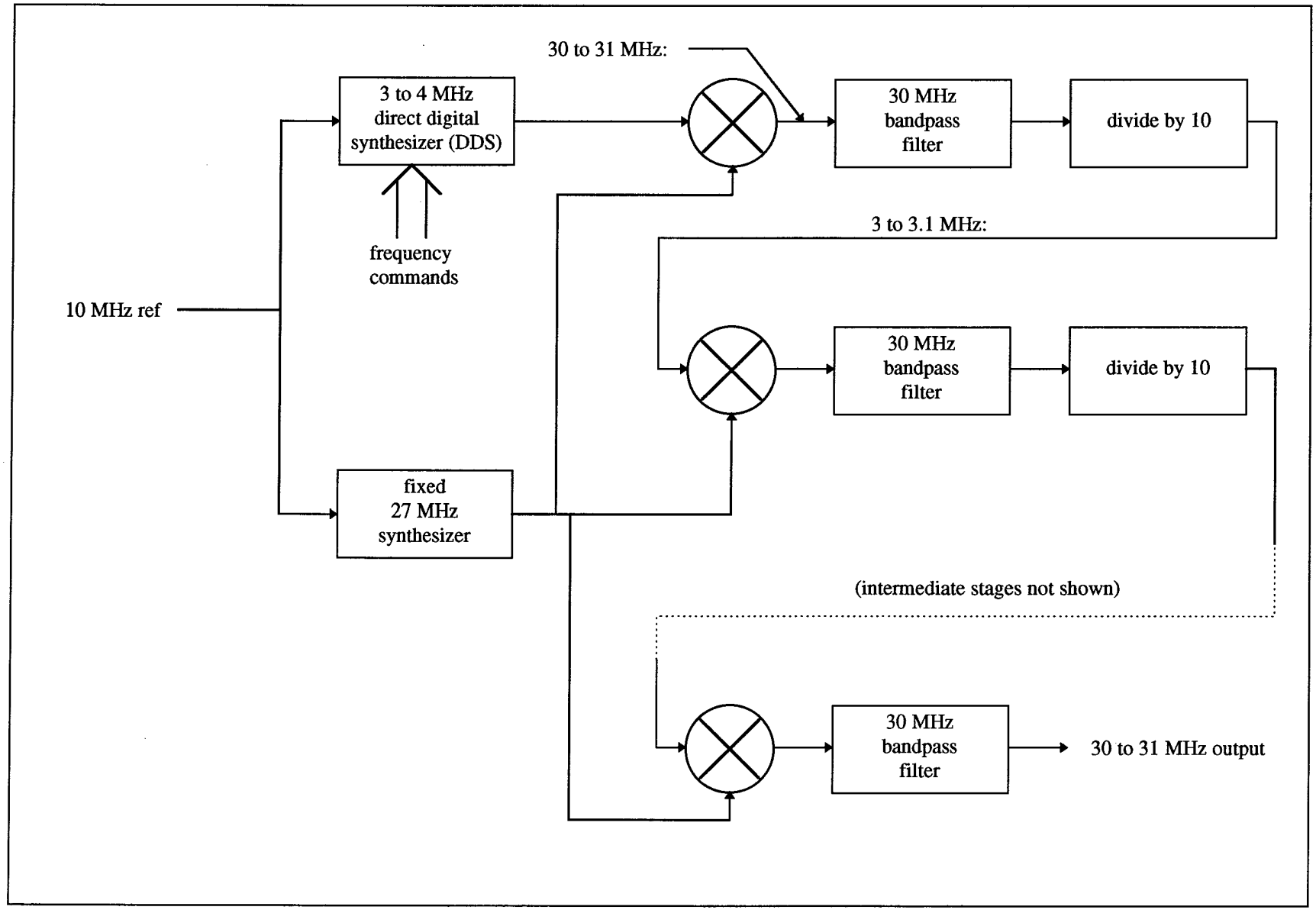


Figure 2. Narrow band 5100 type architecture with DDS source

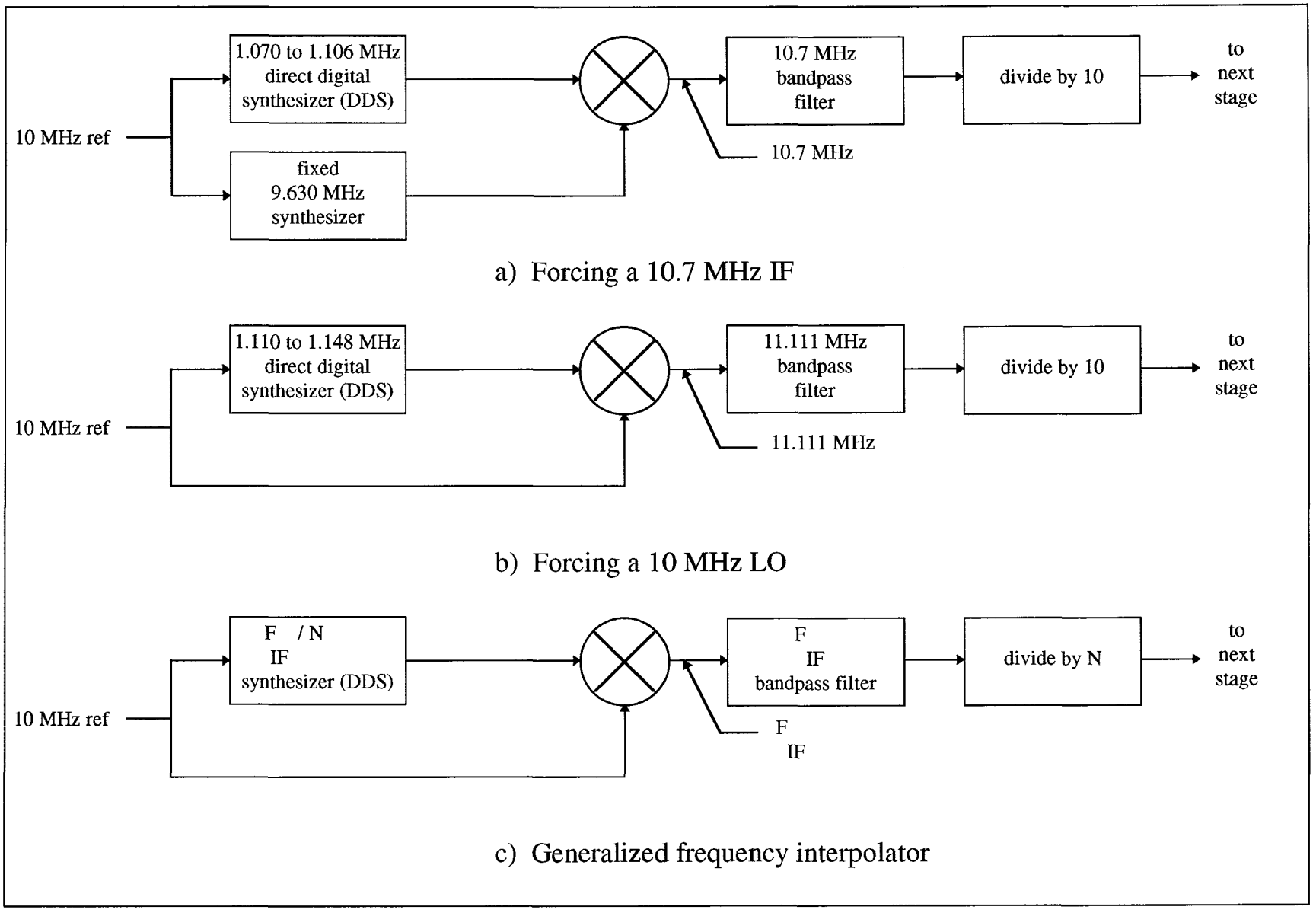


Figure 3: Transformations of 5100 architecture

N:	Ref. freq. for 10.7 MHz IF:
4	8.025000 MHz
5	8.560000 MHz
6	8.916667 MHz
7	9.171429 MHz
8	9.362500 MHz
9	9.511111 MHz

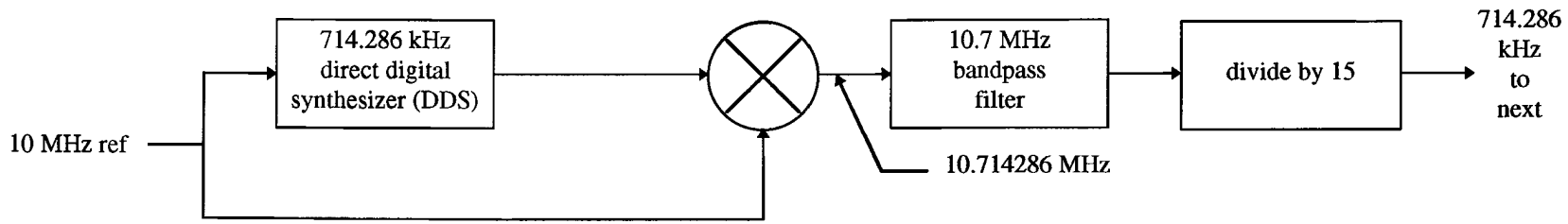
N:	Ref. freq. for 10.7 MHz IF:
10	9.630000 MHz
11	9.727273 MHz
12	9.808333 MHz
13	9.876923 MHz
14	9.935714 MHz
15	9.986667 MHz

N:	Ref. freq. for 10.7 MHz IF:
16	10.031250 MHz
17	10.070588 MHz
18	10.105556 MHz
19	10.136842 MHz
20	10.165000 MHz
21	10.190476 MHz

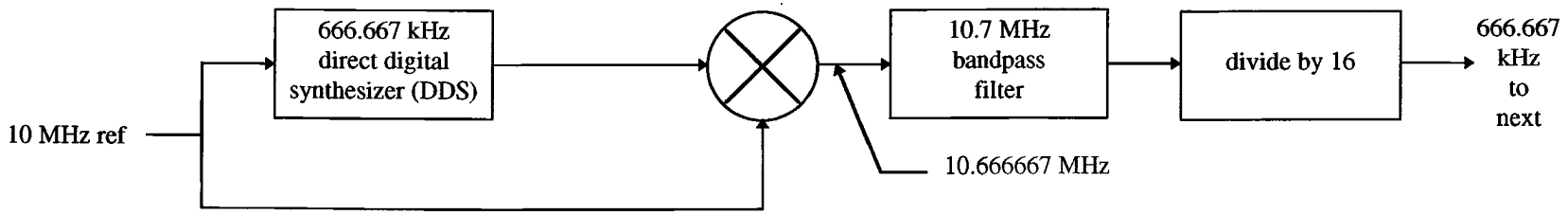
N:	Ref. freq. for 10.7 MHz IF:
22	10.213636 MHz
23	10.234783 MHz
24	10.254167 MHz
25	10.272000 MHz
26	10.288462 MHz
27	10.303704 MHz

N:	Ref. freq. for 10.7 MHz IF:
28	10.317857 MHz
29	10.331034 MHz
30	10.343333 MHz
31	10.354839 MHz
32	10.365625 MHz
33	10.375758 MHz

a) Possible frequencies



b) N=15, 10.714286 MHz IF, 10 MHz ref.



c) N=16, 10.666667 MHz IF, 10 MHz ref.

Figure 4: Usable frequency schemes for 10.7 Mhz ceramic filters

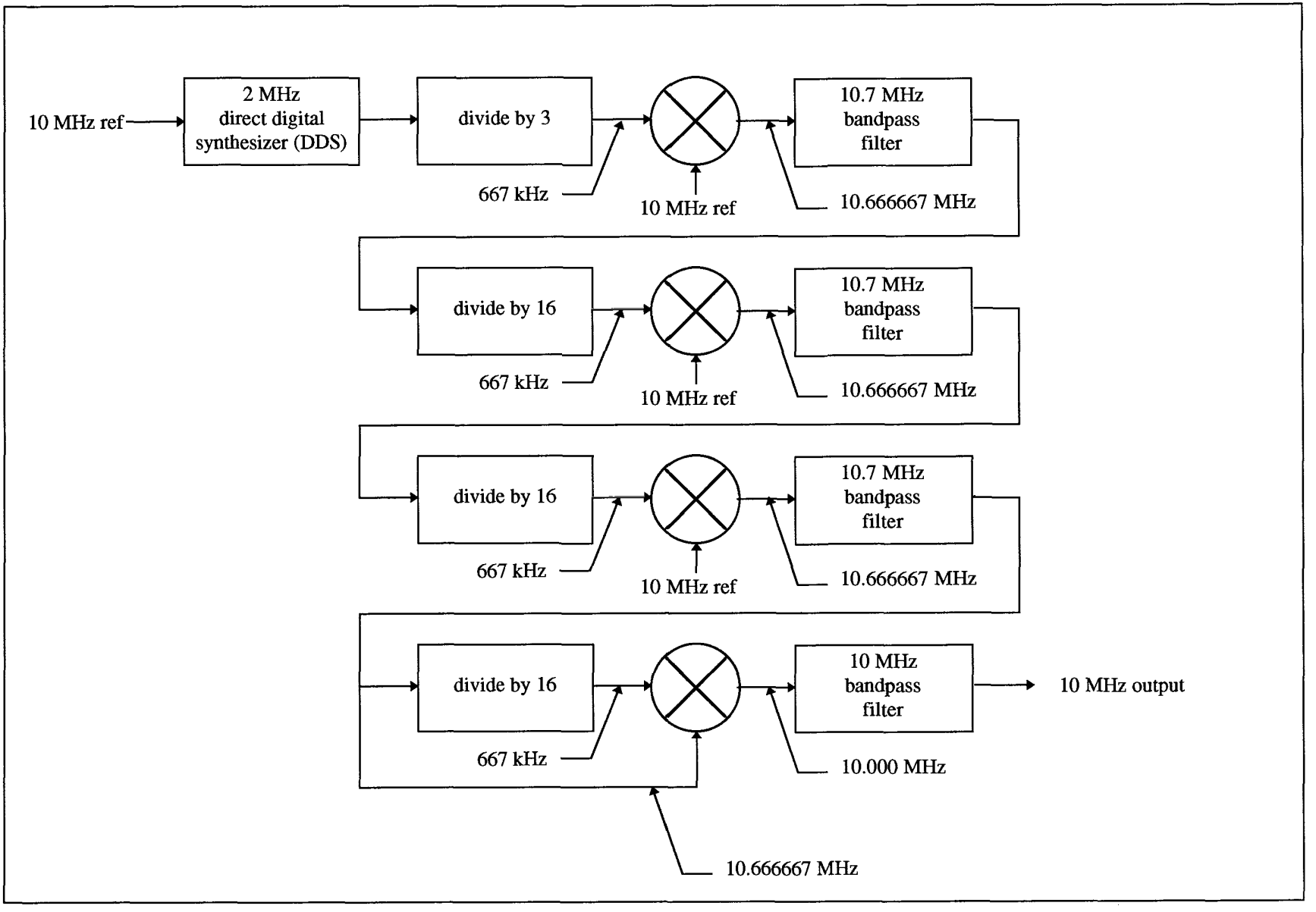


Figure 5. Practical 10 MHz output, 10 MHz ref synthesizer

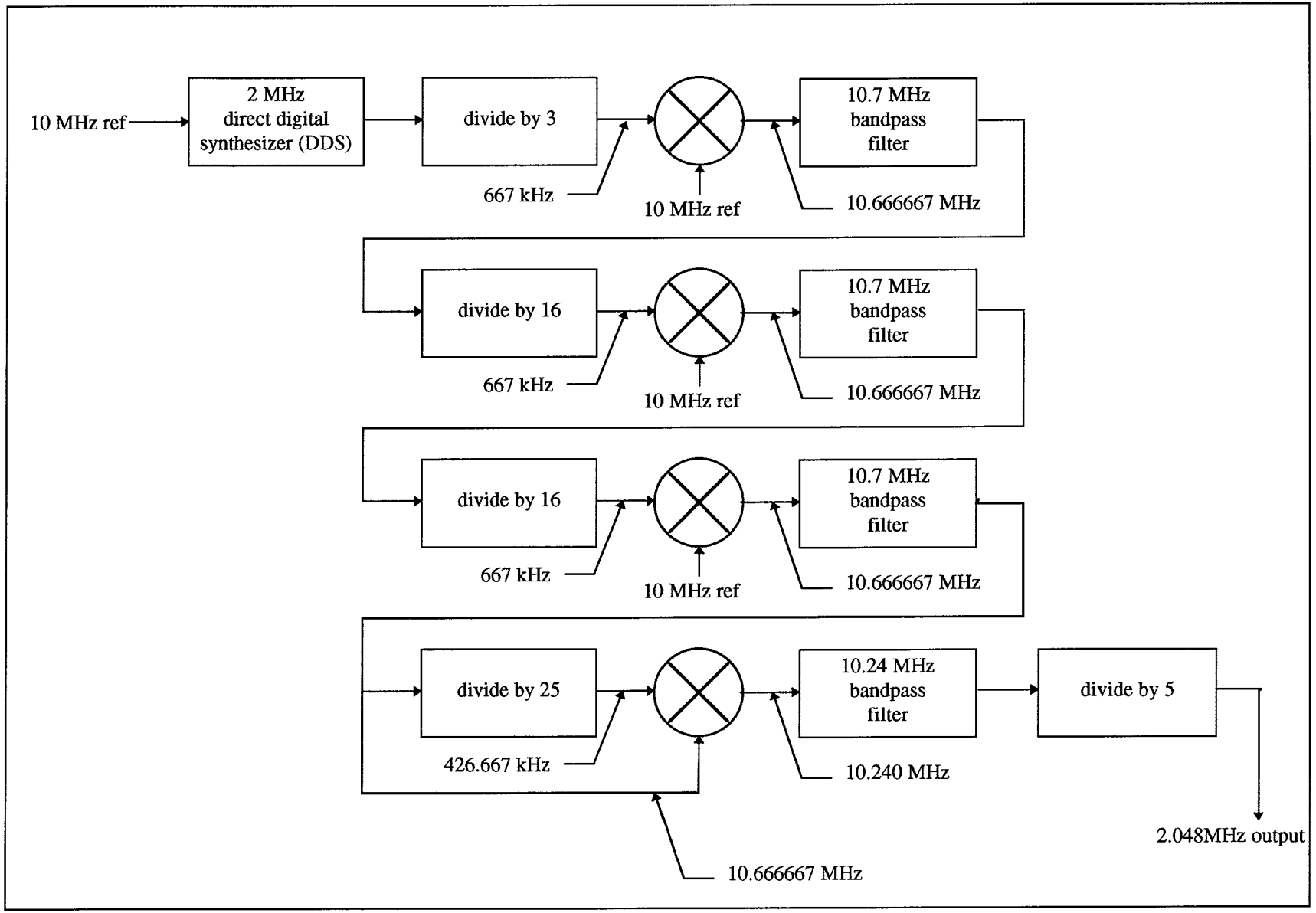
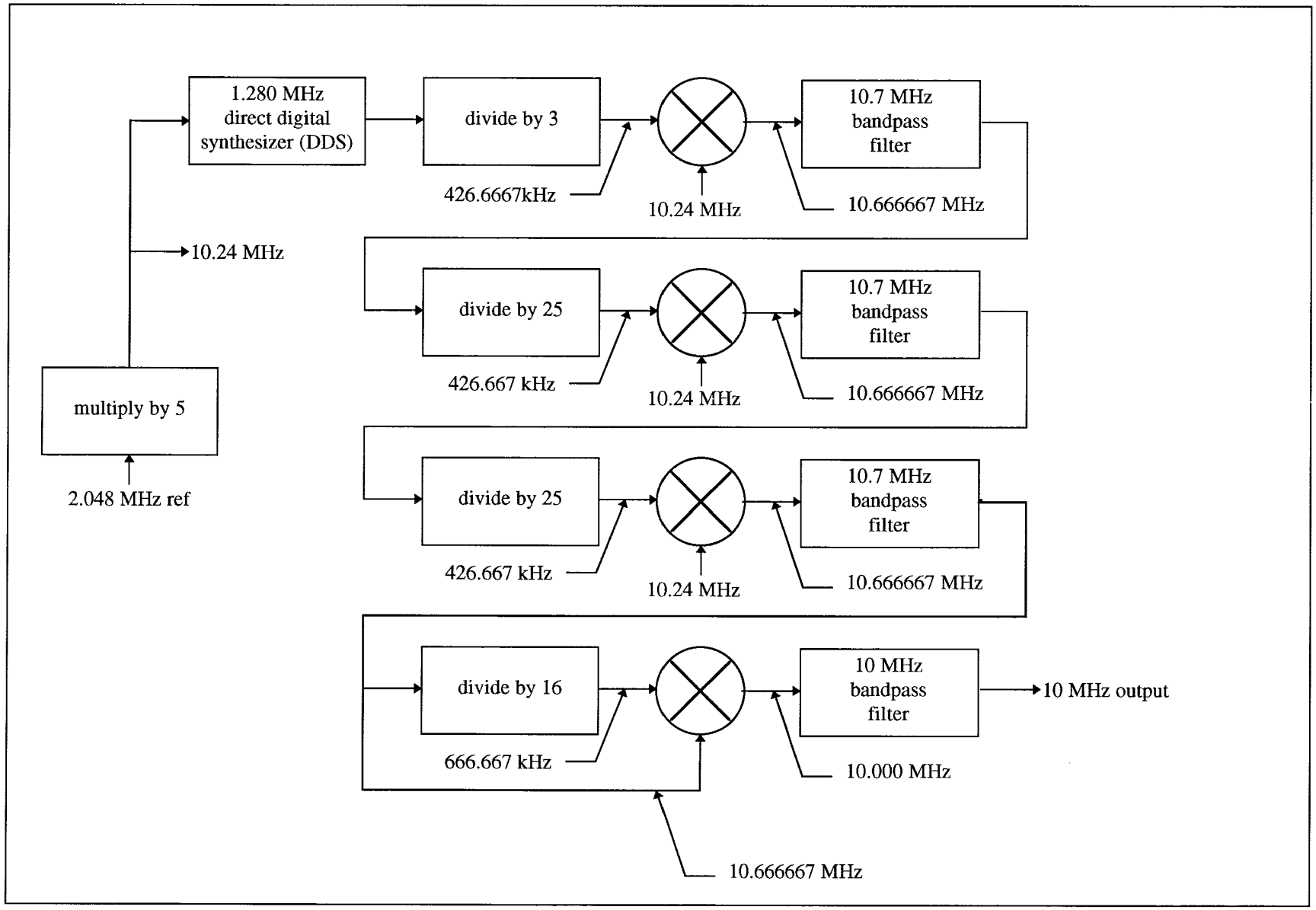


Figure 6. 2.048 MHz output, 10 MHz ref synthesizer



230

Figure 7. 10 MHz output, 2.048 MHz ref synthesizer

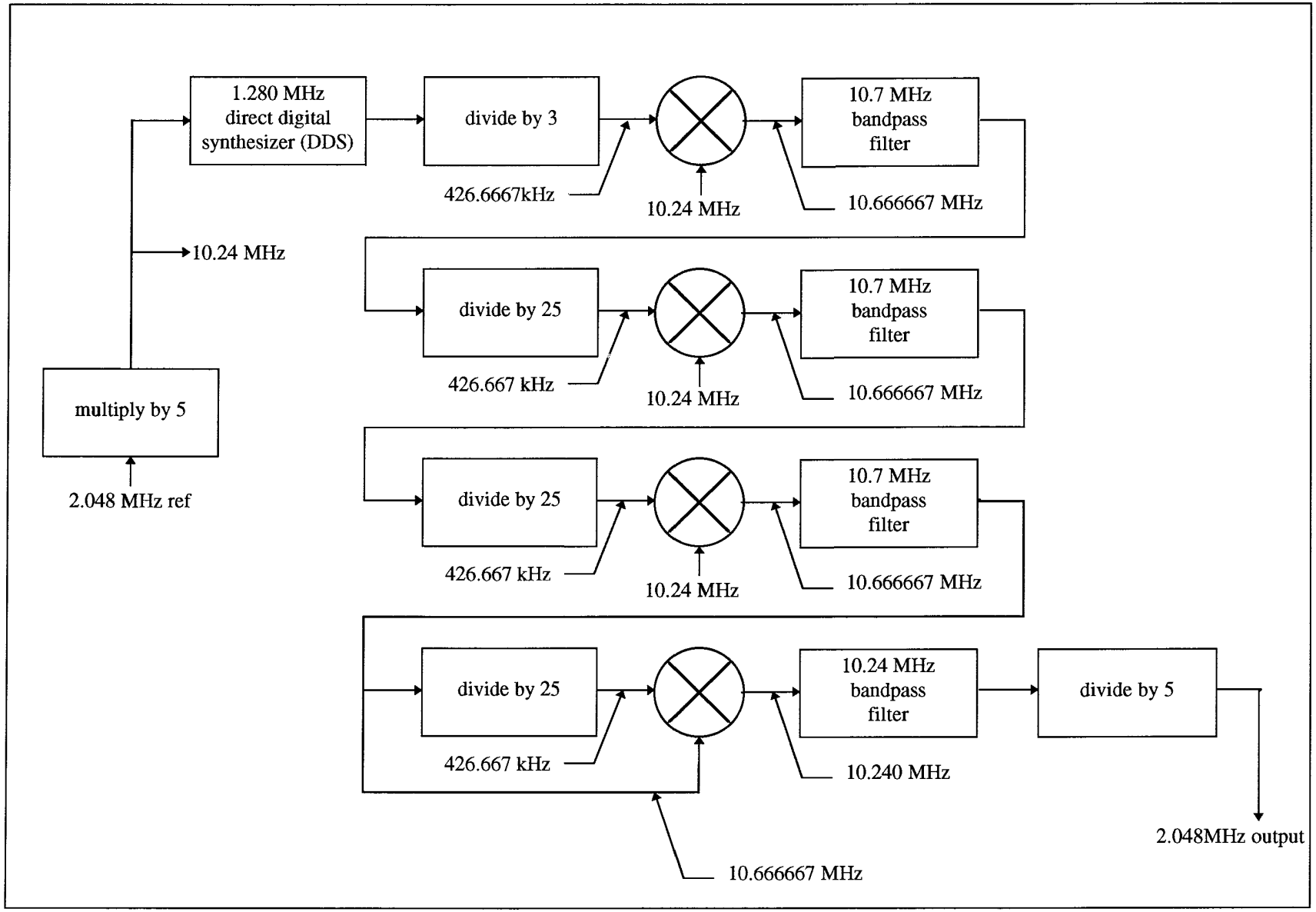


Figure 8. 2.048 MHz output, 2.048 MHz ref synthesizer

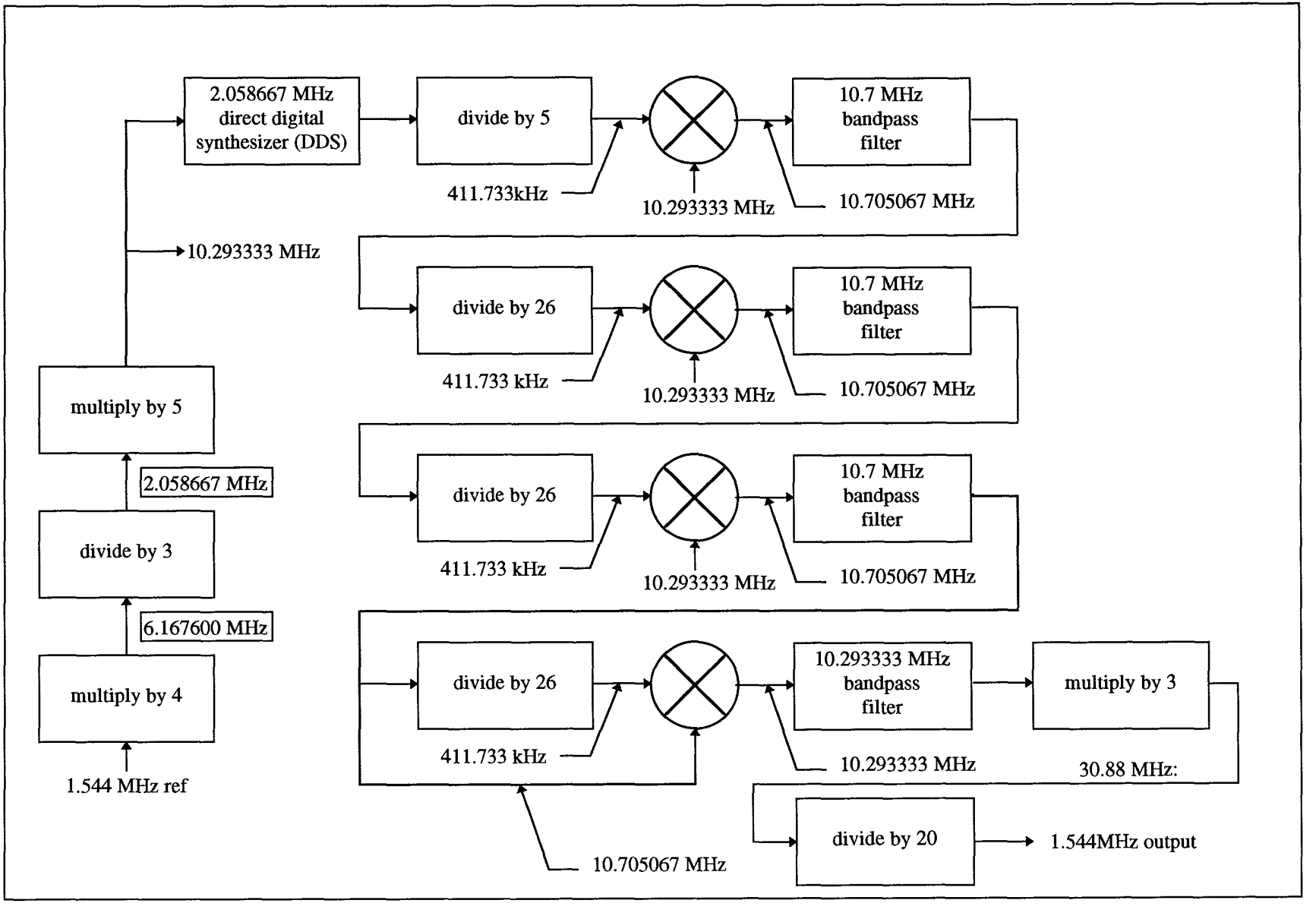


Figure 9. 1.544 MHz output, 1.544 MHz ref synthesizer

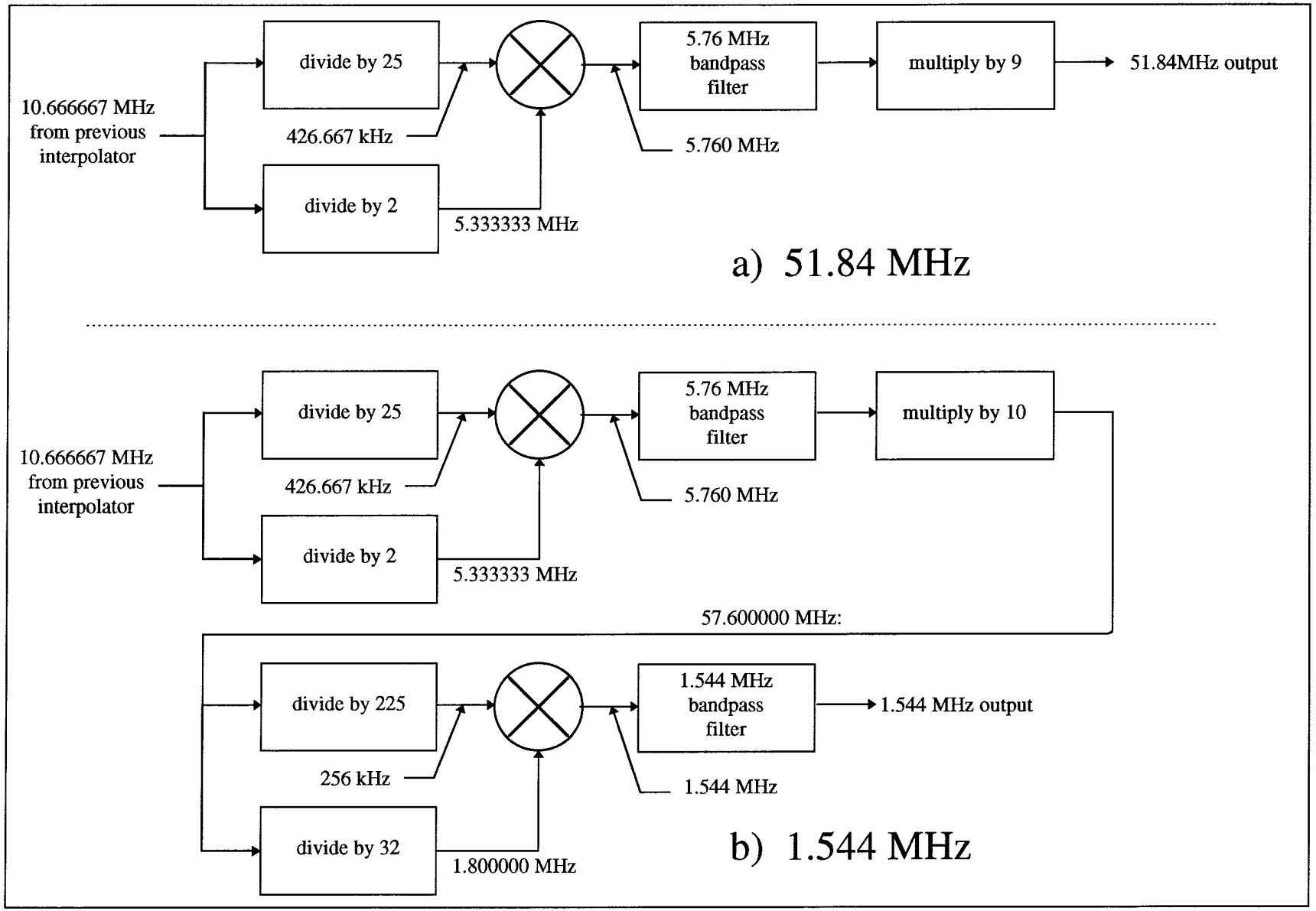


Figure 10. Output sections for 51.84 MHz and 1.544 MHz

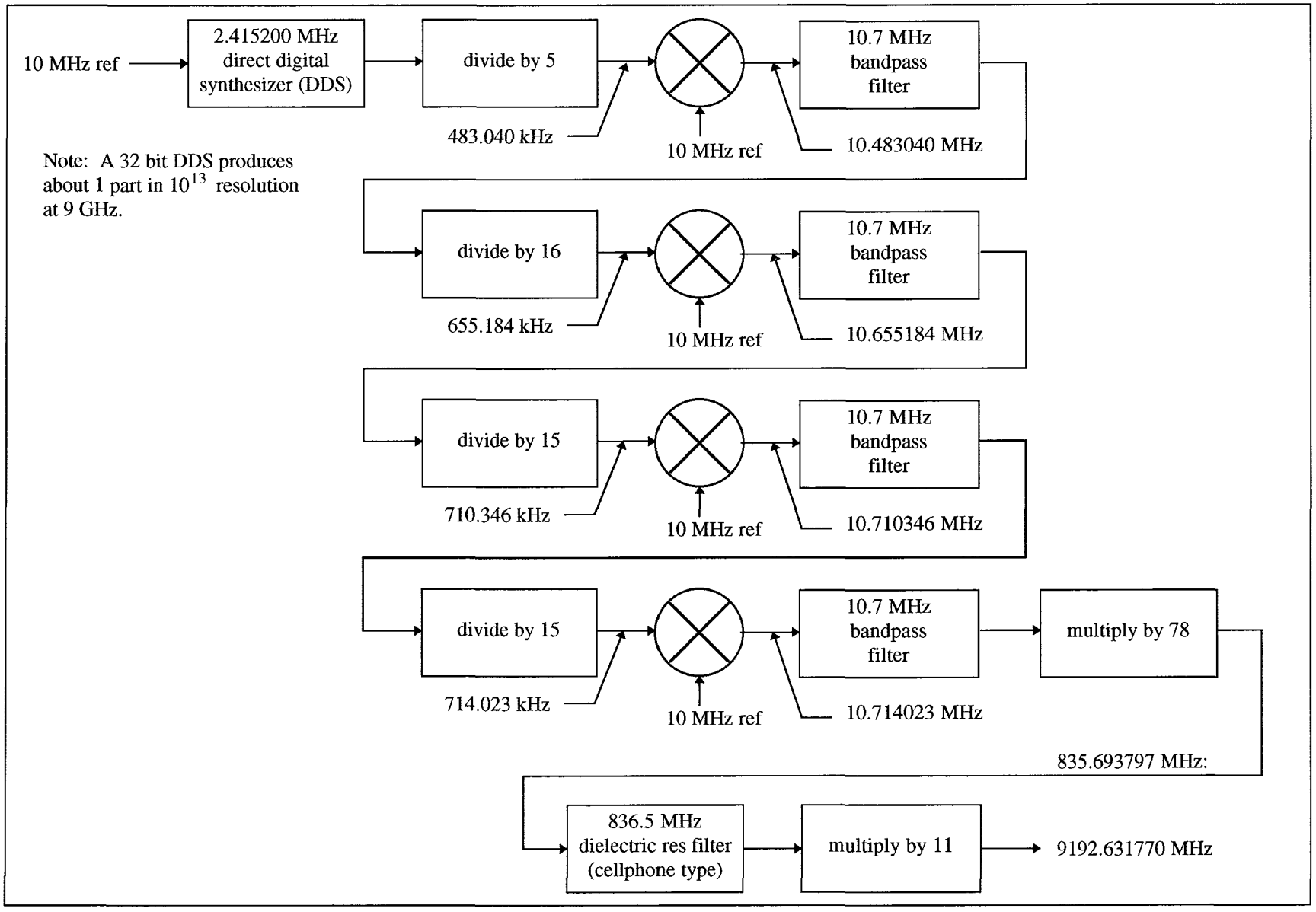


Figure 11. RF chain for cesium standard

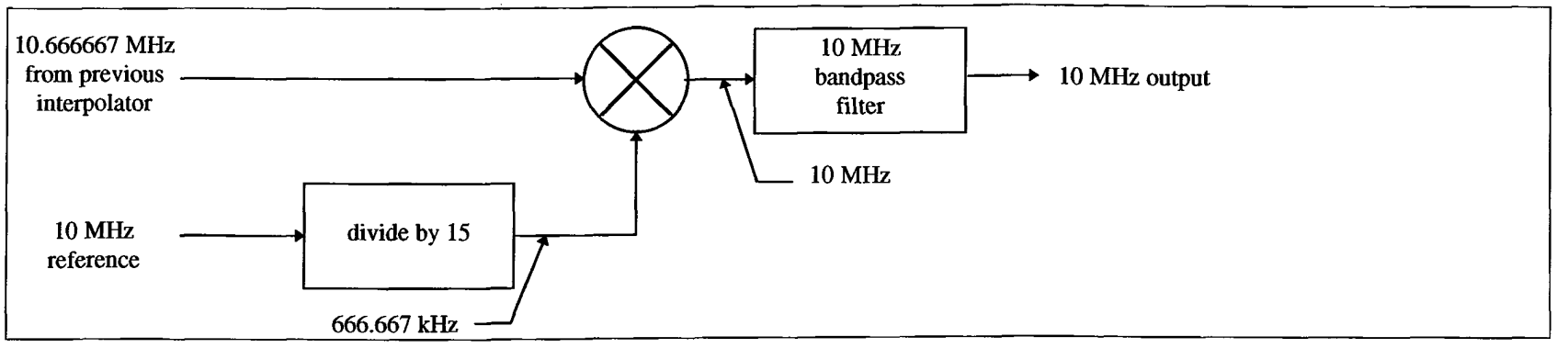


Figure 12. High performance 10 MHz output section

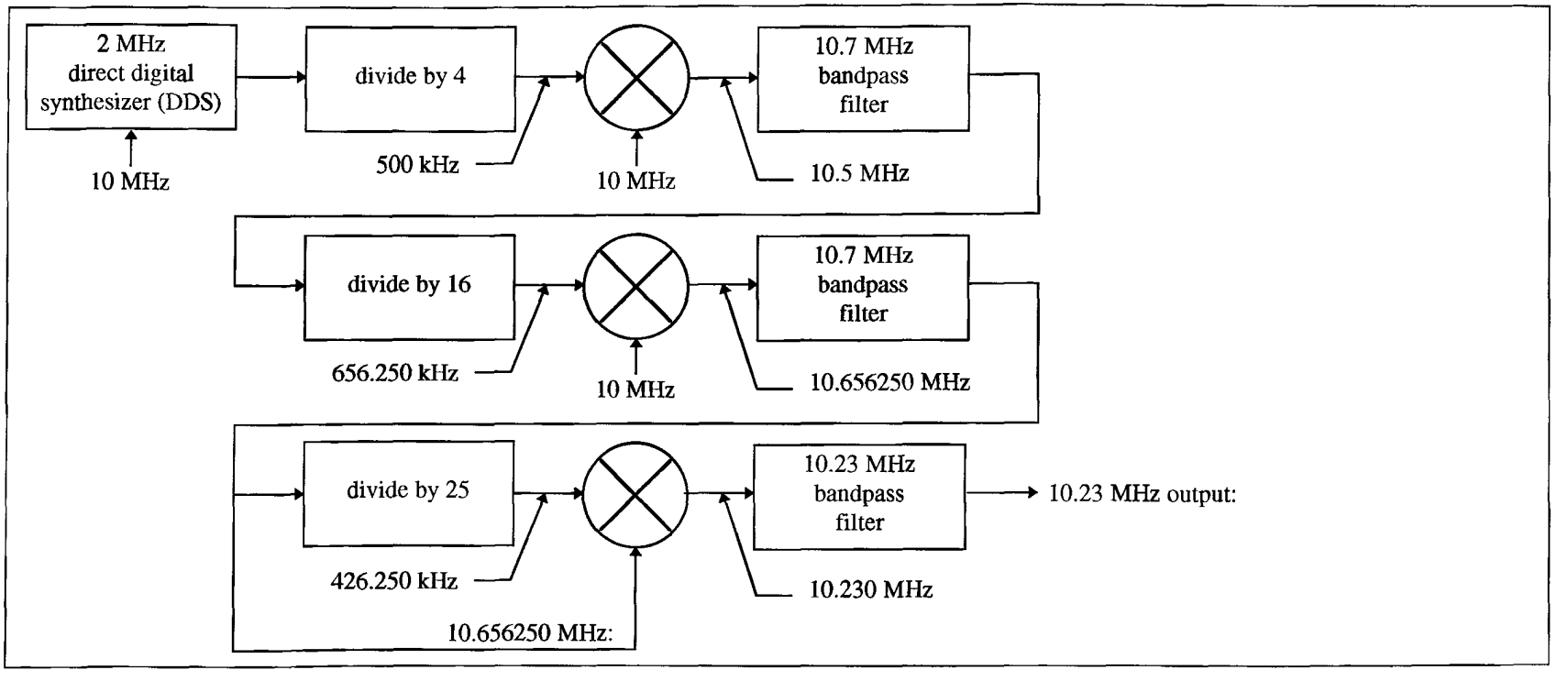


Figure 13. 10.23 MHz output, 10 MHz reference synthesizer

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

NARROW-BAND DIRECT DIGITAL SYNTHESIZERS

V.N.Kochemasov and A.N.Zharov

Moscow Technical University of Communications and Informatics
Aviamotornaya, 8a, Moscow 111024, Russia

Abstract

The direct digital synthesizers (DDS) are widely used in the synthesis of frequencies and signals. If the narrow band is needed, the frequency conversion methods with using phase shifting technique are useful. These methods allows to increase the maximal synthesized frequency, to decrease the cost and power consumption of the synthesizer without lost in accuracy of synthesized waveform.

We offer a new type of synthesizer - direct digital synthesizer with samples commutation (DDS SC), using multilevel digitally controlled phase shifter.

We obtained the mathematical expressions for the Fourier coefficients of the output signal in DDS with sine wave, two level (square wave) and multilevel digitally controlled phase shifters, when the DDS operates at the frequency synthesizing mode and defined the possibility of using these DDS for generation of signal with linear frequency modulation (LFM).

Introduction

The high requirements to the accuracy of signal generation equipment for communication and radar systems [1-3], and developments in integrated circuit technique resulted in DDS creation. The DDS [4,5] are wide spread now due to their main benefiting features: being highly technological in manufacturing, reliability, low phase noise level, destabilize factors busting ability, and small time of frequency switching.

The technology achievements allow to overcome continuously the main DDS shortcomings: low values of the synthesized frequencies and large power consumption. But the difference in operation rates of DDS components (code accumulator, "$\varphi - \sin$" function converter (FC), digital to analog converter (DAC) in multilevel DDS, and code accumulator, digitally controlled delay line, code divider in two level DDS [6-8]) leads to the situation, when the DDS operation rate is defined by the "slowest" unit and the real abilities of the fast units are not fully used.

So, in the case of comparatively narrow band of the synthesized signals and frequencies, there is the interest to the frequency conversion using digitally controlled phase shifters (DCPS). These DCPS can deal with digital samples of sine wave (multilevel DCPS [9]), square wawe signal (two level DCPS [10]) and sine wave signal (sine wave DCPS)

Digital frequency converting synthesizer operation principle is based on the generation of high and low frequency N-phase signal systems, multiplication of the j-signals (j=0,1,2,...,N-1) and summing the multiplication results. As one of such systems one can use two level signals:

$$U_j(t) = \begin{cases} 1 & \text{when } (i+j) \bmod N = 0, \\ 0 & \text{for other } i, \end{cases} \qquad (1)$$

which period is breaked for N equal intervals, where j=0,1,2,..., (N-1) - the signal number; i=0,1,2,...,(N-1) - the interval number; $(i+j)\bmod N$ - the operation of adding by module N. In a signal system (1) only one signal has high level at any moment of time, while the others have low level. In this case the commutator composed of N switch keys operates as multiplier, and the operation of adding N multiplication results is transformed to the operation of sequel passing (with frequency f_c) signals to the commutator output.

When using two level signals (1) in the high frequency system, and codes of sine wave samples in the low frequency system, the commutator of the DDS provides the sequel (with f_c frequency) passing of N samples to the DAC input. We call such a device "the DDS with samples commutation". In this synthesizer the value of the maximal synthesized frequency $f_{s.max}$ depends on the commutator and the DAC operation rates (switching frequency f_c) only, and the output frequency bandwidth depends on the accumulator clock frequency f_{cl} and function converter switching rate only. In fact, this synthesizer is a digital form of the single side band frequency conversion circuit.

The second group of narrow band DDS consists of synthesizers, in which the synthesized signal is formed at the output of delay line or phase shifter, controlled by the code from the phase accumulator output [7,12]. In this case signals (1) are used as a low

0-7803-2500-1/95 $4.00 © 1995 IEEE

frequency system, and N-phase two level or sine wave signals form a high frequency system. The maximal synthesized frequency in such DDS is defined by the digitally controlled delay line or the phase shifter maximal operation frequency.

The purpose of this paper is:

1) to find the mathematical expressions for the output spectrum waveform and the spurious spectrum components (SSC) and to define conditions of effective filtering of SSC to obtain their minimal level within the operation band;

2) to define the possibility of using these DDS to generate linear frequency modulation (LFM) signals.

In solving the second problem the LFM signals generation accuracy will be evaluated using the Hamming weight function compression filter output response, when the weight function provides the sidelobe level -42,8 dB [1,13] with rather long signal time T_s to bandwidth W product (base) WT_s.

Direct Digital Synthesizer with Samples Commutation

The DDS SC (Fig. 1) operation principle is based on calculation at clock moments $t_{cl}=n/f_{cl}=nT_{cl}$ the phase code K_φ in the phase forming unit (PFU), and on calculation the corresponding N codes of sine wave samples $K_U(k)=\sin\{2\pi[K_\varphi/L+(k-1)/n]\}$ in the FC1... FCN, where $f_{cl}=f_c/p$ - the PFU clock frequency, f_c - the reference oscillator output frequency, p - integer value, k=1,2,3,...,N, L - phase code accumulator capacity.

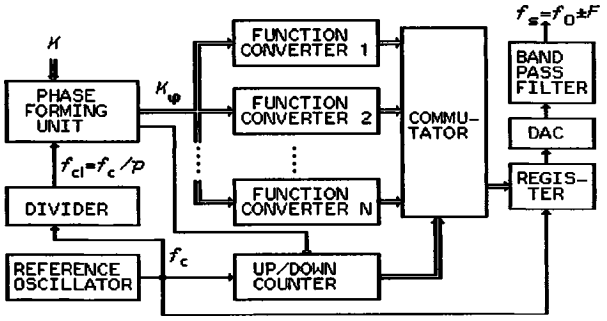


Figure 1. Direct digital synthesizer with samples commutation (DDS SC)

By the up/down counter controlled commutator, the samples are sequentially passing with high commutation frequency f_c through the register to the DAC, which output signal contains the useful frequency component $f_s=f_0+F$ or $f_s=f_0-F$ (dependent on the count direction of up/down counter). Here, f_0 - conversion frequency (central frequency of the band), F - shift frequency of code $K_F=LF/f_{cl}$.

To find the output signal spectrum of DDS SC with any phase quantity N operating in frequency synthesizing mode, we assume that the synthesized signal samples are calculated in discrete time moments with high precision. When $K_F=0$, the phase code K_φ at the PFU output remains constant and is defined by the initial conditions. The output signal is described by the staircase function:

$$u(t,\varphi)=u_0\sin[\varphi+(2\pi/N)\mathrm{ent}\{t/T_c\}, \quad (2)$$

when $i_cT_c<t<(i_c+1)T_c$. Here i_c- the commutation period number, $\mathrm{ent}\{t/T_c\}$ - the integer part of t/T_c.

In frequency synthesizing mode the DDS SC output signal has the maximal period $T=LpNT_c$ and contains $I=LN$ elementary signals, where i-th signal is a result of multiplication of signal (2) (with the initial phase $\varphi=2\pi K_F i/L$) and the weight function

$$w_i(t)=\begin{cases}1 \text{ when } aT+iT_{cl}<t\le aT+(i+1)T_c,\\ 0 \text{ for other } i,\end{cases}$$

where $a=0,\pm1,\pm2,\pm3\ldots$, $T_{cl}=pT_c$.

Assuming $u_0=1$ at (2) and taking in account, that $\mathrm{ent}\{t/T_c\}=k+ip$, where k=0,1,...,(p-1) - the elementary signal stair number, we obtain the DAC output signal Fourier spectral coefficients:

$$C_m=\frac{1}{T}\sum_{k=0}^{p-1}\sum_{i=0}^{I-1}\sin 2\pi\left(\frac{K_Fi}{L}+\frac{k}{N}+\frac{pi}{N}\right)\times$$
$$\times\int_{(iN+k)T_c}^{(iN+k+1)T_c}u(t,\varphi)\exp\left\{-\frac{\mathrm{j}2\pi mt}{T}\right\}\mathrm{d}t.$$

After the integration, taking in account that

$$\sin x=(e^{\mathrm{j}x}-e^{-\mathrm{j}x})/2\mathrm{j}, \quad (3)$$

we obtain:

$$C_m=\frac{1}{\mathrm{j}2\pi m}\sin\left(\frac{\pi mT_c}{T}\right)\exp\left(-\frac{\mathrm{j}\pi mT_c}{T}\right)\times$$
$$\times\left\{\sum_{k=0}^{p-1}\exp\left[\mathrm{j}2\pi k\left(-\frac{mT_c}{T}+\frac{1}{N}\right)\right]\times\right.$$
$$\times\sum_{i=0}^{I-1}\exp\left[\mathrm{j}2\pi i\left(\frac{K_F}{L}+\frac{p}{N}-\frac{mT_cp}{T}\right)\right]-$$
$$-\sum_{k=0}^{p-1}\exp\left[-\mathrm{j}2\pi k\left(\frac{mT_c}{T}+\frac{1}{N}\right)\right]\times$$
$$\left.\times\sum_{i=0}^{I-1}\exp\left[-\mathrm{j}2\pi i\left(\frac{K_F}{L}+\frac{p}{N}+\frac{mT_cp}{T}\right)\right]\right\}. \quad (4)$$

Using ratio [13] for the geometric progression sum

$$\sum_{i=1}^{L} x^i = \frac{1-x^{L+1}}{1-x}, \qquad (5)$$

that sum by i members produces in the first ratio (4) adding member. This sum equals to

$$S' = \frac{1-\exp[j2\pi(NK_F + Lp - m)]}{1-\exp[j2\pi(NK_F + Lp - m)/LN]}$$

and it differs from zero only when $m = m'$, where $m' = LNQ + NK_F + Lp$, Q=0,±1,±2,±3,..., as $NK_F + Lp - m$ is an integer value. Using the Lopitals rule we obtain $S' = LN$ when $m = m'$. In the same way it is possible to show, that in the second adding member (4) sum by i equals $S'' = LN$, when $m = m''$, where $m'' = LNQ + NK_F + Lp$, Q=0, ±1, ±2, ±3,..., and $S'' = 0$ in the case of other m.

Since $m' \neq m''$, equation (4) contains only one differing from zero adding member, and may be transformed into:

$$C_{m,s} = \frac{sLN}{j2\pi m}\sin\left(\frac{\pi m}{LNp}\right)\exp\left(-\frac{j\pi m}{LNp}\right)\times$$
$$\times\sum_{k=0}^{p-1}\exp\left[\frac{j2\pi}{N}\left(s-\frac{m}{Lp}\right)k\right],$$

where s=±1.

Using equations (3) and (5), we obtain:

$$C_{m,s} = \frac{s}{j}|C_{m,s}|\exp\left\{j\frac{\pi}{N}\left[s(p-1)-\frac{m}{L}\right]\right\}, \qquad (6)$$

where

$$|C_{m,s}| = \begin{cases} \left|\frac{1}{2p}\cdot\frac{\sin\left[\frac{\pi m}{NLp}\right]}{\left[\frac{\pi m}{NLp}\right]}\cdot\frac{\sin\left[\frac{\pi p}{N}\left(\frac{m}{Lp}-s\right)\right]}{\sin\left[\frac{\pi}{N}\left(\frac{m}{Lp}-s\right)\right]}\right| \\ \text{when } m = sK_FN + sLp + LNQ, \\ 0 \text{ for other } m, \end{cases} \qquad (7)$$

s=±1, Q=0, ±1, ±2,±3,..., $p = f_c/f_{cl}$, $K_F = LF/f_{cl}$ - the integer values. Here $|C_{m,s}|$ is the magnitude of spectrum component of number m, where the number is connected with frequency f by the ratio $f/f_T = m/LN$.

The normalized frequency values $f/f_{cl} = m/LN$, for that $|C_m| \neq 0$, are defined by the abscissas of the crosspoints of nomogram (Fig. 2) lines build by the ratio

$$m = sK_FN + sLp + LNQ \qquad (8)$$

and the horizontal dashed line crossing the Y-axis at the point corresponds to the chosen synthesized frequency value $f_s/f_{cl} = (f_0 \pm F)/f_{cl} = p/N + K_F/L$.

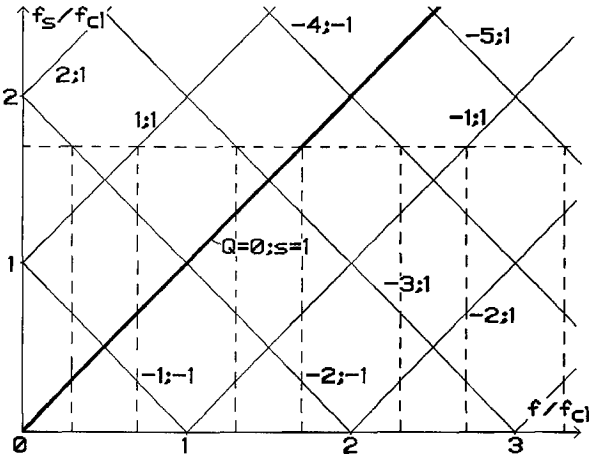


Figure 2. A nomogram that shows the DDS SC output signal spectrum components positions at the frequency axis

In this case, the useful spectrum component position on the frequency axis is defined by the abscissa of the crosspoint of the horizontal dashed line and the nomogram line corresponding to Q=0, s=1 values.

To ease the filtering it is necessary to choose such a normalized synthesized frequency f_s/f_{cl} position, for the spurious spectrum components bands to be located as far as possible from f_s/f_{cl} values. When synthesizing frequencies located symmetrically relative to conversion frequency f_0, the normalized frequencies f_0/f_{cl} positioning is required so, that SSC would also be located symmetrically relatively f_0, that is choosing p from the ratio:

$$p = (R/2 + 1/4)N, \qquad (9)$$

where R=1,2,3... . As p is integer, ratio (9) may be correct for proportional to 4 values N only.

Fig. 2 nomogram allows to define the position of spectrum components at the output of DDS SC operating at the frequency synthesizing mode, but it does not provide the information about its magnitudes.

It is possible to get the illustration of the useful and spurious spectrum components location at the frequency axis and their magnitudes by the summar spectrum analysis. The summar spectrum is produced by the DDS SC output spectrum components superposition. The components are calculated by ratios (6) and (7), when the synthesized frequency changes from $f_0 - F$ to $f_0 + F$, that is in the band W=2F, with the central frequency f_0.

When choosing the phases quantity N different from value proportional to 4, the spurious-free frequency band is not symmetric respective f_0 (Fig. 3,a,b). To avoid masking by the SSC, here and further on there are reproduced only useful spectral components summar spectrum band frame.

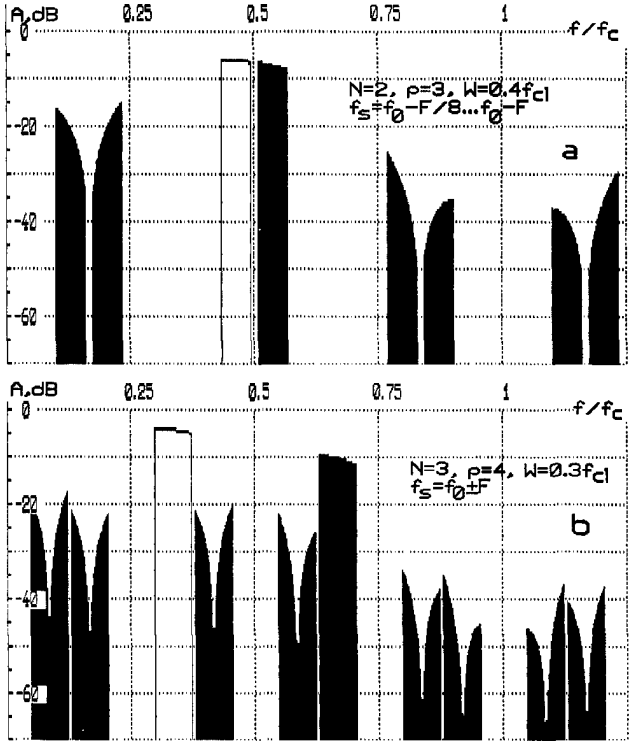


Figure 3. A summar spectrum of signals at the DDS SC output (frequency synthesizing mode) with phase quantity N=2 (a) and N=3 (b)

When N=2 there are the high and low side bands with frequencies f_0+F and f_0-F (Fig. 3,a) in the output signal spectrum - as at the double balance multiplier output. Choosing the phase quantity N=2 is proved, when extremely high frequency synthesing is necessary. If N>2, then in the case of $K_F \neq 0$, only one spectral component (with frequency f_0+F or f_0-F) is formed af the DDS SC output.

The DDS SC variant with N=4 (the digital analog of quadrature single sideband frequency conversion circuit) satisfies the contradictive requirements of the output signal wide band, low level of inband spurious components, low equipment and power consumption in the most complete way. In this case, in accordance with (9), the odd p values are optimal (Fig. 4,a). When p is even (Fig. 4,d), spurious components occur in the synthesized frequency band, and become unsuppressible by the band pass filter.

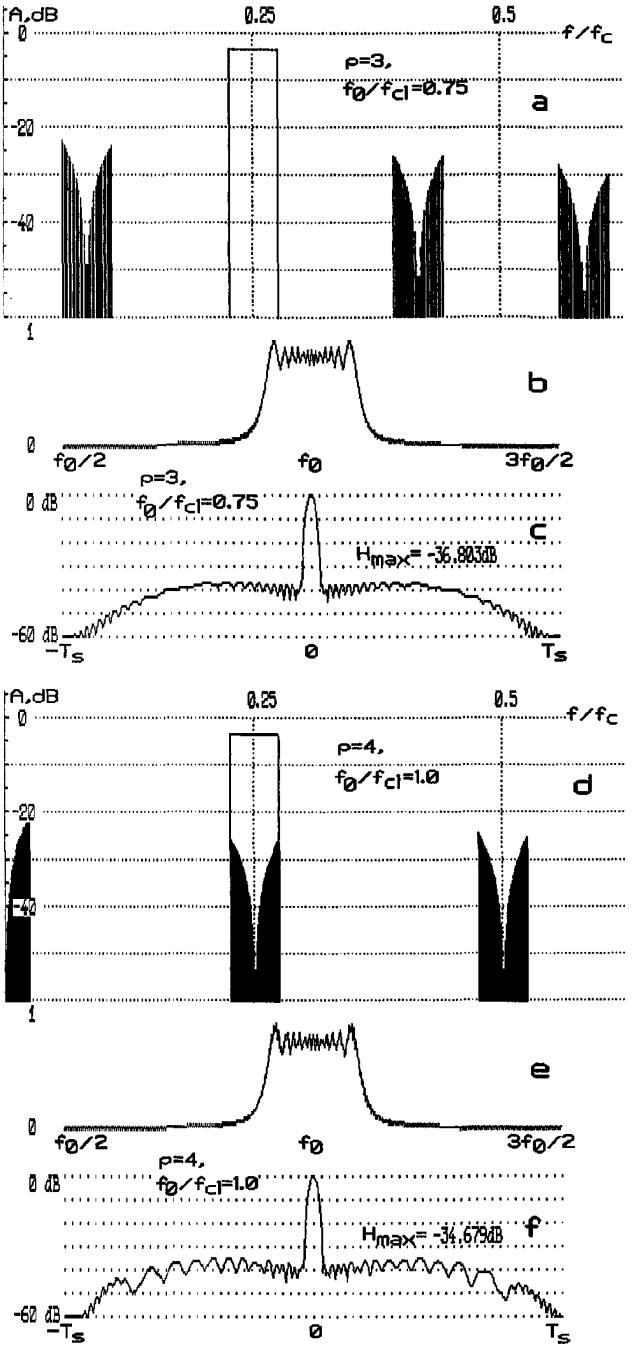


Figure 4. An f_0/f_{cl} effect in DDS SC (W=0,05f_{cl}, N=4, without band pass filter) upon frequency synthesis summar spectrum (a,d), LFM pulse spectrum (b,e) and compressed LFM pulse waveform (c,f): f_0/f_{cl}=0.75, p=3 (a-c); f_0/f_{cl} =1, p=4 (d-f)

In the LFM pulse synthesizing mode that leads to the spurious LFM pulse forming in the same band with the useful LFM pulse has an opposite FM-rate sign (obvious from Fig. 2 nomogram), and lower

magnitude, changing strong in the signal - forming process (Fig. 4,d). This causes the distortions of spectrum shape (Fig. 4,e) and the compressed signal sidelobes pulsation (Fig. 4,f) in comparison with the case when p is odd (Fig. 4,b,c).

At enwidening the synthesized frequencies band (Fig. 5,a,d,g), the neighboring SSC-free bands become narrow, and at $W=0,5f_{cl}$, the SSC-free bands becomes equal to zero. Correspondingly, the filtering conditions become worse.

Though, when DDS SC operates at the LFM pulse synthesizing mode (Fig. 5,b,c,e,f) the spectrum and the compressed signal are practically the same as in the ideal case. When $W>0,5f_{cl}$, the spurious occur in the useful frequency band and can not be filtered.

But such W to f ratio may be acceptable at LFM pulse synthesizing mode (Fig. 5,h,i). The same trend may be noticed in the other odd p cases (Fig. 6, the curves $f_0/f_{cl}=0,25$; 0,75; 1,25; 1,75). Here, the $f_0/f_{cl}=0,25$ curve of the sidelobes level to the LFM pulse normalized frequency $H(W/f_{cl})$ ratio corresponds to the case of $p=1$ (when the DDS SC transforms to the classic multilevel DDS).

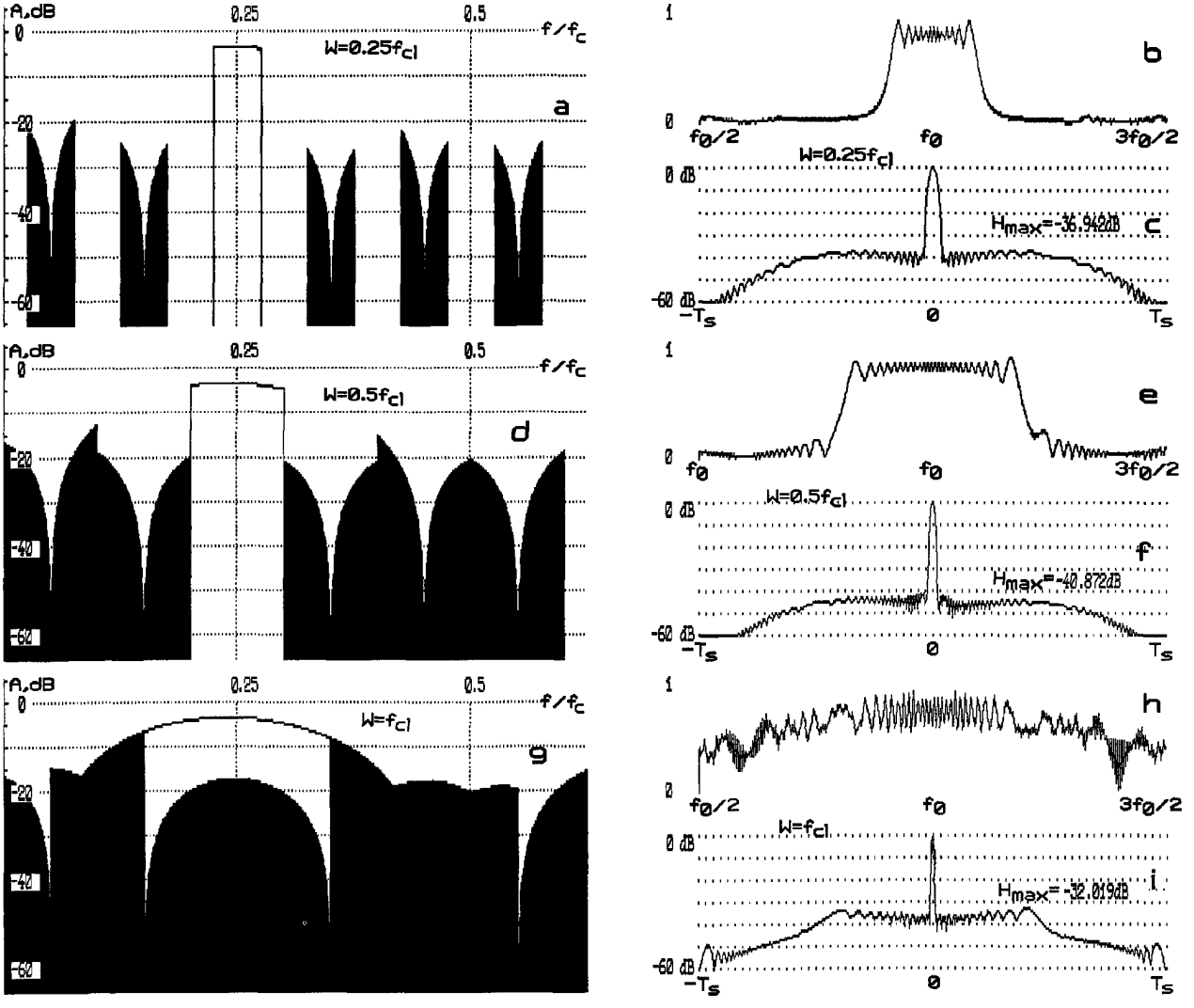


Figure 5. An effect of synthesized signal bandwidth W in DDS SC (N=4, f_0/f_{cl} =1,25, without band pass filter) on frequency synthesis summar spectrum (a,d,g), LFM pulse spectrum (b,e,h) and compressed LFM pulse waveform (c,f,i)

When the frequency deviation is low (up to $0{,}5f_{cl}$), the compressed signal sidelobes level for the even p (Fig. 6, curves f_0/f_{cl}=1; 1,5; 2) is more than for the odd nearby p (Fig. 6, curves f_0/f_{cl} =0.25; 0,75; 1,25; 1,75).

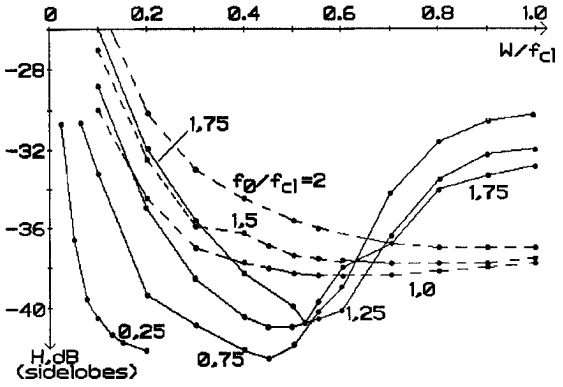


Figure 6. A sidelobe level of compressed LFM pulse, synthesized in DDS SC, versus normalized frequency deviation (for f_0/f_{cl}=0,25;0,75;1,0;1,25;1,5;1,75;2)

The difference may reach 4...5 dB. However, for the great frequency deviation $\left(W > 0{,}5f_{cl}\right)$ the sidelobe level increases with the increasing of W for the odd p, and when $W > (0{,}6...0{,}7)f_{cl}$, the sidelobe level for the odd p is greater than for the even ones. This means that, when DDS SC is used as the LFM pulse synthesizer, the minimal value of compressed signal sidelobe level is provided in the case of $W<0{,}5f_{cl}$ and odd p. If it is necessary to synthesize the maximal deviation signals (up to $W=f_{cl}$), it is preferable to choose the even p.

All the above resumes were made by the assumption, that the sine samples are calculated exactly (at the computer simulation of compressed signal and spectrum there was fixed the FC and the DAC were the 16-bit units: $q_{FC} = q_{DAC} = 16$).

But in real synthesizers the FC and DAC bit number are limited. The FC and DAC bit number in DDS SC effecting the maximal sidelobe level of compressed signal is shown on Fig. 7. For LFM pulse low values of base (for example, WT_s=40), the almost ideal sidelobes level is reached already in 6-Bit FC and DAC (Fig. 7, curves 1 and 3). When the base become greater, the almost ideal sidelobes level is reached in FC and DAC with greater bit number (Fig. 7, curves 2 and 4). The Fig. 7 $H(q_{FC})$ and $H(q_{DAC})$ ratios analysis shows, that it is preferable to choose the FC and DAC bit number so that the relation $q_{FC} - q_{DAC} \geq (1...2)$ becomes true.

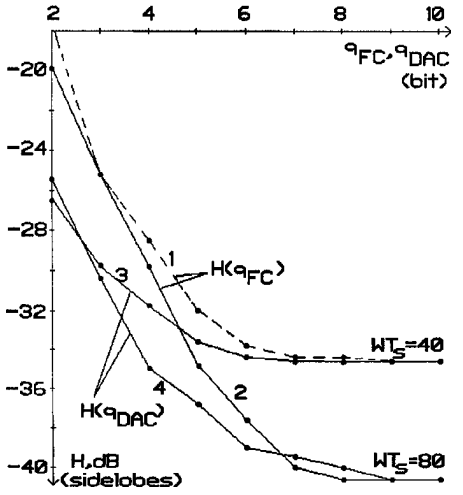


Figure 7. An FC (curves 1, 2) and DAC (curves 3, 4) bit number effect on the maximal sidelobes level of the compressed LFM pulse (for WT_s=40; 80)

This conclusion corresponds to the arithmetic units bit number choice recommendations given in [15] for the classic multilevel DDS.

Direct Digital Synthesizer Based Upon Two Level Digitally Controlled Phase Shifter

The DDS with two level DCPS (Fig. 8) operation principle is based upon the high frequency L-phase system of two-level signals (with pulse/pause ratio 2) forming and the sequential switching of the signals with frequency f_{cl}. The PFU formed low frequency L-phase system is assembled by two-level signals (1) with duration $T_{cl} = 1/f_{cl}$. As in DDS SC the multiplication of l-th signals of high and low frequency systems is provided by a switch and adding of all the L multiplication results transforms to their sequel passing to the output.

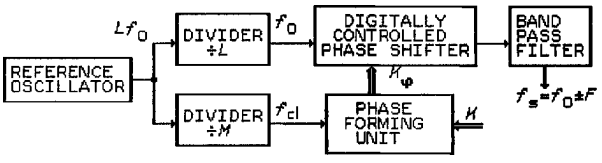


Figure 8. Direct digital synthesizer based upon two level digitally controlled phase shifter

In this case, when $K_F \neq 0$, the phase of DCPS output signal of frequency f_0 is periodically (with the period $T_{cl} = 1/f_{cl}$) incremented. The increments may be positive or negative, dependent on K_F sign. As a result, the output signal phase middle value equals to the required one.

Let us define the output signal spectrum of DDS with two level DCPS operating in the frequency synthesizing mode, assuming that phase accumulator capacity equals to the DCPS phase gradations quantity.

When $K_F = 0$, the initial conditions dependent phase shift $2\pi l/L$ (where l=0,1,2,...,L-1), corresponds to the PFU output constant phase code. In this case the output square wave signal (with the repetition period $T_0 = 1/f_0$) Fourier spectral power representation gives the following result:

$$u(t,l) = \begin{cases} \sum_{k=-\infty}^{\infty} \frac{2}{j\pi k} \exp\left(\frac{j2\pi kl}{L}\right) \exp\left(\frac{j2\pi kt}{T_0}\right) & \text{for odd } k, \\ 0 \text{ for even } k. \end{cases}$$

If $K_F \neq 0$, the phase shift remains constant only in the output two-leveled sequence elementary signal duration (T_{cl}) limits. In this case, the phase shift value $l = (l_C - 1)K_F$ depends on the elementary signal number l_s (l_s=1,2,3,...,L), and on the K_F value. The maximal repetition period of this two-level signal is $T = LT_{cl}$.

The Fourier spectral coefficients of this signal may be written as:

$$C_m = \frac{1}{T} \cdot \sum_{l_s=1}^{L} \int_{(l_s-1)T_{cl}}^{l_s T_{cl}} \sum_{k=-\infty}^{\infty} \frac{2}{j\pi k} \exp\left\{\frac{j2\pi k(l_s-1)K_F}{L}\right\} \times$$

$$\times \exp\left\{\frac{j2\pi kt}{T_0}\right\} \cdot \exp\left\{-\frac{j2\pi mt}{T}\right\} dt.$$

After the integration:

$$C_m = \frac{1}{\pi^2 L} \cdot \sum_{k=-\infty}^{\infty} \frac{1 - \exp\{j2\pi(kf_0/f_{cl} - m/L)\}}{k(kf_0/f_{cl} - m/L)} \times$$

$$\times \sum_{l_C-1}^{L} \exp\left\{\frac{j2\pi(l_s-1)}{L} \cdot \left(kK_F + \frac{kLf_0}{f_{cl}} - m\right)\right\}.$$

Using (5) and representing f_0/f_{cl} as the ratio f_0/f_{cl}=M/L, we obtain:

$$C_m = \frac{1}{\pi^2} \cdot \sum_{k=-\infty}^{\infty} \frac{1 - \exp\{j2\pi(kM-m)/L\}}{k(kM-m)} \times$$

$$\times \frac{1 - \exp\{j2\pi(kK_F + kM - m)\}}{1 - \exp\{j2\pi(kK_F + kM - m)/L\}}.$$

Since $(kK_F + kM - m)$ - an integer value, then C_m differs from zero when

$$m = k(K_F + M) - QL, \qquad (10)$$

where Q=0,±1, ±2 ,±3,... Using the Lopitals rule, we obtain:

$$C_m = \frac{L}{\pi^2} \cdot \sum_{k=-\infty}^{\infty} \frac{1 - \exp\{j2\pi(kM-m)/L\}}{k(kM-m)}. \qquad (11)$$

Let us assume that there exist a pair of values k_0, Q_0 (k_0 - odd, Q_0 - integer), satisfying to the equation (10), i.e.:

$$m = k_0(K_F + M) - Q_0 L. \qquad (12)$$

Assuming that k_0, Q_0 is not an unique solution, let's find the other k, Q being solutions of (10).

From (10) and (12) we obtain, that $k = k_0 + L(Q - Q_0)/(K_F + M)$. For k to be integer and odd, the equation $L(Q - Q_0)/(K_F + M) = 2j$ must be correct. Here j=0,±1,±2,±3,... . In this case $Q - Q_0 = 2j(K_F + M)/L$ - an integer value, what is correct only when j=aL, where a=0,±1,±2,±3,... . So, we can find:

$$\begin{aligned} k &= k_0 + 2al, \\ Q &= Q_0 + 2a(K_F + M). \end{aligned} \qquad (13).$$

So, C_m differs from zero, when (11) is correct, where k and Q are defined by (13), where k_0, Q_0 - any couple of integer values, corresponding to the equation (12); in eddition, k and k_0 are odd. In this case (13) may be substituted into (11), and summing by k may be changed for summing by a:

$$C_m = \frac{L}{\pi^2}\left(1 - \exp\left\{j2\pi \frac{k_0 M - m}{L}\right\}\right) \times$$

$$\times \sum_{a=-\infty}^{\infty} \frac{1}{(k_0 + 2aL)[(k_0 + 2aL)M - m]}$$

Using the relation [14]:

$$\sum_{a=-\infty}^{\infty} \frac{1}{(a+x)(a+y)} = \frac{\pi}{(y-x)}[\mathrm{ctg}(\pi x) - \mathrm{ctg}(\pi y)],$$

we obtain the expression for spectral coefficient:

$$C_m = \begin{cases} \frac{1}{2\pi m} \cdot \frac{\left(1 - \exp\frac{j2\pi(kM-m)}{L}\right) \cdot \sin\frac{\pi m}{2LM}}{\sin\frac{\pi k}{2L} \cdot \sin\frac{\pi(kM-m)}{2LM}} \\ \text{when} \quad m = k(K_F + M) - QL \\ 0 \quad \text{for other} \quad m \end{cases} \qquad (14)$$

and for its module

$$|C_m| = \begin{cases} \left| \frac{1}{\pi m} \cdot \frac{\sin\frac{\pi m}{2LM} \cdot \sin\frac{\pi(kM-m)}{L}}{\sin\frac{\pi k}{2L} \cdot \sin\frac{\pi(kM-m)}{2LM}} \right| \\ \text{for } m = k(K_F + M) - QL \\ 0 \qquad \text{for other } m \end{cases} \qquad (15)$$

where k_0 is substituted for k=±1,±3,±5,...; Q=0,±1,±2, ±3... . The equations (14) and (15) were obtained at the condition, that $f_0 / f_{cl} = M / L$, (M - integer) and the phase accumulator capacity equals to L (to phase shifter gradations quantity).

In fact, the condition $f_0 / f_{cl} = M / L$ means, that the spectrum coefficient expression was obtained for the circuit Fig. 8, where the frequency dividers FD1 and FD2 division factors are equal L and M correspondingly, and where Lf_0 frequency is supplied to FD1 and FD2 inputs. The two level phase shifter, for example may be used in such a circuit. This DCPS contains a shift register (wich output connected to its input through invertor) and a commutator (in this case L must be even). This shift register combines the functions of the FD1 and the phase splitter.

To verify the equation (13) let's change k to k+2al in equations (14) and (15). In this case C_m and $|C_m|$ are not changes.

At Fig. 9 a nomogram was build according to the equation (10) (in the same way as Fig. 2 nomogram was build) for the described synthesizer. Using the nomogram one may find the values of normalized frequencies $f / f_{cl} = m / L$, in which case $|C_m| \neq 0$. The lines ±3,±5,±7,... presence indicates the SSC great content in the two-level PFU output signal spectrum. This does not allow to choose the optimal f_0/f_{cl} ratio using the nomogram (in the case of DDS SC the optimal ratio f_0/f_{cl} choice allow to get the maximal synthesized signal band W when SSC level at the band pass filter output not exceed the desired value).

Solving this problem there were build the maximal SSC level curves $A_{max} = 20\lg(|C_{m\max}|/|C_0|)$, dependent on f_0/f_{cl} (Fig.11) ratio. The curves were build for some W/f_{cl} values. Here $|C_0|$ and $|C_{m\max}|$ - are correspondingly the useful and the maximal SSC on the DDS output. The DDS contains the two level DCPS and the 4-order Butterworth band pass filter with 1,25 dB maximal irregularity.

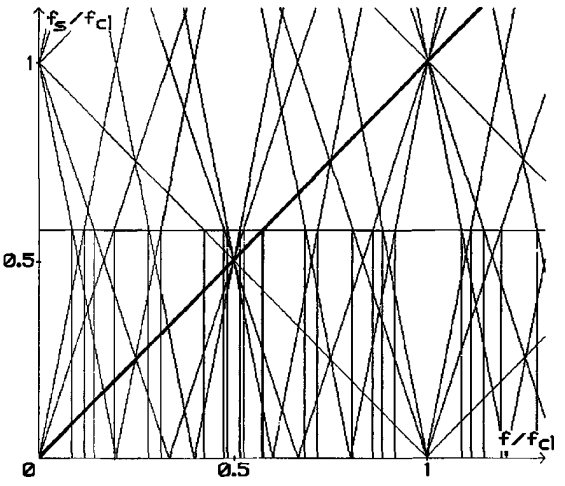


Figure 9. A nomogram that shows the output signal spectrum components positions at the frequency axis in DDS with two level phase shifter

The Fig. 10 curves analysis allows to obtain the empirical ratio

$$f_0/f_{cl}=R/2+1/4 \qquad (16)$$

for the values, for which the $A_{max}(f_0/f_{cl})$ function local minimum exist for any W/f_{cl} ratios. Here R=1,2,3,..., and (16) coincides with (9) for DDS SC.

For the W=0,25f_{cl} the SSC level is about 5dB less in comparison with $f_0/f_{cl}=R/2$ points, and 2dB less in comparison with $f_0/f_{cl}=R/2+1/2$ points. When the bandwidth becoms narrower, the difference in SSC level increases, and at W<0,125f_T it reaches about 18 and 14 dB for the $f_0/f_{cl}=R/2$ and $f_0/f_{cl}=R/2+1/2$ points correspondingly. In addition, there are other local minimums on the $A_{max}(f_0/f_{cl})$ curves when the band becoms narrower. The minimum values are about equal or less than the ones located at the nearest $f_0/f_{cl}=R/2+1/4$ point. However there are no other f_0/f_{cl} values (besides (16)), that could be the minimum points in the case of the great changing of the synthesized signal bandwidth. That makes preferable f_0/f_{cl} choice by the (16) equation, although when the band is narrow (for example, if W=0,03f_T) the variation of f_0/f_{cl} value may decrease the SSC level at about 8 dB or more. The additional SSC level decreasing may be achieved by R value increasing (in assumption that (16) is true). So, the SSC level decreases at about 15 dB, if f_0/f_{cl}=5,25, and at about 25 dB if f_0/f_{cl}=16,25.

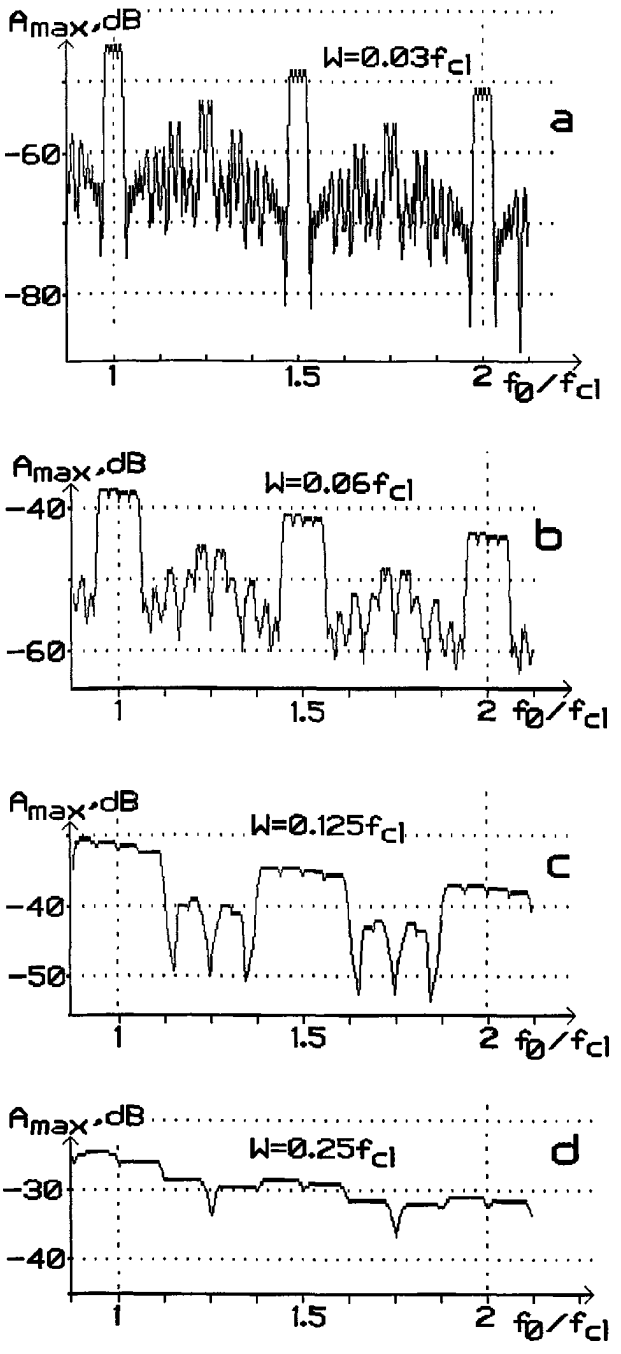


Figure 10. An effect of the ratio f_0/f_{cl} on the inband maximal level of spurious components of signal at the output of DDS with two level phase shifter (frequency synthesizing mode, 4-th order band pass filter): W=0,03f_{cl} (a); W=0,06f_{cl} (b); W=0,125f_{cl} (c); W=0,25f_{cl} (d))

When the DDS with two level DCPS output synthesized frequency band W becomes wider, not only the outband but the inband SSC level becomes greater too (Fig. 11,f,d,g,j). In this case the inband SSC can not be suppressed by the band pass filter. It is the course of the compressed signal sidelobes H increasing in the case of LFM ($W \geq 0,5 f_{cl}$) signals synthesizing (Fig. 11,i,l). When the deviations are less the DDS with two level DCPS output LFM pulse spectrum and compressed signal (Fig. 11,b,c,e,f) are the same as ones of the ideal two level LFM signal.

From the curves at Fig. 10, and from the summar spectrums comparison (Fig. 11,a,d,g,j and Fig. 12,a,d,g,j) it follows, that for all the W band values the f_0/f_{cl} choice in the two level DCPS DDS brings better results, than for example by the ratio

$$f_0/f_{cl}=R/2 \qquad (17)$$

where R=1,2,3,... . However, it is impossible to spread this conclusion to the case of LFM signals synthesizing. It is obvious from the comparison of the spectrums (Fig. 11,b,e,h,k and Fig. 12,b,e,h,k,) and compressed signals (Fig. 11, c,f,i,l and Fig. 12, c,f,i,l). Fig. 11 and Fig. 12 corresponds to (16) and (17) consequently. The curves Fig. 13 (maximal sidelobe level H versus normalized deviation W/f_{cl}) analysis shows the preference of the f_0/f_{cl} ratio choice by the (16) equation, when W<(0,3...0,4)f_{cl}, and by (17) for the greater deviation values.

The output signal spectrum and spectrum module equations (14) and (15) are obtained with assumption, that the phase accumulator capacity is equal to the DCPS phase gradations quantity. In a real frequency DDS the accumulator bit number is defined by the required frequency discretes, and in a DDS with frequency modulation signals - by the required phase to time dependence representation accuracy and may equal 20...32 and more. The phase shifter bit number q_{DCPS} is limited by the reference oscillator frequency value (when using a phase splitter based upon a frequency divider), or by the DCPS manufacturing precision - this bit number, usually, is not greater then 6...8. Due to this, there are disturbances in a synthesized signal, and the disturbances effect the output spectrum and compressed signal form (Fig. 14,a-f). For the Hamming weighting processing and for small LFM pulse base value (WT_s=51,2), already when $q_{DCPS} \geq 6$, the maximal sidelobe level $H(q_{DCPS})$ changes small when the phase shifter bit number increases. When a synthesized signal has greater base value, it is preferable to use 7- or 8-bit DCPS to achieve the minimal possible value of maximal sidelobe level.

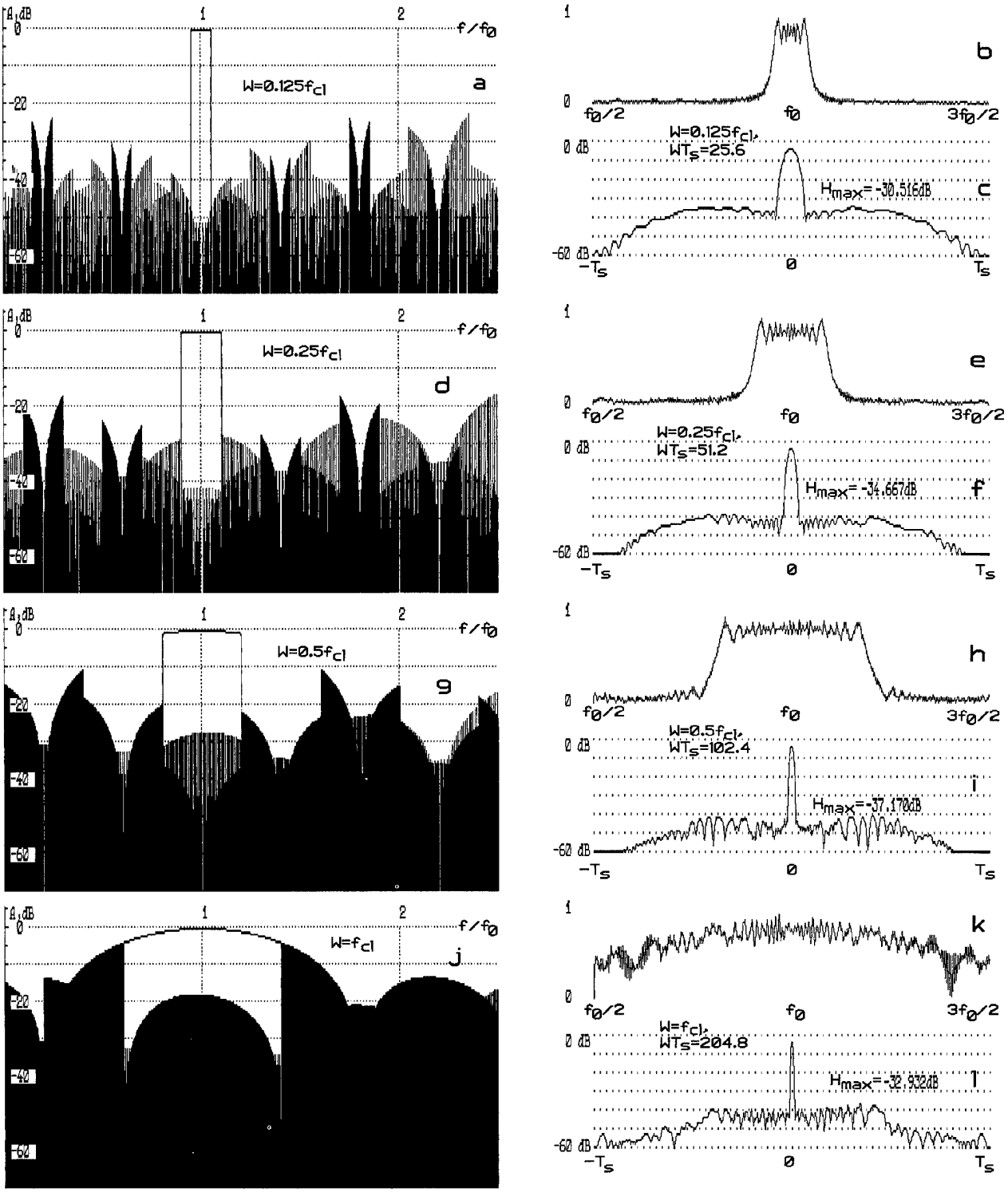


Fig. 11. An effect of synthesized signal bandwidth W in DDS with two level phase shifter (f_0/f_{cl}=1,25, without band pass filter) on frequency synthesis summar spectrum (a,d,g,j), LFM pulse spectrum (b,e,h,k) and compressed LFM pulse waveform (c,f,i,l):
W=0,125f_{cl}=0,1f_0 (a-c); W=0,25f_{cl} =0,2 f_0 (d-f); W=0,5f_{cl} =0,4 f_0 (g-i); W= f_{cl} =0,8f_0 (j-l)

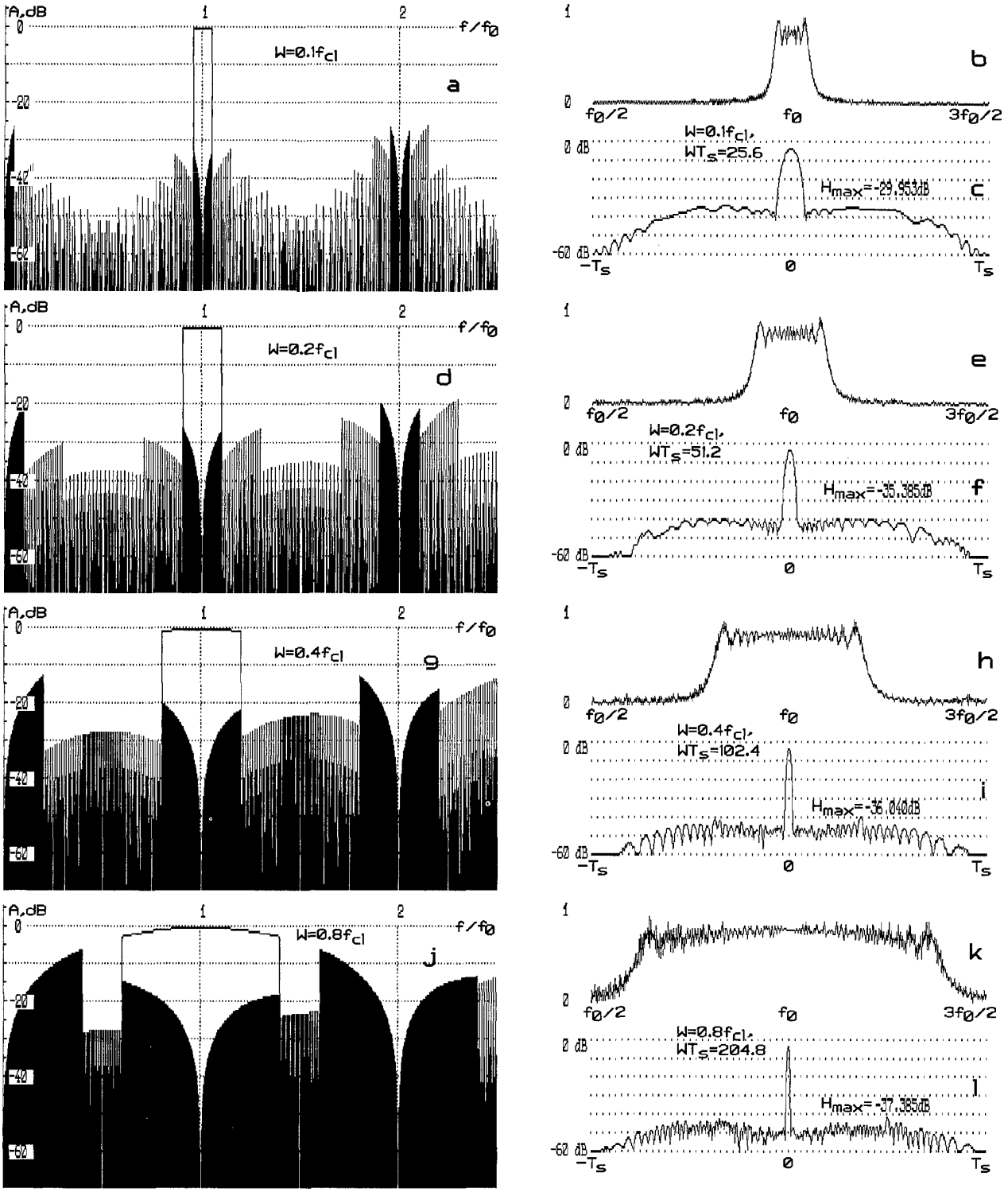


Fig. 12. An effect of synthesized signal bandwidth W in DDS with two level phase shifter (f_0/f_{cl} =1,0, without band pass filter) on frequency synthesis summar spectrum (a,d,g,j), LFM pulse spectrum (b,e,h,k) and compressed LFM pulse waveform (c,f,i,l):
W=0,1f_{cl}=0,1f_0 (a-c); W=0,2 f_{cl} =0,2 f_0 (d-f); W=0,4 f_{cl} =0,4 f_0 (g-i); W=0,8 f_{cl} =0,8 f_0 (j-l)

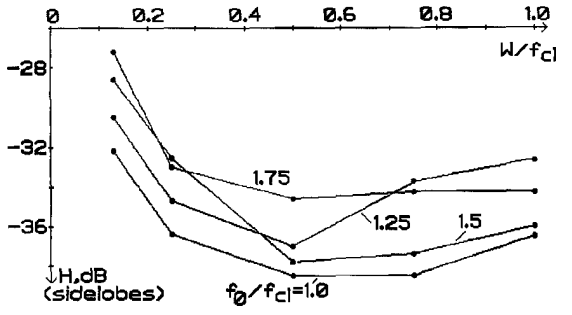


Figure 13. A maximal sidelobe level of compressed LFM pulse, synthesized in DDS with two level phase shifter, versus normalized frequency deviation (for f_0/f_{cl}=1,0; 1,25;1,5;1,75)

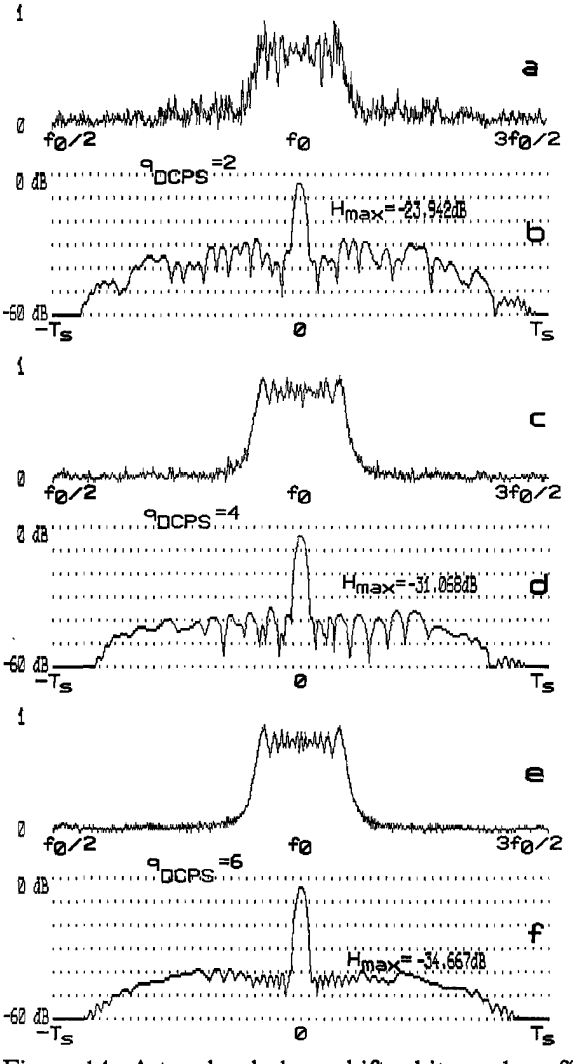


Figure 14. A two level phase shifter bit number effect on the spectrum waveform (a,c,e) and compressed LFM pulse waveform (b,d,f):
q_{DCPS}=2 (a,b); q_{DCPS}=4 (c,d); q_{DCPS}=6 (e,f)

Direct Digital Synthesizer Based Upon Sine Wave Digitally Controlled Phase Shifter

The synthesizer in which sine wave DCPS controlled by the phase code in clock moments (Fig. 8) is examined in [12] for LFM pulse forming mode. This case is mostly realistic, as the phase code is formed at the output of PFU, built upon the base of code accumulators. In such a synthesizer the output signal phase is incremented in each of the clock moments only in the case of $q = q_\varphi$, i.e. the PFU bit number is equal to the DCPS bit number.

Let us derive the equation for the frequency synthesizing mode sine wave DCPS DDS output spectrum, assuming that the DCPS and the PFU bit numbers are equal to each other.

When $K_F = 0$, the initial conditions dependent phase shift $2\pi l/L$ (where l=0,1,2,...,L-1), corresponds to the PFU output constant phase code:

$u(t,l) = \sin(2\pi t / T_0 + 2\pi l / L)$, where $T_0 = 1/f_0$.

In the case of frequency synthesizing ($K_F \neq 0$), the phase shift value remains constant only in limits of one elementary signal duration (T_{cl}) and depends on the signal number l_s (l_s=1,2,3,...,L): l=(l_s-1) K_F. So, the DDS output signal maximal repetition period is T=LT_{cl}.

The Fourier spectral coefficients are:

$$C_{m'} = \frac{2}{T} \cdot \sum_{l_s=1}^{L} \int_{(l_s-1)T_{cl}}^{l_s T_{cl}} \sin\left[\frac{2\pi t}{T_0} + \frac{2\pi l}{L}\right] \exp\left\{-\frac{j2\pi m' t}{T}\right\} dt.$$

Representing f_0 / f_{cl} as $f_0 / f_{cl} = M / L$, after the transformations by analogy to the DDS SC ones, we can obtain: $C_{m'} = C_{m'-1} + C_{m'+1}$, as the simultaneous $C_{m'-1} \neq 0$, $C_{m'+1} \neq 0$ equations being true is impossible. Here :

$$C_{m',s} = -\frac{js}{L} \cdot \frac{\sin\left[\pi(M - m's)/L\right]}{\pi(M - m's)/L} \cdot \exp\left\{\frac{j\pi(M - m's)}{L}\right\} \times$$

$$\times \frac{1 - \exp\{j2\pi s(K_F + M - m's)\}}{1 - \exp\{j2\pi(K_F + M - m's)/L\}},$$

where $s = \pm 1$; $C_{m',s} = C_{m',-1}$, when s=-1 and $C_{m',s} = C_{m',+1}$, when s=1.

Since $(K_F + M - m's)$ is an integer, then $C_{m',s}$ differs from zero only when:

$$m' = sK_F + sM - QL, \qquad (18)$$

where Q=0,±1,±2,±3,... . Using the Lopitals rule one can obtain:

$$C_{m',s} = js\left|C_{m',s}\right| \exp\{-j\pi s(M + m's)/L\},$$

where

$$|C_{m',s}| = \begin{cases} \left|\dfrac{\sin[\pi(M - m's)/L]}{\pi(M - m's)/L}\right| \\ \text{for } m' = sK_F + sM - QL, \\ 0 \quad \text{for other } m', \end{cases} \qquad (19)$$

$s=\pm 1$; $Q=0,\pm 1,\pm 2,\pm 3,\ldots$.

At Fig. 15 shown summar spectrum of output signal of DDS with sine wave phase shifter, obtained in accordance with (19).

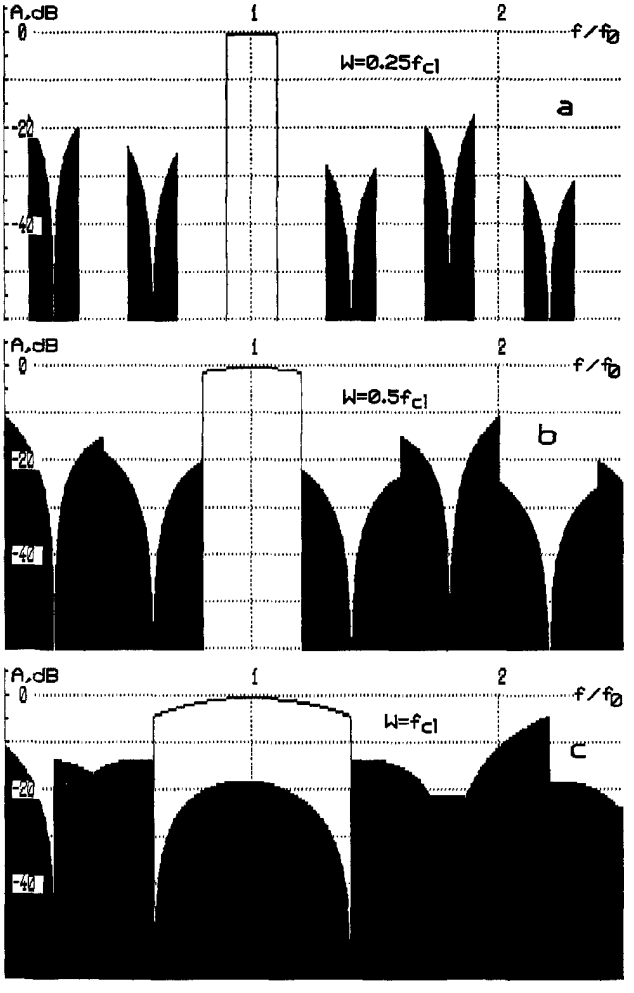


Figure 15. An effect of bandwidth W on output signal summar spectrum in DDS with sine wave phase shifter (frequency synthesizing mode, f_0/f_{cl}=1,25, without band pass filter): W=0,25f_{cl} (a); W=0,5f_{cl} (b); W=f_{cl} (c)

Taking (18) in account, the equation (19) may be represented as:

$$|C_{m',s}| = \left|\frac{\sin[\pi(Q - K_F/L)]}{\pi(Q - K_F/L)}\right|.$$

The (8) and (18) equations being similar leads to the sine wave DCPS synthesizer output spectrum components frequency definition nomogram coinciding with the DDS SC Fig. 2 nomogram at all. It is true, if to take in account, that the $f_s/f_{cl} = m/n$ equation is correct for the second synthesizer, and the $f_s/f_{cl} = m'$, $p/N = f_0/f_{cl} = M/L$ - for the first one. The discussed synthesizers summar spectrums occur to be almost equivalent too (Fig. 5,a,d,g and Fig. 15,a,b,c).

The minimal value of maximal sidelobe level for almost any W/f_{cl} is provided by the f_0/f_{cl} ratio choice in the LFM-signals synthesizing mode (Fig. 16).

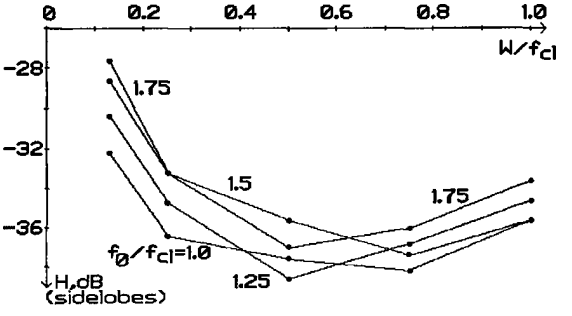


Figure 16. A sidelobe level of compressed LFM pulse, synthesized in DDS with sine wave phase shifter, versus normalized frequency deviation (for f_0/f_{cl}=1,0; 1,25; 1,5; 1,75)

Conclusions

1. As a computer simulations shows, to obtain the compressed signal minimal sidelobes level it is necessary to choose the $W<0,5f_{cl}$ frequency deviation with the odd p, in condition of using the N=4 phase quantity DDS CS as the LFM pulse synthesizer. In this case the useful and spurious spectrum component bands do not overcross, and the compressed signal sidelobes level is as in the ideal case. If $W>(0,6...0,7)f_{cl}$, then it is preferable to choose the even p. In this case the sidelobes level is 4...5 dB grater than that one of the given base ideal signal. Avoiding this recommendation execution may lead to the sidelobes level increasing up to 5...6 dB in each case.

2. Allowing to reach almost the ideal case compressed signal sidelobes level the FC and DAC bit number choice depends on the LFM signal base WT_s (for example $q_{FC} = q_{DAC} \geq 6$ and $q_{FC} = q_{DAC} \geq 8$ when WT_s=40 and WT_s=80). The decreasing of q_{FC} effects the sidelobes level more than the decreasing of q_{DAC}.

3. The two level DCPS DDS nomogram lines ($\pm 3,\pm 5,\pm 7...$) presence indicates the output signal spurious spectrum discrete components great quantity.

The graph of the spurious component level versus normalized conversion frequency f_0/f_{cl} has a number of local minimums, which quantity and positions at the frequency axes depends on the synthesized frequency band W. The $f_0/f_{cl}=R/2+1/4$ points are the exclusions, as there exists the minimum at the points for any band W. The latter is very useful when choosing the band width for great W.

4. When the two level DCPS DDS output synthesized frequencies band W increases, the band-pass filter unsuppressible inband SSC level grows besides the spurious spectrums band increment. This is the reason for the compressed signal sidelobes level grows in the case of LFM signals synthesizing. Though the satisfactory results (sidelobes level not more than -34...-35dB) are being obtained in the synthesized signal deviation changing up to $W=f_{cl}$ value. The f_0/f_{cl} ratio choice by the equation $f_0/f_{cl}=R/2+1/4$ (R=1,2,3,...), when $W<(0,3...0,4)f_{cl}$ (for the two level DCPS DDS), or when $W<0,6f_{cl}$ (for the sine DCPS DDS) allows to obtain some (about 3...4 dB) gain in the compressed signal sidelobes level. For the greater deviation values one must choose $f_0/f_{cl}=R/2$.

5. At the condition, that the signal processing is made by the compression filter, the LFM signal bandwidth being near to the maximal one ($W=f_{cl}$) can be provided by the two level or sine wave DCPS DDS. If the compression filter is not able to provide the ideal compressed LFM signal sidelobes level less than -34...-36 dB, then the two-level or sine DCPS bit number increasing more than 6 does not effect strong (almost at all) the sidelobes level.

6. When the two-level DCPS DDS operates in frequency synthesizing mode, the output signal band W depends on the requirements to the SSC level, being mainly defined by the inband components. So, -50 and -60 dB SSC level is provided in the band of $0,125\,f_{cl}$ and $0,06\,f_{cl}$ correspondingly. The W value for the sine DCPS DDS must be chosen in the same way as for the DDS SC, that is $W\leq(0,25...0,3)\,f_{cl}$ for the filtering conditions provision. So, the two level DCPS DDS may be used for the narrow band frequency synthesizing cases (Doppler synthesizers, frequency shift devices, narrow-band FM synthesizers and the others).

7. When the synthesized signal band is narrow, the substitution of the classic two-level DDS [6] for the searched structure based on two-level DCPS allows to decrease the equipment spending and the power consumption, and to increase the maximal synthesized frequency.

References

[1] Cooke C.E., Bernfield M. Radar Signals. An Introduction toTheory and Application, Academic Press, 1967.

[2] Кочемасов В.Н., Белов Л.А., Оконешников В.С. Формирование сигналов с линейной частотной модуляцией. -М.: Радио и связь, 1983.

[3] Кочемасов В.Н., Долбня Е.В., Соболь Н.В. Акустоэлектронные фурье-процессоры /Под ред. В.Н.Кочемасова - М.: Радио и связь, 1987.

[4] Tierney J., Rader C.M., Gold B. "A Digital Frequency Synthesizer", IEEE Transactions on Audio Electroacoust., AU-19, March 1971, p.43-57.

[5] Котов В.С. "Синтезаторы частоты, основаниые на сложении импульсных последовательностей" //Радиотехника. - 1971. - Т.16, № 5. - С. 64-68.

[6] Кочемасов В.Н., Фадеев А.Н. "Цифровые вычислительные синтезаторы двухуровневых сигналов с компенсацией фазовых ошибок" Радиотехника. - 1982. - Т.37, №10. - С. 15-19.

[7] Раков И.А., Кочемасов В.Н. "Точиость представления данных в арифметических блоках цифровых вычислительных синтезаторов сигиалов", Изв. вузов. Радиоэлектроника. - 1987. Т.30, № 12. -с.49-55.

[8] Reinhardt V., Gold K., McNab K., Bustamante M. "A Short Survey of Frequency Synthesizer Techniques", 40th Annual Frequency Control Symposium, May 1986.

[9] Кочемасов В.Н., Раков И.А. "Цифровой вычислительный синтезатор на основе фазовращателя с коммутацией отсчетов", Электросвязь. - 1988. - № 2. - С. 56-60.

[10] Vettori G. "Digital Generation of a Linear FM Pulse of Given Time-Bandwidth Product", Alta Frequenza. -1971. -Vol. XL, № 2. -P.140(48E)-146(54E).

[11] Pat. 3792378 (USA), cl. 331-178. "Digitally Controlled RF Sweep generator". A.S.Hughes, S.A.Taylor. Published 12.02.74.

[12] Iglehart S.C. "Some Results on Digital Chirp" IEEE Trans. on Aerospace and Electronic Systems: 1978. - Vol. AES-14. - № 1. -P. 118-127.

[13] Оконешников В.С., Кочемасов В.Н. "Сжатие частотномодулированных сигналов с небольшим произведением девиации частоты на длительность импульса". Зарубежная радиоэлектроника. - 1987. -№1. -С. 82-94.

[14] Прудников А.П., Брычков Ю.А., Маричев О.И. Интегралы и ряды. -М.: Наука, 1981.-800 с.

[15] Nicholas H.T., Samueli H. "An Analysis of the Output Spectrum of Direct Digital Frequency Synthesisers in the Presence of Phase-Accumulator Truncation" 41st Annual Frequency Control Symp. - 1987. - P. 495-502.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A HIGH-SPEED PHASE-FREQUENCY DETECTOR ASIC WITH PROGRAMMABLE DIVIDERS FOR LO GENERATOR AND INDIRECT SYNTHESIS APPLICATIONS

Robert Howald and Thomas Marra
Lockheed-Martin Astro-Space
230 Mall Blvd., King of Prussia, Pa. 19406

Abstract

A novel device has been developed by Lockheed Martin Astro-Space (LMAS) for its Local Oscillator (LO) Generator and Frequency Synthesizer product lines. The ASIC is a digital, quad-D type, complementary output, phase/frequency detector with built-in programmable digital dividers on both the reference and RF inputs. The device has been used as a universal element of both direct divide and mix-down phase-locked loop (PLL) topologies. Applications requiring basic indirect synthesizers have developed into virtually off-the-shelf designs. The ASIC is currently implemented both in flight frequency generators and synthesizers.

Introduction

Microwave frequency synthesis techniques for military and space applications have evolved considerably in the past ten to fifteen years. Fixed frequency LO's commonly used direct multiplication of a high stability crystal oscillator reference, or multiplication of a low noise oscillator at some higher frequency phase-locked to a crystal oscillator reference. In order to take advantage of digital circuits within the PLL in the latter approach, the circuitry required was limited by the highest frequency ECL logic available. The maturity of GaAs devices and GaAs integrated circuits have spurred development of simple, microwave fundamental oscillators, such as microstrip oscillators and dielectric resonator oscillators (DRO's). Additionally, high speed logic has allowed designers to incorporate digital topologies into PLL circuits at even higher frequencies. The evolution of design techniques for true synthesizer applications is more drastic, and has been driven primarily by the development of the direct digital synthesizer (DDS). The DDS has become the cornerstone of any application which requires simultaneously fine resolution, fast switching speed, and low noise performance. In short, the DDS is ideal for fast-hopping, spread spectrum synthesizer applications. However, it can be overkill for applications requiring moderate switching, basic channelization selection, or a simple fixed, but programmable frequency source. The device described here serves very well the needs of both modern fixed LO generators, and basic indirect synthesizer schemes.

ASIC Description

Device Description

The phase detector ASIC uses a high speed, quad-D type, digital phase detector for frequency and phase comparison. The quad-D phase detector is an ideal choice for a digital phase detector since it does not have a 'dead zone' and provides rapid lock times even for large frequency offsets. The phase detector ASIC also has two high speed frequency dividers capable of dividing by any integer from one to 1023. One frequency divider is used on the reference input and the other is used on the VCO feedback input. The reference frequency divider can be bypassed by selecting a control input. In this mode, an alternate input of the ASIC is used for the reference frequency input.

The phase detector ASIC also contains a coarse phase detector circuit which provides extended pull-in for frequencies which have large offsets from the reference. The range at which the coarse phase detector operates is determined by a two bit input which is the coarse threshold ratio (CRSTHR[1:0]). This two bit input signal selects a ratio of 2, 4, 8, or 16 for coarse detector operation. A ratio of 2 results in coarse detector

0-7803-2500-1/95 $4.00 © 1995 IEEE

operation when the two frequencies at the quad-D phase detector differ by a factor of 2 to 1. Similarly, for a ratio of 4, 8, or 16 the frequency difference would be a factor of 4 to 1, 8 to 1, or 16 to 1 respectively. A frequency lock circuit was also implemented in the ASIC. This frequency lock circuit indicates when the two inputs to the quad-D phase detector are within 0.4% of each other.

The phase detector is capable of directly comparing or dividing all frequencies up to its maximum input. There are three frequency inputs to the ASIC. They are FREF, FIN, and FREFDIR. Each of these three frequency inputs is capable of converting a 0.25 Vp-p sine wave input into an appropriate CMOS level signal. This is accomplished using a high speed amplifiers on each of these inputs. The high speed amplifiers on the frequency inputs allow a multitude of input frequency signals to be used, such as ECL, TTL, CMOS, and analog.

The phase detector ASIC typically draws 20 mA at 5 volts. Measured maximum current draw is 30 mA at 5 volts, or 150 mW. The ASIC is implemented using the Harris AVLSI CMOS process. This process has a total dose capability of 1 Mega-RAD, making it ideal for space applications. It is packaged in a 64 pin ceramic flatpack. The dimensions of this package are 0.92 inches square including leads.

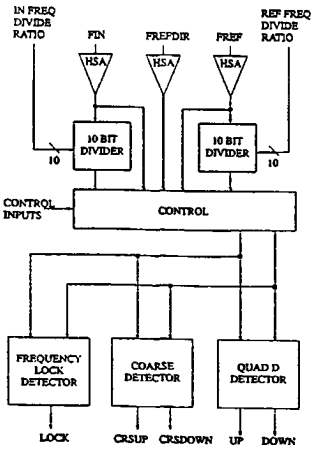


Figure 1

The output of the control block goes to the quad-D phase detector. The quad-D phase detector generates the UP and DOWN error signals which are used by the charge pump to control phase locked operation. Similarly, the coarse detector circuit generates the CRSUP and CRSDOWN outputs. The third and final output block is the frequency lock detector which indicates when the frequencies going into the quad-D phase detector are within 0.4% of each other.

Block Diagram

Figure 1 is the block diagram of the phase detector ASIC. The three frequency inputs can be seen in the block diagram. These inputs are FIN, FREF, and FREFDIR. The frequency inputs go into the ASIC through the High Speed Amplifiers (HSA). Two of the frequency inputs, FIN and FREF, go from the HSA's into the 10 bit dividers. The 10 bit dividers are by-passable and are designed to operate at high speeds. The output of the frequency dividers and the high speed amplifiers go into a control block which switches these signals to the appropriate input of the quad-D phase/frequency detector. This level of control allows for either a positive or negative slope VCO to be used with the phase detector without rewiring its outputs. In addition, the control section is used to bypass one or both frequency dividers if a divide ratio of one is required for one or both of the frequency inputs.

I/O Functional Definition

INPUTS (27 pins)

DIVREF[9:0]	The binary number that determines the divide ratio for the reference frequency divider.
DIVIN[9:0}	The binary number that indicates the divide ratio for the feedback frequency divider.
FREF	The reference input frequency.
FIN	The feedback input frequency.
RESET	The control signal that initializes the divide counters to their appropriate dived ratios as defined by DIVREF[9:0] and DIVIN[9:0].

CRSTHR[1:0]	Determines threshold value for the coarse pump circuit. Selects the ratio of the frequency difference between the quad-D phase detector input frequencies that must be achieved before the coarse pump circuit is activated.
PULSEL	The select input used to determine the mode for the pulse width of the quad-D phase detection circuit. In one mode, the pulse width will be minimum. In the other mode, the pulse width is settable via an external capacitor.
HL	High-side/Low-side control signal for quad-D phase detector input which allows either a positive or negative slope VCO to be utilized by switching the high-side and low-side inputs input the quad-D circuit.
FREFDIR	The reference input frequency when the reference frequency divider is bypassed.
FREFSEL	The control signal used to bypass the FREF frequency divider.

OUTPUTS (7 pins)

UP	Output signal to VCO charge pump.
DOWN	Output signal to VCO charge pump.
DOWNN	Complement of DOWN output signal.
PUL	Output loading on this pin determines the minimum pulse width on UP and DOWN when PULSEL is not selected for minimum pulse width.
LOCK	Indicates when the inputs to the quad-D phase detector are in frequency lock +/- 0.4%.
CRSUP	Coarse phase detector output signal.
CRSDOWN	Coarse phase detector output signal.
OFFMIX	Offset mixer output signal.
FDIVOUT	Reference frequency divider output signal.

Measured Electrical Characteristics

Input Power: As previously described, the ASIC has a high speed squaring amplifier on each of its signal inputs. The squaring amplifier design acts to generate output logic levels compatible with the internal ASIC functional logic. The circuit is a bipolar design with a high input impedance. This makes it easily compatible with logic inputs of all levels, and with AC coupled RF inputs when the coupling capacitor is preceded by a 50 ohm shunt resistor to ground. The simplest interface for all inputs is to simply AC couple into the phase detector, allowing the squaring amplifier to provide the gain, if necessary, to drive its internal logic. Additionally, the limiting action of the amplifier eliminates AM spurious components.

Component hardware designs have successfully implemented the phase detector with various input types, the most common being ECL, HCMOS, and RF inputs. It follows from the inherent high frequency nature of the chip that it has been applied most commonly in circuits which use high frequency topologies. Minimum RF input drive is a significant parameter to characterize, particularly for high-rel applications. It is often critical to implement an ultra-high stability reference, such as an oven-controlled crystal oscillator (OCXO). These SC-cut units are typically sine wave outputs, thereby providing a very low noise floor. Outputs in the 7 dBm to 13 dBm range are common. When these units must drive multiple PLL's in an LO generator scheme, and must provide the redundancy required of high-rel designs, the crystal oscillator power can quickly be split to enough modules to become a problem, even with buffer amplifiers provided. Furthermore, some excellent PLL IC's, which use other forms of high speed logic, provide low-level RF outputs. The programmable NEC UPB587 is an example of such a device.

The circuit design of the input amplifier was geared towards an input range of -3 dBm to +3 dBm, nominal. Test results show a minimum input RF level, for a mid-range 64 MHz input frequency, of about -10 dBm at room temperature. Since it is presumed that the input circuit is operating in its linear range for this low level input, performance margin over temperature implies

about a -8 dBm minimum input power for this input frequency. This agrees with gain performance described in the device description section. This input range is sensitive to good RF mounting techniques of the IC at higher frequencies, and does degrade beyond the target design frequency of 100 MHz, as will be discussed in the following section. The high speed select pin should also be implemented above 50 MHz.

Maximum Input Frequency: The phase detector design had a minimum input frequency goal of 100 MHz, and therefore had some built-in design margin. (i.e. circuit simulations were run at a frequency above 100 MHz). Current LMAS frequency generator topologies implement discrete crystal oscillator designs utilizing overtones near 130 MHz. This makes stretching the phase detector input frequency very attractive. For designs using digital prescalars, every additional division by two increases the phase noise at offsets where phase detector logic is dominant by 6 dB. Or, if the phase detector remains in a divide-by-one mode, the last prescalar noise floor will increase by 6 dB. Additionally, the phase noise contributed by the reference itself is likely to be better for a non-multiplied 130 MHz unit compared to a lower frequency crystal multiplied to 130 MHz. Therefore, tests were undertaken to determine the maximum usable input frequency of the phase detector.

To provide the most stringent phase detector operating conditions, the ASIC was run with each side in the divide-by-one mode. On the reference input side, this can be accomplished by either programming a one on the logic inputs, or by properly setting a select bit on the chip which bypasses the dividers. The select bit is called "frefsel", and when excited high, the input signal is routed to "frefdir" (pin 4) instead of "fref" (pin 2). For these tests, the input signal was routed to the former. By using the ASIC without the divider functions, the phase detection operation itself is performed at the frequency of the reference, once frequency pull-in occurs. As a final note, the tests were performed with the ASIC in a clamp down test fixture. The fixture pins were soldered to a standard .0625" G10 printed circuit board with microstrip 50 ohm lines for RF traces. Thus, performance represents a conservative measure of what could be expected when the ASIC (a "flatpack") is solder mounted in an actual circuit application, using good RF grounding techniques.

The PLL implemented used two Hewlett-Packard (HP) synthesizers. An HP8642 connected in the FM mode was used as the VCO, driven by the loop amplifier to provide a fixed VCO gain in Hz/V. The nominal input power of 0 dBm was used for both signal inputs. With the "pulsel" pin on the phase detector set to low speed operation (by which an external capacitor can be used to set the pulse width at phase-lock), the ASIC operated trouble-free through 80 MHz. The maximum specified frequency using the low speed setting is 50 MHz. The circuit's ability to lock became intermittent at turn-on when the frequency approached 100 MHz. When the "pulsel" was toggled to the high speed mode, operation was flawless through 120 MHz. Above 120 MHz, performance began to drop off. To maintain lock, the input power was increased, primarily on the reference input side. At 130 MHz, an input power of +10 dBm was required into a 50 ohm terminated "fref" port to allow phase lock. Without the terminating resistor, an input power near 0 dBm was still adequate to maintain lock. However, in an application where the device is driven by a 50 ohm RF amplifier, this may not be a reasonable solution. An isolator, or a higher power driver amplifier followed by an attenuator to minimize the mismatch may be required for proper operation of the RF amplifier. Without the terminating resistor, whose voltage divider action resulted in too much signal level loss, phase-lock was achieved up to 135 MHz reliably. The PLL would not lock under any circumstances for a 137 MHz input frequency. Thus, 135 MHz represents the upper limit of input frequency.

Phase Noise: A final key performance criteria is the ASIC logic noise floor. In general, PLL's which use digital prescalars are limited in the ultimate noise floor they can achieve relative to their all-analog PLL counterparts. Thus, while digital prescalars represent a simple implementation, and while digital phase/frequency detection provides several advantages over analog phase detection, there is a price to pay in phase noise performance inside the PLL loop bandwidth. Because of this noise floor, the loop bandwidth generally will be narrower in digital schemes, since the phase noise crossover will occur at a lower offset frequency. This results in the phase noise roll-off from the voltage-controlled oscillator (VCO) beginning at a closer, noisier offset for the VCO, degrading performance.

Because this ASIC has built-in dividers, its own logic floor becomes more significant. This is because when the dividers are implemented, such as in a frequency synthesis application, they will decrease the phase noise contribution at the output due the noise accumulated

before the divider. Thus, for a total divide ration of Nmax = N1(first divider)•N2(ph. detector), the noise floor of the first divider will be increased relative to its contribution at the output by 20•log(Nmax/N2), whereas the phase detector floor itself will be increased by 20•log(Nmax). Therefore, for applications where the phase detector's dividers are used, its own noise floor characteristics often determine the logic noise floor performance inside the PLL loop bandwidth.

For ease of measurement, the phase detector noise floor was measured at a 5 MHz comparison frequency by building a simple PLL whose only logic was the ASIC itself. The reference is a high stability, low noise, ovenized unit, whose phase noise was measured for loop characterization. A 100 MHz VCO implemented with an HP8642 was used to provide a significant divide ratio with which the logic noise floor would be increased high enough to be measured accurately (26 dB). A wide loop bandwidth was used to provide noise data for further offsets. A plot of the measured characteristics are shown in Figure 2. From this plot, the following represents a close estimate of the noise characteristics of the ASIC for a 5 MHz comparison, and a 100 MHz input frequency:

Offset	dbc/Hz
10 Hz	< -120
100 Hz	-132
1000 Hz	-135
10 kHz	-135
100 kHz	-135

The performance at 10 Hz is dominated by the reference, so the less than -120 dBc value represents that the ASIC noise could not exceed this value without showing a significant contribution at the overall output (i.e. for the most part, the reference masked the phase detector performance at this offset). The remaining numbers are very close to actuals. It is obvious that the device noise floor is about -135 dBc/Hz.

Performance degrades for higher frequencies of phase comparison, as determined by higher frequency PLL's with direct phase detection (divide-by-one) and other known component noise characteristics. Measured data indicates that for phase comparison at about 65 MHz, the logic noise increases by about 5 dB for the offset range around 1 kHz. This is likely a nearly across-the-board increase. However, for offsets of 100 Hz and less, the logic noise is not typically the most significant contributor.

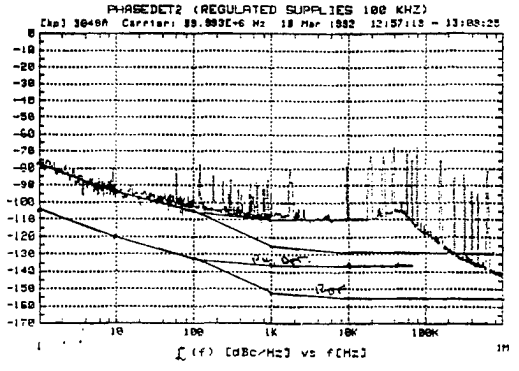


Figure 2

Applications

LO Generators

Direct Divide PLL's: Direct divide topologies used in some current X-Band LO designs for remote sensing satellite applications are shown in Figure 3 and Figure 4.

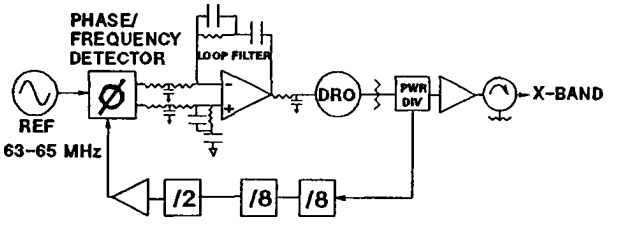


Figure 3

The ASIC is ideally suited to direct divide PLL schemes, because it can be programmed to any integer divide value between one and 1023. This partially alleviates constraints set by available divide ratios of standard prescalars, and simplifies topologies by avoiding the need for dual modulus schemes or custom divider designs implemented using several shift register IC's. Its high input frequency capability is useful in allowing a direct divide microwave oscillator the

flexibility to easily select a divide scheme which results in an input to the phase detector that falls within the input bandwidth. The bandwidth also represents a range over which high stability references (i.e. crystal oscillators) can be directly phase compared to. Because direct divide schemes imply digital prescaling, the logic compatibility of the design also becomes valuable, as does its ability to obtain fast lock-up time at turn on and reset, due to its frequency detect capability.

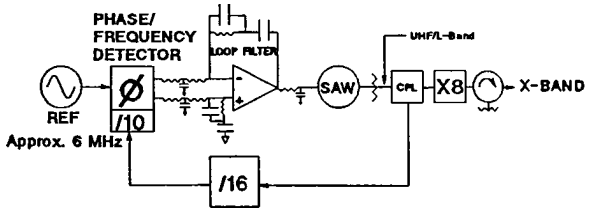


Figure 4

The wide bandwidth and large divide range also make the ASIC an ideal cornerstone for a local oscillator family for which direct divide performance is adequate, as it is for many robust modulations used in satellite communications. The broad bandwidth of the accompanying microwave prescalars, and the ability to program the phase detector divide ratios, allows multiple LO generator designs to take on the same topology with just minor changes. This can be done by simply hardwiring the programming bits for the proper divide ratio. Its compatibility with multiple input signal types minimizes differences among designs, which become easily compatible with sources varying from external, sine wave OCXO's, to on-board logic clocks.

Mix-Down PLL's: LO topologies are not always so easily implemented using direct divide techniques. There are several applications where alternate approaches, such as a mix-down, must be used. The mix-down PLL also has advantages for synthesizer applications, which will be discussed in the next section. One obvious problem for direct divide approaches is the unavoidable degradation of phase noise with divide ratio. The lower the total divide ratio can be made, the better phase noise will be. Furthermore, many applications require an LO to be coherent to a single 5 MHz or 10 MHz high stability standard. Thus, divide ratio constraints which were partially mitigated by a high frequency, self-selected frequency standard, is once again a problem. Adding to the potential troubles is that the desired output frequency may not be easily related numerically to this frequency standard. Clearly, the availability of all divide ratios in the phase detector will be a huge benefit here, providing flexibility in the implementation of divide scheme. However, microwave prescalars which may precede the ASIC are basically limited to powers of two ratios, limiting the possible overall divide ratios. While dual-modulus prescalars can be used to provide other ratios, with input frequencies exceeding 1 GHz, they often consume significant power (ECL) at RF frequencies, and add complexity to the implementation. Some inconvenient LO frequency requirements, without use of other fancy techniques, would require phase comparisons at very low frequencies, exacerbating the divide ratio problem. For such cases, a mix-down PLL approach offers another avenue of flexibility for implementing a low-noise microwave oscillator. The availability of the full range of integer divide ratios in the phase detector, combined with some numerical analysis, makes this problem a straightforward one to solve.

Many applications require more than a single LO, and often many LO's. While it is prudent to maintain design similarity amongst the units, it is sometimes costly to implement all designs with X-Band parts and technologies. Instead, one conveniently chosen unit can be used as a mixing reference for the remaining units. At the mixer output, these other units need only implement common RF hardware. Another important advantage is that it allows the designer to potentially implement one single very low noise PLL, and allow it to dominate the phase noise characteristics of the remaining LO's. Such a topology designed for a satellite payload is shown in Figure 5. In these mixed units, the remaining PLL circuitry after the mixer operates at lower frequencies, eliminating at least one, and usually two prescalars. The design in Figure 5 used a dual loop approach for the single low noise LO. This was fed to each of the other units to lower the overall loop divide ratio for those PLL's, thereby limiting the logic noise floor effects. The phase detector's ability to provide the integer divisions makes the single mix-down reference approach straightforward to implement. It would otherwise be more difficult to determine an LO frequency that would provide a numerically convenient mixer output to lock to the fixed satellite reference.

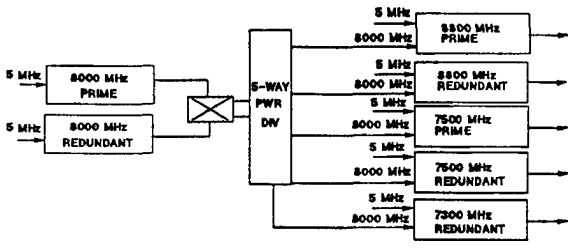


Figure 5

Indirect Synthesis

PLL Synthesis: A direct divide PLL used as a synthesizer achieves output frequency changes by varying the programmable divider values. The built in dividers on this ASIC obviously can be used in this type of application. Direct divide, indirect synthesis is most useful using this ASIC for synthesizers in the UHF-type of range or less, where minimal prescaling is required to get within the frequency range of the phase detector. This is because, as the synthesizer output frequency increases, more prescaling done in front of the phase detector will result in a larger output step size without correspondingly dividing the reference. This lowers the comparison frequency, and, correspondingly, the output frequency. It makes it difficult to achieve any significant synthesizer bandwidth.

For higher frequency synthesizers, where direct divide is an option with regard to other important synthesizer parameters, LMAS has instead a 2 GHz GaAs programmable divide-by-N. This device was also developed within the company for synthesizer applications. It, too, has integer divide capability for every value between 56 and 2047, and can be jointly implemented with the phase detector ASIC if necessary. It must be kept in mind that indirect synthesis has the unfortunate characteristic of varying PLL parameters as divider values change. This often requires additional circuitry within the PLL to compensate by changing loop gain as necessary to maintain noise and switching performance [1].

Mix-down PLL topologies, similar to those described above, are quite common in synthesizer schemes where moderate resolution is required, but not so much as to justify a DDS. A mix-down topology for an X-Band flight downconverter application based on an L-Band synthesizer is shown in Figure 6.

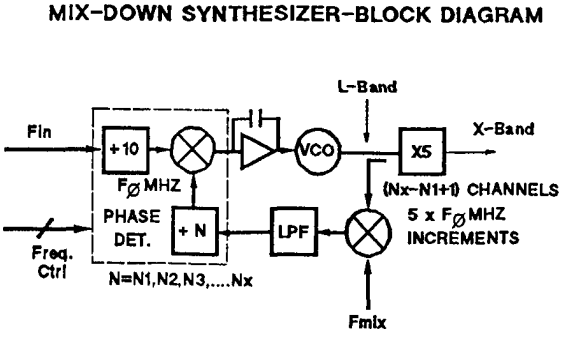


Figure 6

The mix-down approach allows the designer to use a relatively low phase comparison frequency without absorbing a high divide ratio and associated noise floor increase due to logic noise multiplication in the PLL. A low noise mix down signal, such as can be obtained through an analog multiplication or PLL, can mix the synthesizer output to a lower frequency, and limit the total division in the loop. The equation for the output frequency of this approach, when the mixing signal is lower in frequency than the synthesized output, is

$$Fout = (N/R) \cdot Fin + Fmix,$$

where

N = PLL feedback loop total divide ratio
R = Reference input divide ratio.

This equation also indicates the previously described advantage of the mix-down approach with respect to finding convenient numerical relationships. That is, the ability to add (or subtract) in the PLL presents a degree of freedom in the search for a topology that results in a rich set of factorials easily implemented with available or easily implemented divider circuits.

Switching Performance:

As previously described, the ASIC has a quad-D configuration for phase detection, a type of 3-state phase detector. A dual-D configuration represents the simplest implementation of a 3-state phase detector. It is outperformed by the quad-D configuration for frequency acquisition purposes when the frequency separation is very large. When the difference is less than a factor of two, the acquisition performance varies little. Thus, as a conservative estimate, and for many practical applications, the quad-D configuration is closely related to the dual-D 3-state phase detector for analysis purposes. The quad-D and dual-D are identical in the phase detection mode, and have the same gain and phase detector characteristic.

The exclusive-OR (EXOR) and two-state phase detectors represent the first types of simple logic phase detectors that were widely used. The three common phase detector types (sinusoidal, sawtooth, two-state) rely on the DC voltage associated with the asymmetry of the waveform that occurs when the unlocked, modulated VCO output is phase compared to a reference. The asymmetry of the voltage waveform results in it having a DC component that charges the integrator capacitor, eventually pulling the VCO on frequency. However, the AC waveform is not an efficient tuning control input, as its average voltage is relatively small. Indeed, in a low SNR or in poorly design loops, offsets and other noise voltages can overcome this pulling capability if the noise is of the same order of magnitude as this average. (As an aside unrelated to frequency acquisition, two and three-state phase detectors are generally troublesome in high-noise situations themselves because of their inherent memory, which allows the effect of a noise impulse to linger.)

The three-state topology, which most simply can be considered a two-state detector with reset control via the flip-flop outputs, provides the additional capability of frequency detection. This occurs in the dual-D configuration because the duty cycle of the output pulse waveform is proportional to the difference in input frequencies. When filtered, this results in a more meaningful DC component to drive the VCO control line. The result can be frequency acquisition times two to three orders of magnitude better than the sinusoidal phase detector, depending on loop bandwidth and frequency offset. A couple orders of magnitude improvement is also obtainable relative to the EXOR phase detector.

The quad-D differs from the dual-D in that it enables the maximum pulling voltage to remain on the output line when the comparison frequencies differ by a factor of two by employing additional storage elements (two more FF's, and thus a "quad-D"). The dual-D, 3-state device described above outputs a duty cycle difference, which is averaged. This output, in fact, becomes asymptotic and less information-bearing for VCO outputs that exceed twice the input frequency (see [2],[3]). The quad-D provides a pure DC output on either the up or down coarse output for the duration of the time that the two times threshold is exceeded. This is more efficient than having a voltage proportional to the frequency difference. Below this factor of two there is little difference. There are, again, issues associated with maintaining loop parameters over broad divide ratios [1], but these will not be addressed here.

The above descriptions make no use of the advantages of the additional coarse tuning outputs available from the ASIC. Since these outputs are separate pins, they allow the designer the ability to slew the VCO more quickly than is dictated by the loop time constants. With circuitry to sum or switch this voltage into the control line directly, as a type of pretune scheme, an additional level of flexibility is achieved. Careful design practice is a must here, as applications using this topology inherently have the broadest band VCO's (unless several switched VCO's of more modest bandwidth are implemented), which are also the most sensitive to garbage on the control line.

A simple example shows the switching advantage of the 3-state design, using the acquisition time solutions derived in [2]. This example is similar to a flight L-band synthesizer design such as that shown in Figure 6. Assume a coverage of 40 MHz of bandwidth with a 1 MHz comparison frequency. With the relatively wideband VCO used here, a 50 kHz loop 3-dB bandwidth is roughly associated with the optimum phase noise crossover point given the mixing signal logic noise contributions in this topology. Assume a 1500 MHz mixing signal, and a desired output covering 1530-1571 MHz. Therefore, the value of N varies from N=30 to N=71, maximum. Switching from the high end to the low end (worse comparison frequency difference) means instantaneously changing the programmable divide from N=71 to N=30 while the VCO output is 1571 MHz. This results in a phase comparison frequency of 2.67 MHz, being compared against 1 MHz. The frequency acquisition time for the

four types of phase detectors described, assuming loop damping of one, when switched to N=30, is:

Detector Type	Acquisition Time (msec)
Sinusoidal	35.83
EXOR	22.38
Two-State	5.40
Three-State	0.21

Lockheed Martin Space Products

Lockheed Martin (LM) is one of the leading aerospace technology company in the world today. This is due in large part to the successes of its individual aerospace components which have produced quality hardware for many years. This paper describes another in a long line of technological advances that has been the trademark of the company. The ASIC described here provides a cornerstone for a unified set of LO generator and frequency synthesizer topologies that helps reduce the number of costly, unique designs. Virtually off-the-shelf, proven topologies help to minimize schedule and technical risk associated with the design task.

The phase detector has allowed the company to carve out LO generator product lines in common frequency bands including phase-locked voltage controlled crystal oscillator sources and phase-locked SAW oscillator sources from 30 MHz up to about 1 GHz. A family of phase-locked DRO's beginning at S-Band and extending up through Ku-Band has also been built around this device by taking advantage of its uninterrupted range of programming divide ratios, and its high maximum input frequency.

Similarly, in the moderate resolution frequency synthesizer arena, LM has taken advantage of the phase detector's inherent programmability to greatly simplify design complexity. Recent systems including L-Band, C-Band, and X-Band synthesizers have been designed around the integer programmability of the phase detector.

References

[1] Howald, R., "Introduction to Analog and Direct Digital Frequency Synthesis", RF Design, January 1995.

[2] Wolaver, D., Phase-Locked Loop Circuit Design, Prentice Hall, Englewood Cliffs, NJ 07632.

[3] Sharpe, C.A., "A 3-state phase detector can improve your next PLL design", EDN, September 1976.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

GPS/LORAN IN AN URBAN ENVIRONMENT-OSCILLATOR STABILITY CONSIDERATIONS

Benjamin Peterson, Richard J. Hartnett, and Geffrey Ottman
Department of Engineering
U. S. Coast Guard Academy, New London, CT 06320
Lance Miller
Science Applications International Corporation, 4001 N. Fairfax Dr., Arlington, VA 22203

Abstract

During the summer of 1994, the Advanced Research Projects Agency and the United States Coast Guard Academy collected extensive radionavigation data in the New York City area. The purpose was to determine the feasibility of using existing radionavigation systems for tracking applications in an urban environment. The data includes the accuracy and availability of the Global Positioning System (GPS) and LORAN, a comparison of electric field (whip) and magnetic field (loop) antennas at both LORAN and Differential GPS frequencies, and the increased availability of both GPS and LORAN by adding a precise clock input to the receivers. The data was collected among the narrow streets and tall buildings in the Wall Street area, among the tall buildings but wider streets of Third Avenue, in the relatively more open streets and smaller buildings of the Bronx, in the vicinity of the large metallic structure of the George Washington Bridge, and under the cover of foliage along the New Jersey side of the Hudson River.

Of particular interest in this paper is the study of how the availability of radionavigation fixes can be substantially enhanced in these urban areas by the integration of a precise clock with LORAN and GPS navigation information. More specifically, a clock allows a geographic fix to be obtained from as few as two Time of Arrival (TOA) measurements. In the case of LORAN in NYC, Seneca and Nantucket are 10-12 dB stronger than Carolina Beach, hence a precise clock provides a significantly higher fix availability since only two stations (plus the clock) are used in obtaining the two TOA fix. Similar improvements in fix availability are achieved in the integrated LORAN/GPS and the GPS modes by using a precise clock. Preliminary results of implementing a two satellite fix using an external Cesium reference and a GPS Builder Kit manufactured by GEC Plessey are presented.

A secondary advantage of using a precise clock occurs in receiver reacquisition after signal loss in the urban environment. Quite simply, in the case of GPS, one can tolerate longer outages without having the GPS receiver reenter the "search" mode, since a precise oscillator can stay within one chip (one usec) for a longer period of time (like durations corresponding to brief signal outages due to shielding). Should the oscillator drift beyond one chip, acquisition is still quicker since one need only search over a handful of Doppler bins when using a precise clock. By having two GPS Builder Kits side by side, each looking at the same signals, with one referenced to a Cesium and one referenced to its own clock, we examine acquisition and reacquisition improvements in a very controlled way for performance comparisons.

Introduction

In an earlier paper [1], the results of a comprehensive study of GPS and LORAN fix availabilities in an urban environment were presented. While the basic conclusions reached in that study will be repeated here for completeness, the reader is referred to [1] for more details. In that paper the data analysis was divided into two general categories: (1) analysis of fix availability for a number of types of fixes, and (2) analysis of the repeatable accuracy for the LORAN fixes.

In the analysis of fix availability, we took (binary) data provided by the GPS and LOCUS LORAN receivers (with E-field antenna) which documented whether or not they were tracking a particular satellite or LORAN station. Recognizing that an H-field antenna can offer superior performance at LORAN frequencies in urban environments, we included an additional element in the LORAN availability which was based on measured SNR from an HP89410A Vector Signal Analyzer connected to an H-field crossed loop antenna. We used known azimuths and elevations of the satellites and/or azimuths of the LORAN stations to calculate PDOP's for a number of fix categories, and compared these PDOP's to a threshold.

In performing the analysis we made several assumptions. Integrated GPS/LORAN fixes assume a fully integrated receiver (not implemented in this experiment) where the offset between LORAN and GPS time plus any bias (due to receiver front end, etc.) has been determined and stored when redundant information was available. Second, we treat a LORAN TOA as being equivalent to a GPS

0-7803-2500-1/95 $4.00 © 1995 IEEE

pseudorange. Finally, in those fix categories which make use of a precise clock, we assume that any clock bias has been determined in an open area with redundant signals and stored. Tables 1 and 2 summarize the fix availabilities of GPS and LORAN.

	PDOP < 3	PDOP < 6	No limit
Wall St. 3D	0.4%	1.0%	6.9%
Wall St. 2D	8.0%	11.2%	21.8%
Wall St. 2D w/clk	11.8%	34.9%	57.1%
Bronx 3D	43.0%	71.1%	81.1%
Bronx 2D	76.5%	85.2%	89.3%
Bronx 2D w/clk	87.1%	91.1%	97.4%
Third Ave. 3D	1.3%	33.1%	55.3%
Third Ave. 2D	47.0%	60.0%	84.2%
Third Ave. 2D w/clk	65.4%	77.3%	96.0%
GW Bridge 3D	59.3%	83.3%	88.7%
GW Bridge 2D	89.8%	92.1%	93.5%
GW Bridge 2D w/clk	92.8%	94.7%	95.0%
Foliage 3D	3.9%	9.5%	16.5%
Foliage 2D	39.5%	44.2%	53.2%
Foliage 2D w/clk	53.1%	77.5%	82.4%

Table 1. GPS Fix availabilities.

	# sta	E-field	H-field
Wall St.	2	4.1%	93.3%
	3	0.7%	53.7%
Bronx	2	95.6%	99.2%
	3	90.0%	85.8%
Third Ave.	2	75.8%	99.4%
	3	64.5%	81.6%
GW Bridge	2	97.4%	99.7%
	3	96.6%	95.7%
Foliage	2	99.4%	100.0%
	3	90.3%	98.5%

Table 2. LORAN Fix availabilities.

From our work in [1] as summarized in Tables 1 and 2, we make the following conclusions:

(1) While the LORAN E-field signal in deep urban canyons is virtually unusable, the H-field signal is considerably stronger and exhibits repeatable phase.
(2) The availability of LORAN H-field signals in the urban canyons and either LORAN E- or H-field signals under foliage is better than GPS signal availability.
(3) In many areas (and particularly on Wall St.), only LORAN signals from Seneca and Nantucket (and not Carolina Beach) can be received, so fix availability is considerably enhanced by the addition of a precise frequency reference, enabling one to obtain fixes from only two TOA's.
(4) The addition of a precise frequency reference to a GPS receiver and the integration of LORAN and GPS also enhance fix availability.
(5) A twelve channel all-in-view GPS receiver performs significantly better than a six channel receiver.
(6) Tests in receiving LORAN signals from an H-field antenna in the Wall Street scenario in December of 1994 suggest that it is possible to select the correct LORAN pulse zero crossing virtually all the time if a heading sensor is used, and almost all the time if no external sensor is employed.

Integrated LORAN/GPS Receivers

The concept of integrated LORAN/GPS receivers is not new [2-4], but to date no truly integrated receivers have been marketed. In current LORAN/GPS receivers, "integrated" really means combining a GPS position with a position derived from intersecting hyperbolic LORAN lines of position. A truly integrated receiver, on the other hand, would measure TOA's of all available LORAN and GPS signals relative to the receiver oscillator, and solve a full (presumably over-determined) matrix solution. Such true integration takes advantage of the fact that LORAN, while not strictly synchronized to GPS time, is (by virtue of having three Cesium time standards at each transmitting station) very stable relative to GPS for extended periods of time. Unfortunately the offset between LORAN and GPS time is in general non-zero, and must be explicitly estimated in a location where redundant information is available, thus truly integrated fixes are not available until that occurs. In Northern Europe, LORAN is being expanded and the time of transmission of all stations will be tightly synchronized to UTC. The Northern European system may provide the impetus to develop integrated LORAN/GPS receivers for urban environments.

GPS Receivers in Urban Environments

After completion of the 1994 New York data collection phase [1], we decided that we would focus future efforts on GPS receiver issues, since the future of U.S. LORAN was rather uncertain. Based on the data from [1] summarized in Table I, a few areas looked promising, and virtually all were impacted by either clock accuracy or stability issues.

Clearly the implementation of a GPS receiver that would produce two dimensional fixes with only two satellites would improve availability. Such an application demands a clock whose stability over the period during which only two satellites are available is such that the

clock error is less than or equal to the acceptable pseudorange error. Since our goal is accuracy comparable to SPS with SA on, this clock error is on the order of 100 ns. Fixes with two satellites for tens of minutes would be possible with 10^{-10} stability, and 10^{-11} stability would permit such fixes for several hours. Other issues of importance in urban environments (such as acquisition and reacquisition of satellite signals) relate to clock accuracy and stability, but do not need nearly as demanding specifications.

The issue of clock stability in GPS receivers has been addressed by several other groups [8-10]. When [8] was written in 1983, plans were for an 18 satellite constellation to be operational in 1988. This 18 satellite constellation would have had extensive outage periods for conventional 3D (4 satellite) navigation, and the author analyzed accuracy as function of outage duration and clock stability. Reference [9] contains analytic solutions for the steady state optimal estimate (Kalman filter) covariance and in particular looks at the improved performance in situations with bad geometry. It is well known that because the satellites used (by receivers on the earth's surface) all must lie within a visible "cone" beginning at some (elevation mask) angle above the horizon, there is high correlation between altitude errors and clock errors. In [10] the authors find that by using very stable clocks, the altitude error on aircraft precision approaches can be substantially reduced.

To study clock issues in GPS receivers we use a modified GPS Builder™ Kit from GEC Plessey Semiconductors (Figure 1). With this kit, the researcher has control of all aspects of the receiver, from the acquisition to the final position processing, because 'C' source code is provided as part of the receiver software. For purposes of this work, we modified the receiver hardware to allow for an external frequency reference.

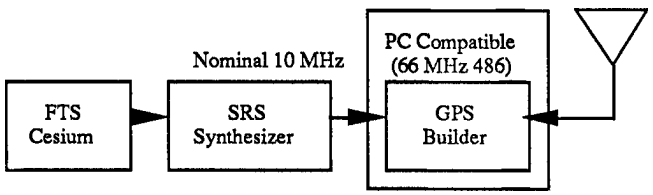


Figure 1. Block diagram of GPS Builder Kit system.

Acquisition Times for Receiver

The first consideration of navigation in an urban environment is that the receiver antenna obtains only short duration, periodic "looks" at satellites of interest. To improve receiver performance, satellite signals have to be quickly acquired, thereby allowing information to be used to obtain a position solution The first step is to examine how the receiver acquires each satellite for use.

The receiver code (correlation) tracking loop attempts to match the C/A code replica generated by the receiver's correlator to the C/A code that modulates the satellite's signal. Each satellite modulates its signal by a unique pseudorandom code, so in order to decipher the satellite information its code must be matched perfectly. This process takes place in two stages, namely "search" and "track." The search is accomplished by the correlator generating its code at a slightly higher rate than the satellite's. The receiver then performs a search of the Doppler bin in which the signal is predicted to be, and then it searches the surrounding bins. When the signal is found and the two signals match up the signal power from the correlator increases. The local C/A code generator adjusts itself to stop the "sliding" relative to the satellite's C/A code, and an early-late tracking loop synchronizes the correlator's code replica with the satellite's [11].

The second step is the phase lock. The phase lock synchronizes the receiver's internal oscillator to the phase of the satellite's signal. A bit sync or bit edge detection is next performed. This matches the signal with its phase shifts to determine each bit. 300 of these bits are then assembled into a subframe. The receiver performs various checks on the preamble and parity to determine the beginning and integrity of each subframe. Finally the frame sync is achieved and the subframe parser begins to decode the information contained in the GPS message.

An actual example of this process for one satellite is shown in Figure 2. Each stage is represented on the vertical axis as an increasing value, with the maximum value, five, corresponding to the satellite signal being usable in the solution.

The effects of using an external Cesium oscillator were examined to determine if this process could be improved. In the initial acquisition stage the receiver estimates Doppler based on a stored almanac and assumed position, and then searches in Doppler bins spaced 500 Hz (or 3.2×10^{-7}) apart, starting with the assumption of a perfect clock. Since the standard TCXO provided with the GPS Builder™ was within this tolerance, the standard receiver performed as well as the Cesium in the initial tests. To illustrate the effects of the clock on acquisition, several tests were conducted with the Cesium referenced synthesizer operating at different offsets from 10 MHz. As the offset increases the receiver has to search through more Doppler for the satellite signal. This causes the position fix to be delayed considerably. These results are summarized below.

Offset	Time to 3-D Fix
0 Hz	~45 seconds
±20 Hz	~1.5 minutes
±30 Hz	~2 minutes
±40 Hz	~3 minutes
±50 Hz	No fix obtained

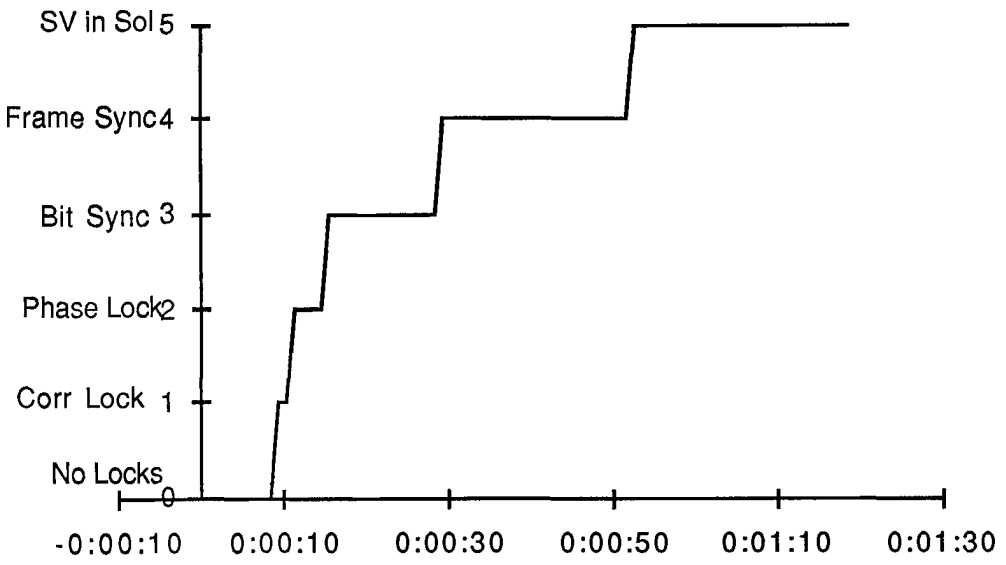


Figure 2: Typical Acquisition Stages versus time (hr:min:sec)

The pattern of increasing time to fix as a function of the oscillator error is evident. It is cut off at ±50 Hz due to the maximum oscillator error of 4 ppm (parts per million) software constraint being exceeded. From the data above one can recognize that a more precise clock provides a quicker usable fix.

Coast Time Improvements

Coasting allows the receiver to tolerate short intervals of signal loss. By propagating the satellite's tracking state forward, the time before a loss-of-lock condition is declared can be controlled. This enables the receiver to lock onto the satellite's signal immediately when it returns to view [11]. During this time the receiver will determine if the satellite is still a valid addition to the position solution. With high buildings blocking large sections of the horizon in urban environments, the only feasible time to receive satellite information is during the short times available at street intersections. If the receiver can maintain these signals (i.e. "coast") from city block to city block, it becomes possible to have continuous coverage by the satellite.

Because the receiver propagates the correlation and phase lock forward to allow for coasting, a stable clock is important. Although the receiver can coast for minutes, if upon the signal's return to view the locks aren't synced, the locks will be lost, necessitating a full reacquisition. In that case all solutions during the coast time using this satellite become invalid due to error.

The effects of a precise oscillator on coasting ability were examined by comparing the performance of the standard receiver with the performance of a receiver with an external Cesium reference. In addition, the software was modified to extend the allowed coast time from 5 seconds to 20 seconds. Blockages were simulated by masking sections of the horizon with a large metal sheet. Blockages of 3, 7, 17, and 23 seconds were conducted to the east, west, and south. The plot of satellite state versus time into the scenario shows the effects of the Cesium reference on the coast time (Figure 3).

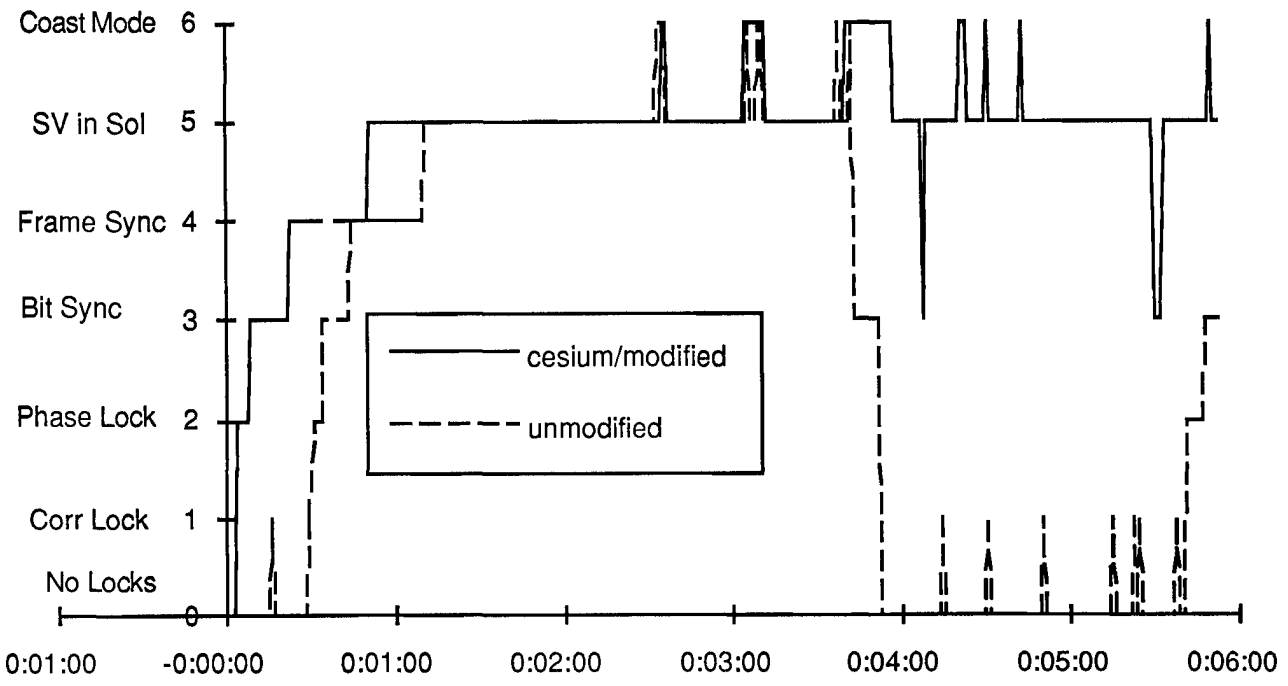


Figure 3. Satellite State versus time (hr:min:sec), Blockages of 3, 7, 17, and 23 seconds.

Observe from Figure 3 that both receivers entered the coast mode as expected for the shorter blockages at 2.5 and 3 minutes into the scenario, and immediately returned to a usable signal when the satellite returned to view. When the 17 second blockage occurred at 3.5 minutes into the scenario, the standard receiver lost all locks after a short coast period. It then attempted to reacquire the satellite but was unable to do so due to the later blockage which occurred at 4.5 minutes. Note that it takes about two minutes before the satellite is fully recovered and usable for the solution. The Cesium enhanced receiver enters the coast mode for the duration of the 17 second blockage. When the satellite is returned to view, the frame sync has been lost, however it is almost immediately recovered, and the signal is usable once again. In this case of using a precise reference, the consequences of a 17 second signal loss are very minor, and would hold no ill effects on the position fixing.

Two Dimensional Fixes from Two Satellites

The solution for a three satellite, two dimensional (2-D) fix has long been implemented by receivers. This is accomplished by constraining the altitude to some known value. For the GPS Builder's position solution, this entails using the known radius of the earth to act as a pseudorange from an imaginary satellite at the center of the earth.

For a two satellite fix, the receiver is assumed to already be in the 2-D mode due to imposing an altitude constraint. With an extremely accurate clock, the clock drift can be assumed to be zero. This eliminates the time correction component of the solution and leaves three unknowns. With the two satellites and the third augmented satellite, position can again be fixed. If the clock is in error the effects should be readily apparent as the position walks along a straight line bisecting the azimuths of the satellites used (Figure 4). When the clock moves ahead of true GPS time the pseudoranges of the satellites will be increased at each solution increment. This will cause the position to move away from the satellites and the true position. The same holds true for slower clocks, except the position will move towards the satellites as the pseudoranges are decreased (Figure 4).

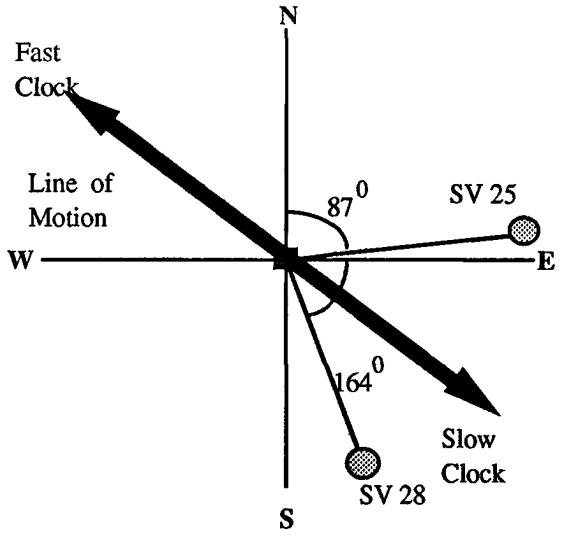


Figure 4. Two Satellite Position - Effects of Clock Error.

To test this, the receiver was operated with modified software (to allow for a 2 satellite fix), but with the standard TCXO frequency reference. Satellites were de-selected until two satellites (SV 25 and SV 28) remained. The results of this test are shown in Figure 5. Note the similarity between predicted (Figure 4) and actual (Figure 5) results.

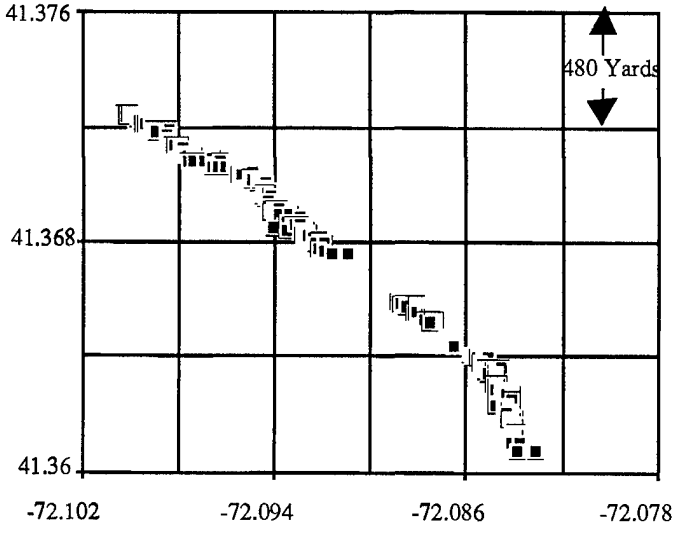


Figure 5. Movement of two satellite fix over 27 minutes, standard (TCXO) oscillator.

The final test for determining the accuracy of a two satellite fix was the application of the Cesium reference with the oscillator error entered at zero. The same process of de-selecting satellites was followed leaving two satellites separated by 78°. When plotted over 40 minutes, Figure 6, the receiver using the two satellites and the Cesium reference exhibits accuracy comparable to that of SPS (100 meters, 2DRMS).

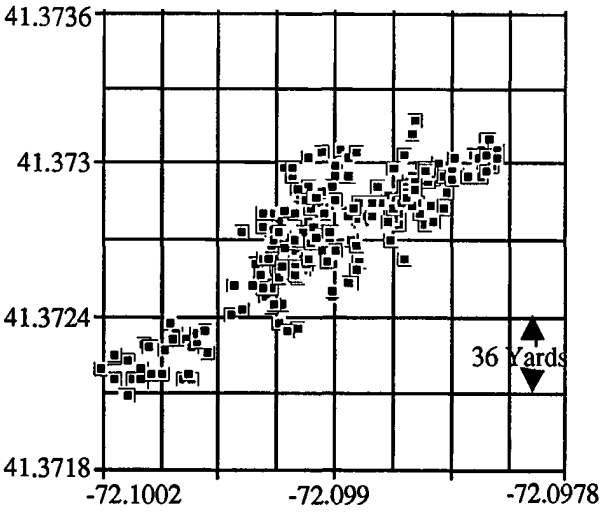


Figure 6. Movement of two satellite fix, Cesium reference.

Conclusions and Future Directions

Improving GPS fix availability in an urban environment was examined with three specific objectives: improving satellite acquisition times, extending the time period the receiver can coast, and obtaining position fixes with only two satellites. All three were shown to be powerful extensions of the standard receiver when used with a precise time reference such as a Cesium oscillator.

Future goals will be the continuing analysis in all three areas. With two receivers operating at the same time under the same conditions, acquisition time and coasting effects can be better examined and improved. The further expansion of the code via a Kalman filter will enable the receiver to use signals acquired at various times to perform an electronic "running fix". In addition, in an extension of the FFT techniques presented in [12], we will look at position solutions based on block processing of 1 msec long time records once satellite ephemeris and clock offset data have been obtained and stored. This will allow for GPS navigation with only very brief glimpses of satellites and greatly improve GPS availability in urban environments.

Acknowledgements

The New York City (1994) phase of this project was sponsored by the Advanced Research Projects Agency under project CO74. The efforts of Prof. Michael McKaughan and LT Steven Bartlett, the loan of the loop antenna by

Megapulse, Inc, and the generous support of the Ferryboat Maintenance Facility at the Coast Guard Base on Governors Island are sincerely appreciated.

The GPS Builder Kit (1995) phase of this project was sponsored by the USCG Navigation Center. We would like to thank Dr. Ian Kitching from GEC Plessey who offered us helpful suggestions regarding hardware modifications of their product to accommodate a precise reference.

References

[1] Benjamin Peterson, Michael McKaughan, Lance Miller and Steven Bartlett, "Evaluation of Radionavigation Systems in an Urban Environment," Condensed version in Proceedings of Institute of Navigation National Technical Meeting, Anaheim, CA, January 1995. Long version available as NTIS report number PB95-129391.

[2] G. LaChapelle and B. Townsend, "GPS/LORAN-C: An Effective System Mix for Vehicular Navigation in Mountainous Areas," , Navigation, Vol 40, No. 1, Spring 1993.

[3] P. Braisted, R. Eschenbach, and A. Tiwari, Combining LORAN and GPS-The Best of Both Worlds, GLOBAL POSITIONING SYSTEM, Vol. III, pp. 235-240, Institute of Navigation , 1986.

[4] K. L. Van Dyke, "The Status and Future of GPS/LORAN Interoperability," Proceedings of the 19th Annual Wild Goose Association Technical Symposium, pp. 123-130, Long Beach, CA, October, 1990.

[5] "Airport and and Airway Safety and Capacity Expansion Act of 1987, " Public Law 100-223, Section 310, Congress of the U. S., 1987.

[6] Specification of the LORAN-C Signal, COMDTINST M16562.4A, Commandant (G-NRN), USCG, May 1994.

[7] 1994 Federal Radionavigation Plan, U. S. Departments of Defense and Transportation, NTIS Report DOT-VNTSC-RSPA-95-1/DOD-4650.5, 1995.

[8] M. A. Sturza, "GPS Navigation Using Three Satellites and a Precise Clock," GLOBAL POSITIONING SYSTEM, Vol. II, pp. 122-132, Institute of Navigation , 1984.

[9] E. M. Copps, "An Aspect of the Role of the Clock in a GPS Receiver," GLOBAL POSITIONING SYSTEM, Vol. III, pp. 44-53, Institute of Navigation , 1986.

[10] P. N. Misra, and R. L. Ferranti, Clock Aided Satellite Navigation for Precision Approaches, presented at IEEE International Frequency Control Symposium, San Francisco, June 1995.

[11] GPS Builder™ Designer's Guide, GEC Plessey Semiconductors, GPS Group, Swindon, Wiltshire, UK, 1994.

[12] Benjamin B. Peterson, Eric A. Chamberlin , Terry R. Vogler, Peter Daly and Stuart Riley, "GPS/GLONASS Time Difference Measurements and Test Bed Development," Proceedings of Institute of Navigation GPS-94, Salt Lake City, September, 1994, also NTIS report number PB95-130019.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PREDICTION OF THE TIME ACCURACY AND INTEGRITY OF GPS TIMING

G. Jeffrey Geier, T. Michael King, Howard L. Kennedy
Russell D. Thomas, Brett R. McNamara

Motorola Position and Navigation Systems Business
8201 E. McDowell Road, Scottsdale AZ 85252

ABSTRACT - **The application of Receiver Autonomous Integrity Monitoring (RAIM) algorithms internal to the GPS sensor have added an extra level of system reliability to timing applications that cannot tolerate system outages. This paper addresses the time integrity issue from GPS, introduces an algorithm that has been implemented in a low cost GPS core receiver, and provides analysis of the algorithm performance under unusual satellite conditions.**

The algorithm detects and removes from the solution those satellites that exceed a pre-determined time residual threshold. Scientific error simulations have been used to characterize the detection and isolation performance of the algorithm, and are reported on. The threshold selection is designed to support a detection and isolation probability of at least 99.99%, and a maximum false alarm rate no greater than 1 false alarm every 5.7 days.

A multi-channel GPS satellite simulator was used to generate typical GPS satellite outages, and the performance of the algorithm was tested and results are reported. Accuracy of GPS with and without the T-RAIM algorithm is reported as well as the receiver's ability to protect the 1PPS time reference output pulse from the influence of the failed satellite.

One of the by-products of the T-RAIM algorithm is its ability to predict the time error performance of the receiver under the influence of SA with and without failed satellites. The paper reports on the algorithm's ability to predict the overall system time accuracy for possible exploitation by the end user.

I. INTRODUCTION

The use of GPS based timing in the wireless telecommunications industry has exploded in recent years. As the availability of radio frequency spectrum in the most desirable parts of the band become more scarce, and the commercial value for that spectrum ever increases, system providers look to new technologies to maximizing the use of the available spectrum. With TDMA (Time Division Multiple Access), many more communications channels are created by time-sharing a single frequency, and by doing so, multiply the amount of available spectrum. In fact, TDMA techniques can increase the utilization of the communications spectrum by more than a factor of 10 for voice band data without affecting the audio quality to the end-user, thus preserving and extending an extremely precious and finite resource.

GPS is already playing a key roll in creating a common timebase to synchronize TDMA and other communications systems. These systems exploit the relative and absolute time accuracy of GPS in order to enhance the overall communications system performance. The following is a brief description of several applications.

Simulcast transmission is the process of incorporating two or more geographically separated transmitters all broadcasting identical and simultaneous information on the same frequency to provide contiguous coverage over an area larger than would be possible for a single transmitter broadcasting alone. This approach requires precise time control of the base band signals for the simulcast process to perform the desired function in order to time align the modulation timing clocks in the radio transmitters. GPS is exploited by Paging and Mobile Radio infrastructure systems to extend the coverage area through the simulcast technique in order to broadcast to a wide geographical area. Most Simulcast Paging and TDMA communications systems require relative timing accuracy on the order of 1 to 10 microseconds, easily achieved through the use of GPS.

New paging transmission protocols such as FLEX, (high-speed paging) also requires absolute timing alignment to a single timing reference (UTC) in addition to the relative timing accuracy needed for simulcasting. The absolute timing reference allows paging receivers to power up only when information can be expected to be received, thereby conserving precious battery power in the pager. The UTC time mark from GPS allows the world-wide paging infrastructure to be time synchronized globally to benefit the mobile pager user as he travels from city to city, with the benefit to the user being much longer battery life.

Examples of digital TDMA communications systems that are already operational are MIRS (Motorola Integrated Radio System) and GSM (Global System for Mobile). MIRS is a cellular telephone-like system that provides

0-7803-2500-1/95 $4.00 © 1995 IEEE

voice dispatch, wireless telephone, text messaging, and future data capabilities all within the same system, and GSM is the standard for cellular telephone infrastructures in Europe and elsewhere. Both use TDMA technology to maximize the use of the allocated spectrum.

II. THE INTEGRITY ISSUE

As more and more communications systems rely on GPS for their timing source, it is clear that the reliability of the communications link is a function of the entire GPS system. Even though GPS has demonstrated superb reliability and availability, a few operational glitches have occurred that affected the time and position solutions (i.e., the GPS receiver output). At least one of these has contributed to a complete shutdown of a prototype TDMA system until the problem was cleared by the GPS Control Segment some 30 minutes later (the affected satellite was set unhealthy). A major and somewhat unadvertised flaw in the GPS system design is that there is no apparent way for the Control Segment to immediately remove a bad satellite from the constellation, the process time to do so is typically between 5 and 45 minutes.

Because the probability of these types of events is nonzero, most communications system engineers think twice about placing a strong dependency on GPS-derived time and look to every possible backup system and technique to greatly reduce the likelihood of complete system shutdowns. In this litigious society, it's the guy with a life-threatening problem who can't place his "911" call due to a system outage that causes the most worry.

III. NORMAL GPS TIMING PERFORMANCE

Under normal conditions, GPS performs well and provides time accuracy data that exceeds the requirements of the communications link. The normal CA code GPS system errors are dominated by the intentional SA dither and epsilon components, for which the combined one sigma user range error to a particular satellite is about 33 meters, or about 110 nanoseconds (assuming the receiver is placed on a known location). Modern low cost GPS receiver cores generally track m satellites at a time, and have the potential to average the time measurements from all satellites tracked. The averaging process reduces the resulting one sigma time error by the factor $1/\sqrt{m}$. For example, Figure 1 shows the time error histogram for 18 hours of 1PPS output clocks (as compared to the United States Naval Observatory Master Clock) of a commercially available low cost 6 channel GPS sensor [1]. The observed standard deviation of the data was 43 ns, in line with the $1/\sqrt{6}$ noise reduction predicted from averaging the independent time estimates from 6 satellites.

IV. ABNORMAL GPS TIMING PERFORMANCE

Characterizing the behavior of GPS under abnormal conditions is difficult to do as one has to wait a long time to collect several events. In addition, the DoD does not share very readily what happened during abnormal events, stating only that a particular satellite or satellites were not available for a period of time.

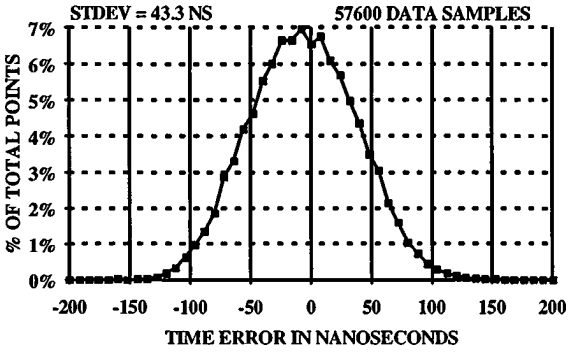


Figure 1. Probability Density Function of 18 Hours of CA Code GPS 1PPS vs. USNO MC.

Any number of error conditions that force one or more satellite ranges outside of the normally observed error profile has the potential of introducing larger than expected timing errors into an end application. The errors can come from any number of sources, including (but not limited to) satellite clock failures, ephemeris prediction or upload errors, satellite clock prediction failure, Control Segment errors, or simply from undetected software bugs in the Control or Space Segment.

The failure conditions all point to one of two classes of failures as observed by the receiver. A ramp error can occur, where the time error to one or more satellites grows from zero at some fixed or variable rate, or a step error in which the time error of the system snaps to a large value. Typically, the step error condition occurs when a receiver acquires an already failed satellite before the Control Segment can set it unhealthy, the receiver passes this condition to the user in the form of a time error step. In addition, as the satellite "hands over" from the current set of ephemeris to the next set of ephemeris, an error in the new ephemeris set could cause the receiver to predict the satellite location in error, which will cause a step error in time as viewed by the GPS receiver.

V. REMEDIES

There are at least two ways to decrease a system's susceptibility to GPS outages. First, disciplining a relatively stable oscillator with GPS affords the luxury of being able to detect larger than normal GPS errors, reject them, then "coast" on the stable oscillator until the Control Segment can fix the condition. Figure 2 shows a simple block diagram of how this can be accomplished using a low cost GPS core.

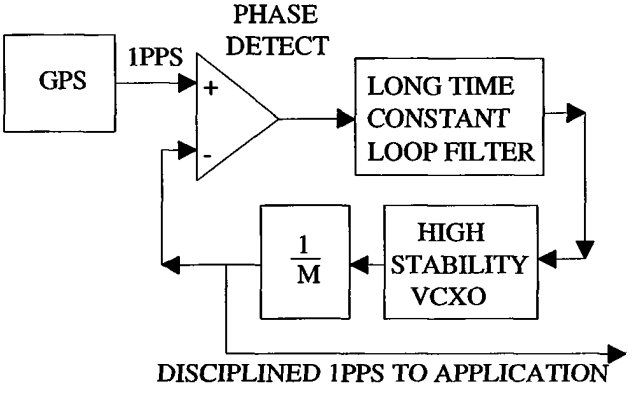


Figure 2. Typical System Block Diagram to Discipline a Stable Oscillator with a Low Cost GPS Sensor.

The system continuously measures the phase error between the GPS 1PPS output and a 1PPS signal created from integer division of a high stability local oscillator. The system forms a phase locked loop using a very long time constant loop filter that is most likely accomplished in software. The filter has a wide initial bandwidth to quickly skew the replica 1PPS signal to near zero phase error, then the bandwidth significantly narrows over time to an effective time constant of hundreds to thousands of seconds, depending on the stability of the oscillator. After convergence, memory within the filter has the ability to reject GPS data (i.e., don't update the loop filter) when the phase error residual exceeds a pre-set value (as would occur when GPS failed for a period of time) and the system then coasts on its previous good inputs. Obviously, the cost of the stable oscillator can greatly exceed the basic cost of the GPS receiver.

Second, there are a class of RAIM algorithms that can be applied to the time problem internal to a GPS receiver. These algorithms then allow the receiver to automatically detect and remove failed satellite(s) from inclusion in the resulting time solution. This second approach is a much lower cost alternative (it's only software). The formulation and testing of a particular T-RAIM algorithm is the topic of the remaining portion of this paper. The most demanding time-integrity applications can be met by applying both techniques simultaneously.

VI. FAULT DETECTION, ISOLATION, AND CORRECTION TECHNIQUES (FDIC).

GPS FDIC has been an area of intense interest for the last several years. Initial work was motivated by the strict integrity requirements of the Federal Aviation Administration (FAA), based on the effects which GPS satellite failures could have on flight safety for aircraft which used GPS for supplemental navigation. This work led to the development of RAIM algorithms [2, 3], which use the redundancy available in the GPS constellation to detect failures. The FDIC techniques have been applied to the more general GPS navigation problem [4, 5, 6], with emphasis on generation of solution quality measures, in addition to isolation from the effects of satellite failures.

Fault detection which exploits the redundancy available in the GPS measurements is generally based on the construction of a parity vector [2]. The parity vector is a linear combination of the (redundant) measurements which is orthogonal to the measurement "range space", i.e., the set of all possible error free measurements:

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{v} \qquad (1)$$

$$\mathbf{p} = \mathbf{P}\mathbf{r} \qquad (2)$$

where: $\mathbf{r}$ = GPS pseudo range measurement residual vector (dimension m)
m = number of satellites in track
$\mathbf{H}$ = measurement gradient matrix (dimension m x 4)
$\mathbf{x}$ = position (and clock) solution correction vector (dimension 4)
$\mathbf{v}$ = pseudo range measurement error vector (dimension m)
$\mathbf{p}$ = parity vector, dimension m-4
$\mathbf{P}$ = transformation matrix (dimension m-4 x m)

The transformation matrix P must be orthogonal to the measurement gradient matrix for $\mathbf{p}$ to qualify as a parity vector, thus

$$\mathbf{PH} = [0] \qquad (3)$$

Making use of equation (3) in equations (1) and (2) results in equation (4).

$$\mathbf{p} = \mathbf{P}\mathbf{v} \qquad (4)$$

Thus, the parity vector is solely a function of the measurement errors, and independent of the solution error vector **x**. In general, the magnitude of the parity vector can be used to detect failures, or inconsistencies, in the redundant measurements, and its orientation can be used to isolate the faulty measurement. Construction of the P matrix is non-trivial. Fortunately, there is an alternative formulation of the problem which permits direct computation of the parity vector in the measurement space: this form of the parity vector will be referred to as the fault vector [2].

The magnitude of the fault vector is equal to the magnitude of the parity vector, but it is of dimension m. The advantage of using the fault vector is that it can be constructed directly from the residuals of the solution correction vector:

$$\mathbf{f} = (\mathbf{r} - \mathbf{H}\mathbf{x}^*) / \sigma_{PR} \qquad (5)$$

where: $\mathbf{x}^* = \mathbf{H}^*\mathbf{r}$ (the least squares solution for **x**)
$H^* = (H^T H)^{-1} H^T$ (pseudo-inverse of H)
σ_{PR} = one-sigma pseudo range measurement error.

Note that, in using the least squares approach in Eq. (5), it is implicitly assumed that each pseudo range measurement is equivalent in accuracy (in a statistical sense). This assumption is valid for non-differentially corrected pseudo ranges, since SA is the dominant error effect, and is equally distributed across the GPS satellites. Given the fault vector constructed using Eq. (5), the detection and isolation process is provided by the following equations:

$$\text{if } (\mathbf{f}^T\mathbf{f} > T) \text{ then a failure is detected} \qquad (6)$$

$$i_{max} = \max_i (f_i^2 / S_{ii}) \qquad (7)$$

where: $S = I - H H^*$ (fault vector covariance matrix, dimension m x m),
i_{max} = the index of the most likely failed satellite

Selection of the failure detection threshold T involves a tradeoff between missed detection probability and false alarm rate: a threshold which is too low will result in a high false alarm rate, a "nuisance factor" for users of the system; on the other hand, a threshold which is too high can lead to missed detections, with generally more serious implications. A detailed discussion of possible methods of threshold calculation (involving fixed false alarm rates or detection probabilities) is beyond the scope of this paper: consult [7] for an excellent discussion of the subject.

In addition to its use in failure detection, the magnitude of the fault vector can be used in estimating the solution accuracy when no failures are present. The so-called aposteriori statistic of unit variance [4, 5] can be found as:

$$u = \mathbf{f}^T\mathbf{f} / (m - 4) \qquad (8)$$

This statistic represents a measure of the departure of the residuals from their expected magnitudes, and can be used to scale the apriori (theoretical) statistic for solution error variance, $(H^T H)^{-1}\sigma_{PR}^2$.

VII. APPLICATION TO GPS TIMING

The general FDIC equations, provided in the previous section, simplify considerably when applied to operations at a fixed, surveyed site [6]; only a single unknown remains to be determined, the offset of the GPS receiver clock. The equations for computation of the components of the fault vector are given by:

$$f_i = (r_i - R_a) / \sigma_{PR} \qquad (9)$$

where: $R_a = (r_1 + r_2 + \ldots + r_m) / m$

Eq. (9) can be substituted into Eqs. (6), (7), and (8) to obtain the required equations for fault detection, isolation, and accuracy prediction. The threshold for failure declaration can be found from the specified time alarm limit and the desired detection probability. In the next section, the algorithm's performance will be characterized as a function of the number of satellites in track.

VIII. TEST AND EVALUATION

A Monte Carlo simulation was utilized to estimate the detection and isolation probabilities and false alarm rate of the T-RAIM algorithm. A large number of trials are required, since the missed detection and false alarm probabilities are (by design) very low. For each trial, independent sets of random numbers are used to simulate the errors associated with each simulated pseudo range measurement: a white measurement noise, a bias error representative of the epsilon component of SA, and a second order Markov process representative of the dither component of SA. A satellite clock failure is modeled as

a ramping pseudo range error superimposed on one of the satellites. A maximum number of eight receiver channels are modeled, corresponding to Motorola's VP ONCORE design. Table 1 indicates the measured probabilities as a function of the desired time alarm threshold and the number of tracked satellites:

Table 1. T-RAIM Algorithm Performance Summary

Time Alarm (nsecs)	#SV detect	#SV isolate	P_d, P_i	P_f
T > 1000	2	3	>99.99%	< 2.e-4%
1000 > T > 700	3	3	>99.99%	< 2.e-4%
700 > T > 500	4	4	>99.99%	< 2.e-4%
500 > T > 300	8	8	>99.99%	< 2.e-4%
300 > T > 250*	8*	8*	>99.99%	< 2.e-3%*

* The minimum required false alarm rate of 2.e-4% cannot be achieved for this time alarm range.

The number of satellites appearing in the table are the minimum which must be tracked to ensure the indicated detection, isolation, and false alarm probabilities. Note that the minimum detection (and isolation) probability corresponds to less than 1 missed detection every 10,000 failures; given the infrequency of GPS satellite failures, this makes a missed detection virtually impossible. Except for the lowest range of threshold values, the false alarm probability corresponds to less than 1 false alarm every 5.7 days. Given that track is maintained on eight satellites, the time alarm can be set as low as 300 nsecs without compromising performance.

In addition to assessing the algorithm's detection and isolation performance, the simulation has been used to assess the accuracy of the solution quality measure (in the absence of failures). Table 2 summarizes the results of the apriori (theoretical) and aposteriori (observed by the simulation) performance of the algorithm.

Table 2. Effectiveness of Solution Quality Measures

Statistic	Apriori	Aposteriori
< 1 sigma	84%	82%
< 2 sigma	98%	96%

The statistics appearing in Table 2 are based on several hundred thousand Monte Carlo samples, and indicate that the aposteriori statistic provides a slightly more realistic measure of performance under nominal (i.e., unfailed) conditions; e.g., theoretically, 95% of the simulated samples should be below their 2 sigma value. More importantly, however, the aposteriori statistic indicates the effects which satellite failures have on the solution accuracy. Figure 3 illustrates this property. Note the close correspondence between the aposteriori statistic and the timing error as the satellite failure (represented as a 10 m/sec ramp) progresses to the end of the simulation, when a failure is declared. On the other hand the apriori statistic, since it is not dependent on the data, shows no evidence of the failing satellite.

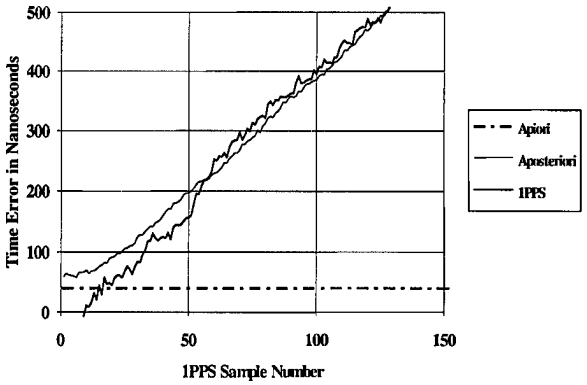


Figure 3. Aposteriori Statistic as a Measure of Solution Error.

Finally, it is worth discussing the affect of multiple satellite failures. The failure detection and isolation techniques presented work best when only a single satellite has failed; however, multiple satellite failures can be readily accommodated. Since satellite failures are unlikely events, it is virtually impossible for 2 satellites to fail at the same time; therefore, the approach presented for T-RAIM can be used to detect subsequent satellite failures once the original, failed satellite has been removed from the solution. Of course, the algorithm's effectiveness will be diminished as each failed satellite is removed from the solution, since the level of redundancy will be reduced.

IX. IMPLEMENTATION DESCRIPTION

The algorithm described above was installed into the firmware of a commercially available low cost GPS receiver module, the Motorola VP ONCORE 8 channel receiver. This receiver has found its way into many time critical applications, including those mentioned earlier. The following description summarizes the details of the implementation by describing the parameters that the user can control for his particular application, and specifics on what important system parameters are made available to the user for observation and use by the end-application.

X. USER CONTROLLABLE PARAMETERS

The VP-ONCORE GPS receiver T-RAIM setup and status message allows the user to modify the T-RAIM algorithm control parameters including the desired time alarm limit. The user can modify the following parameters:

- Output message rate. The user can request the T-RAIM status message one time (polled) or continuously at a user-specified update rate.
- T-RAIM algorithm enable. The user can enable or disable the T-RAIM algorithm.
- T-RAIM alarm limit. The user can set the maximum alarm rate allowable between 100 nanoseconds and 65.535 microseconds, in steps of 100 nanoseconds.

Along with the current status of the user-selected parameters, the VP-ONCORE GPS receiver will output the following parameters to the user for exploitation in the application:

- Time solution status. The receiver indicates the reliability of the time solution. Besides a status of OK and ALARM, the status could be UNKNOWN because the alarm limit is set too low, the T-RAIM algorithm is disabled, or there are insufficient satellites being tracked to make a positive determination.
- T-RAIM algorithm status. The receiver indicates the ability to detect and/or isolate faulty satellite signals. Refer to Table 1 for an indication of the number of satellites required for detection and isolation for a given alarm limit.
- One sigma time solution accuracy estimate. The receiver computes an estimate of the overall accuracy of the time solution based on the measured range residual statistics.
- Negative sawtooth residual. The underlying noise due to the granularity of the internal clock generating the 1PPS signal is deterministic on every pulse. The receiver computes and outputs the residual.
- Individual time solutions for each of the M satellites. The receiver outputs the individual time solutions for each of the satellites being tracked.

For a more detailed description of the T-RAIM I/O message, see [8].

XI. FIRMWARE MODIFICATIONS AND REAL-TIME ISSUES

The receiver firmware was modified using the fault vector approach as described above. When the time solution is in error and the T-RAIM routine can isolate the offending satellite, it will set an internal satellite health indicator to unhealthy which causes the firmware to remove the satellite from the tracked list. Furthermore, upon isolation, the firmware recalculates the receiver time and 1PPS controls without the offending satellite, thus protecting the 1PPS signal. The alarm is raised, and the time sigma estimate is calculated with the offending satellite included to allow the user the ability to detect the error occurrence and suspected magnitude of the error (up to the limit of the time sigma output field).

The firmware requires two seconds to remove the satellite from the tracked list during isolation, so the alarm condition will exist for two seconds while the isolated satellite is flushed from the channel. The 1PPS pulse is protected during this time as the internal software controls that position the 1PPS pulse are re-computed based on all satellites except the satellite causing the alarm condition.

The calculated time output by the receiver at time T_{k+1} is the result of a least squares algorithm using pseudo range measurements taken at time T_k. Although the satellite is detected in error at time T_{k+e} (some time between time T_k and T_{k+1}) the satellite is actually removed from the visible list one second later at time T_{k+1+e}. Thus, the satellite is not fully removed from track until time T_{k+2} (but it is essentially removed immediately via recalculation of the 1PPS control parameters using all but the offending satellite within interval T_k). Satellites that are removed from the track list (i.e., those set unhealthy by the T-RAIM algorithm) will be set healthy and examined again at the top of the hour to determine if the offending condition has been cleared by the Control Segment. If the Control Segment sets the offending satellite unhealthy during this period, the receiver will not use the satellite in its solution upon reexamination at the top of the hour.

XII. T-RAIM PERFORMANCE EVALUATION, TEST SETUP.

Figure 4 shows the test setup block diagram utilized in testing the T-RAIM algorithm. All testing was performed in a laboratory environment utilizing the following instrumentation:

- 2 VP ONCORE GPS receivers, loaded with the T-RAIM firmware (one represented the unit-under-test and the other was the truth model).
- 2 GPS Receiver Control Computers running the GPS Control Software (IBM PC Compatibles).
- Northern Telecom STR2760 Eight Channel GPS Satellite Simulator.
- HP5371A Frequency and Time Interval Analyzer.
- HP300 Computer (Logged Time Difference Data via the HPIB Port from the HP5371A.

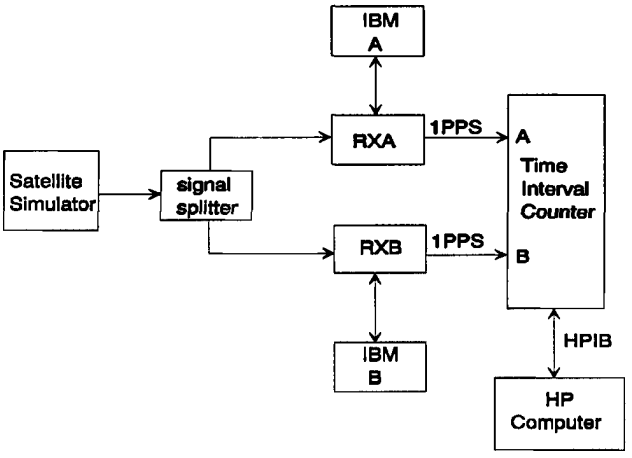


Figure 4. Block Diagram of the Test Setup Used to Test the T-RAIM Firmware Implementation.

The GPS satellite simulator is an important test equipment component that can simulate the L-band spread spectrum signals of up to eight different GPS satellites. Internal interfaces give the user the ability to program particular scenarios such as user location, platform motion, signal blockage, high dynamics, and other conditions not easily duplicated in a lab. It allows for scenario development and re-play, providing the GPS developer a means of testing of GPS receiver hardware and software under conditions that don't occur very often in the real-world. It is very useful for testing RAIM algorithms in that it allows the user to program particular faults into one or more simulated satellite signals, and run them over and over again to test and compare the performance of one or more algorithms.

The L-band output of the satellite simulator is fed through a signal splitter so that both the reference receiver and the receiver under test is supplied with the same input signals.

Both of the GPS receivers were placed in *position hold* mode at the position at which the signals were simulated. A scenario was produced in which all eight satellites were initially good (producing measurements containing simulated SA errors with standard deviation of about 33 meters each). One of these satellites (SV5) was then programmed to simulate a failure in one of two modes, a time ramp error consistent with what would happen if the satellite reference clock failed, and a time-step error to represent the effect of the receiver acquiring an already-failed satellite.

Receiver A was designated as the unit under test (UUT) and Receiver B as the reference receiver. Receiver A was programmed to track all eight satellites generated by the simulator, and as such, its 1PPS output signal would replicate the time error induced on SV5 (weighted by 1/8, representing the average over all eight signals). Receiver B was programmed to track all but the faulty satellite and thus produced a 1PPS signal with nominal system errors.

The HP5371A Time Interval Analyzer was used to measure the time difference between the two 1 PPS signals output by the GPS receivers. Since Receiver B is the reference, the difference measured represents the error in the 1PPS output of Receiver A. The Time Interval Analyzer samples the time error on every 1PPS pulse and transmits the resultant data over the HPIB bus to the HP300 computer. The HP300 computer recorded all of the 1 second time difference data during the individual test scenarios for later analysis.

The GPS Receiver Control Computers were used to set the T-RAIM conditions of the receivers and to record output data from the receivers. The T-RAIM algorithm was set to ON with an alarm limit of 600 ns. With an alarm limit of 600 ns, the T-RAIM algorithm would operate, according to Table 1, in the detect/isolate mode when at least 4 satellites were tracked.

Testing of the T-RAIM algorithm was accomplished by simulating two different satellite failure scenarios, a ramp error and a step error. For the ramp error condition, SV5 was configured to start accumulating a clock drift error at a particular point in time. The second scenario was a step error condition in which Receiver A was set to acquire a satellite that had already developed a large timing error.

For the ramp test scenario, the satellite simulator was programmed to run a 5 minute test. The test was designed to simulate 8 satellites with SV5 configured to simulate a time drift error. SV5 was programmed with a

time drift error of 300 ns per second, starting at 2 minutes into the test, giving the receivers time to acquire the signals and to provide sufficient "good" data at the start of the test for comparison. This time drift error corresponds to approximately 50 meters per second drift in the range error to the satellite. This condition represented a severe drift rate error, but is useful in quickly demonstrating the performance of the T-RAIM algorithm.

Receiver B was set to ignore SV5 so that the satellite would not cause an error in the reference receiver 1PPS output. Receiver A, the unit under test, was set with the TIME RAIM alarm limit threshold of 600 ns.

The second test scenario was the step test. In the step test, the satellite simulator was programmed so that SV5 contained a fixed offset time error of 5600 ns throughout the 5 minute test. This represents a significant timing error that would create a fixed 1PPS error of 700 ns, just over the alarm threshold.

Again Receiver B was set to ignore the errant satellite throughout the duration of the scenario. Receiver A was also initially set to ignore SV5, but then commanded to include it in the solution, forcing it to see the large time error as a step error when the satellite was acquired.

XIII. TEST RESULTS

Figure 5 shows the results of the 300 ns time error drift on one of the eight satellites during the ramp test. The graph shows the time difference between 1 PPS outputs of the two receivers. Additionally, the sigma error estimate of the receiver under test is plotted on the graph. The graph does not show the output for the entire 5 minute test, only the point at which satellite number 5 begins to accumulate a drift error until the receiver detects and isolates the errant satellite.

The 1 PPS time error, shown on the graph, has an RMS noise of about 60 ns, which is consistent with the expected statistics (the differencing operation multiplies the individual error statistics by $\sqrt{2}$). The measurement noise is caused by the quantization error of the clock used to generate the 1 PPS output signal within each receiver.

At approximately 10 samples on the graph, the 1PPS signal from Receiver A begins to drift away from that of Receiver B. The rate at which the error drifts is approximately 6 ns/s, which is what we would expect from averaging eight satellite signals with one having the programmed ramp of 50 ns/s. At 70 samples into the plot, the error on Receiver A accumulates to the point at which the bad satellite is removed from the time solution, and the 1PPS signal then reverts back to the nominal error condition.

Figure 6 shows the results from the step test scenario. Again, the graph shows the measurement of the time difference between 1 PPS outputs of the two receivers. The sigma error estimate of the receiver under test is also plotted on the graph. The graph does not show the entire 5 minute test, only the time immediately surrounding the event at which Receiver A begins to track SV5.

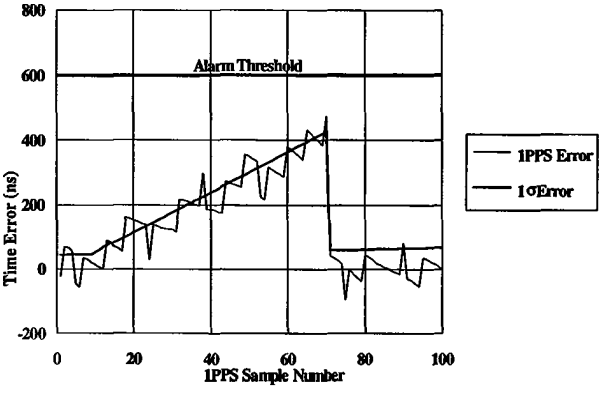


Figure 5. Receiver 1PPS Time Error Plot and Receiver Predicted 1 Sigma Time Error Statistic as Observed during the Ramp Test.

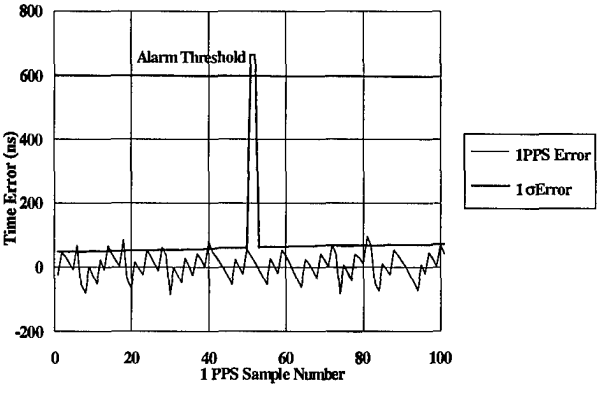


Figure 6. Receiver 1PPS Time Error Plot and Receiver Predicted 1 Sigma Time Error Statistic as Observed during the Step Test.

Receiver A was set to begin tracking SV5 at approximately the 75 sample point on Figure 6. Once the re-

ceiver acquired and locked on to the errant satellite (at sample number 110), the T-RAIM algorithm alarmed, indicating that a satellite in the solution was indeed unhealthy. The sigma error estimate shows the point at which SV5 enters the solution and then is removed 2 seconds later. The error estimate indicates the error that would have been introduced into the timing had the error been allowed to be used in the timing solution. As a result, the 1 PPS output of receiver A was unaffected by the satellite error. Figure 6 shows the 1 PPS error remained at nominal throughout the entire test.

XIV. ERROR PREDICTION VERIFICATION

As described earlier, the timing error can be predicted by the T-RAIM algorithm and was verified using simulated data. Additionally, tests were performed using real data. Sample test results were provided in both the ramp and step tests. As stated earlier, Figures 5 and 6 not only show the difference between the output 1 PPS signals of two receivers, but also the sigma error estimate predicted by the receiver. In the ramp test we see the estimate as a conservative prediction of the error as the errant satellite drifts off nominal. When the prediction reaches the point at which the algorithm removes the satellite from the solution, we see the error estimate return to nominal.

In the step test, we see the time error prediction spike above the selected threshold when the errant satellite is added to the solution. At first blush this would appear to be a false alarm because the errant satellite was removed from the time solution prior to causing any offset in the 1 PPS output. In the step error case, the errant satellite remains assigned to a receiver channel for two seconds while the internal pipeline processing removes it from being assigned to a channel. During this time, the range residual calculations include the errant satellite and therefore the sigma time error output spikes up until the satellite is actually removed from the channel. Notice that the 1PPS time error is not affected during this two second period as the 1PPS controls are recomputed internally upon the alarm condition forcing an immediate ignore function on the detected failed satellite.

The threshold for fault detection is set using Table 1 to calculate threshold level vs. the number satellites in track. The ramp test results show the faulty satellite is removed prior to the threshold being reached. In the results to the step test we see a different response, here the satellite was removed when the first data was used from the satellite. The faulty satellite in this scenario was grossly in error prior to the receiver tracking it and the error was well over the set threshold in the receiver. Once the satellite was tracked and the data used in calculating a solution the T-RAIM algorithm isolated, and immediately removed it from the solution.

Using Table 1, tests were performed to verify the false alarm rate with real data. The Alarm limit set to 500 ns, and long term observations were made on the alarm output signal. According to Table 1, an average of at least 4 satellites being tracked for up to 1 week is required to detect any false alarm conditions. Therefore, demonstration using real data was difficult without observing many weeks of data.

REFERENCES

[1] King, T. M., Miranian, M., Busch, D., "Test Results and Analysis of a Low Cost GPS Receiver for Time Transfer Applications", Proceedings of the Institute of Navigation National Technical Meeting, January 1994.

[2] Sturza, M.A., "Navigation System Integrity Monitoring Using Redundant Measurements", NAVIGATION, Journal of the Institute of Navigation, Vol. 35, No. 4, Winter 1988-89.

[3] Brown, R. G., "A Baseline GPS RAIM Scheme and a Note on the Equivalence of Three RAIM Methods", NAVIGATION, Journal of the Institute of Navigation, Vol. 39, No. 3, Fall 1992.

[4] Geier, G.J., and Sheynblat, L., "The Application of Quality Control to Real-Time Differential GPS Navigation", Proceedings of the Institute of Navigation National Technical Meeting, June, 1993.

[5] Geier, G.J., Musser, D., and Ackroyd, N., "The Application of Integrity Monitoring to Offshore Seismic Survey Operations", NAVIGATION, Journal of the Institute of Navigation, Vol. 39, No. 3, Fall 1992.

[6] Copps, E.M., and Lupash, L.O., "GPS Integrity Monitoring from a Surveyed Site", GPS-ION-92 Proceedings.

[7] Brown, A.K., and Sturza, M.A., "Comparison of Fixed and Variable Threshold RAIM Algorithms", GPS-ION-90 Proceedings.

[8] Motorola VP ONCORE GPS Receiver Users Manual.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

COMMON-CLOCK TWO-WAY SATELLITE TIME TRANSFER EXPERIMENTS

Christine Hackman, S. R. Jefferts and Thomas E. Parker

National Institute for Standards and Technology
Time and Frequency Division, MS 847.5
325 Broadway
Boulder, CO 80303 USA

Abstract

We report the results of a series of measurements designed to determine the magnitudes and noise types of the various sources of instability in the two-way satellite time and frequency transfer (TWSTFT) process. We find that, for our equipment, the earth-station noise predominates at the 1 day time scale. We also report results from a maser-to-maser time transfer conducted between NIST-Boulder and USNO (Washington D.C), which show a stability at 1 day of about 1 ns.

KEY WORDS - Two-way satellite time and frequency transfer, two-way time transfer, time transfer, TWSTFT.

Introduction

Two-way satellite time and frequency transfer (TWSTFT) is used to measure the time difference between two geographically-separated clocks A and B. This is done using a transmitting/receiving earth station at each location as well as a geosynchronous communications satellite [1], [2]. If reciprocity in the atmospheric and satellite transponder delays is assumed [3], [4], one can compute the time difference between the two clocks as:

$$\text{time(A) - time(B)} = \tfrac{1}{2}\{[\text{TIC(A) - TIC(B)}] + [\text{Tx(A) - Rx(A)}] - [\text{Tx(B) - Rx(B)}]\} + \text{Sagnac effect} \quad (1)$$

where

TIC(i) = time interval counter reading at station i
= time between transmit and receive 1 pulse per second (PPS) at station i
Tx(i) = transmit delay through equipment at i
Rx(i) = receive delay through equipment at i.

The first line of Eq. 1 contains the time interval counter (TIC) measurements taken at both sites. Line 2 is the difference between the transmit and receive delay at site A minus that difference at site B. Line 3, the Sagnac effect, depends on the geographic locations of the two earth station antennae and the satellite.

To obtain an accurate time difference (time(A)-time(B)) one needs to know the values of the quantities on the second and third lines of Eq. 1. For accurate frequency transfer it is necessary only that these quantities remain constant over time. This is often assumed in TWSTFT. In 1994, at the National Institute of Standards and Technology (NIST), several ongoing experiments to evaluate the stability of the quantity [Tx(A) - Rx(A)] - [Tx(B) - Rx(B)] (line 2), were begun. In this paper we report our findings on various aspects of the time and frequency stability of a co-located TWSTFT earth-station pair and also assess the performance of TWSTFT between two geographically-separated stations.

Method

Operationally, TWSTFT is used to determine time(A) - time(B) where clocks A and B and their respective earth stations are geographically separated. In our primary measurement, two separate earth stations have been set up at the NIST facility in Boulder, and UTC(NIST) serves as the clock for both earth stations. We then perform TWSTFT between our own two earth stations. In this case, Clock A = Clock B = UTC(NIST) so that time(A) - time(B) = 0. In addition, because the antennae are located atop the same building, the Sagnac effect goes to zero. Applying these facts to Eq. 1 yields:

$$\tfrac{1}{2}\{\text{TIC(B) - TIC(A)}\} = [\text{Tx(A) - Rx(A)}] - [\text{Tx(B) - Rx(B)}] \quad (2)$$

By observing $\tfrac{1}{2}\{\text{TIC(B) - TIC(A)}\}$ over a period of time,

U.S. Government work not protected by U.S. copyright

the day-to-day stability of the earth-station pair can be evaluated. It should be noted, however, that common-mode environmental instabilities may not be observed.

One of the earth stations used in this measurement is that which is routinely used for time transfer between NIST and other laboratories in North America and in Europe. This earth station features a 3.7 m diameter antenna and its transmit/receive frequencies can be adjusted in steps of 120 Hz. The second earth station, a less-expensive "VSAT" earth station package, has an antenna 1.8 m in diameter. Its transmit/receive frequencies can be tuned in steps of 1 MHz. These earth stations are shown in Figs. 1 and 2. SBS-6, a geosynchronous communications satellite located at 95 degrees west longitude, was used as the transfer satellite.

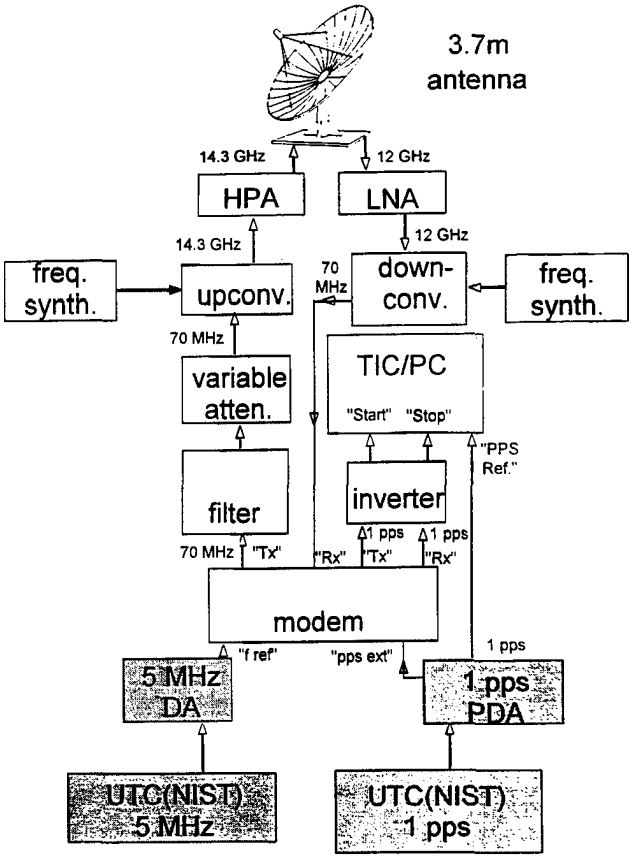


Figure 1. The NIST TWSTFT earth station used for time transfer with other metrology laboratories. The shaded rectangles represent equipment common to both this and the VSAT earth station. PDA is an abbreviation for pulse distribution amplifier, TIC/PC is the time-interval-counter and DA is a distribution amplifier.

Measurement 1

Here we evaluated the noise contributed by the measuring system, i.e., the time interval counters, spread spectrum modems, etc. [5]. These evaluations were made by measuring the time intervals between two 1 PPS signals for 5 minutes (300 measurements) once each day and then calculating an average value for each session. This process was repeated approximately three times per week (usually Monday, Wednesday and Friday) over a period of 116 days.

The data in these measurements, and some of the other measurements to be discussed later, were unevenly spaced. This complicates the process of calculating $\sigma_x(\tau)$ [6]. The $\sigma_x(\tau)$ values were computed by using the adjacent time values as if they were evenly spaced to compute $\sigma_x(\tau = 2.3, 4.6, 9.2,$ etc. days), where $\tau = 2.3$ days is the average interval between data points. By

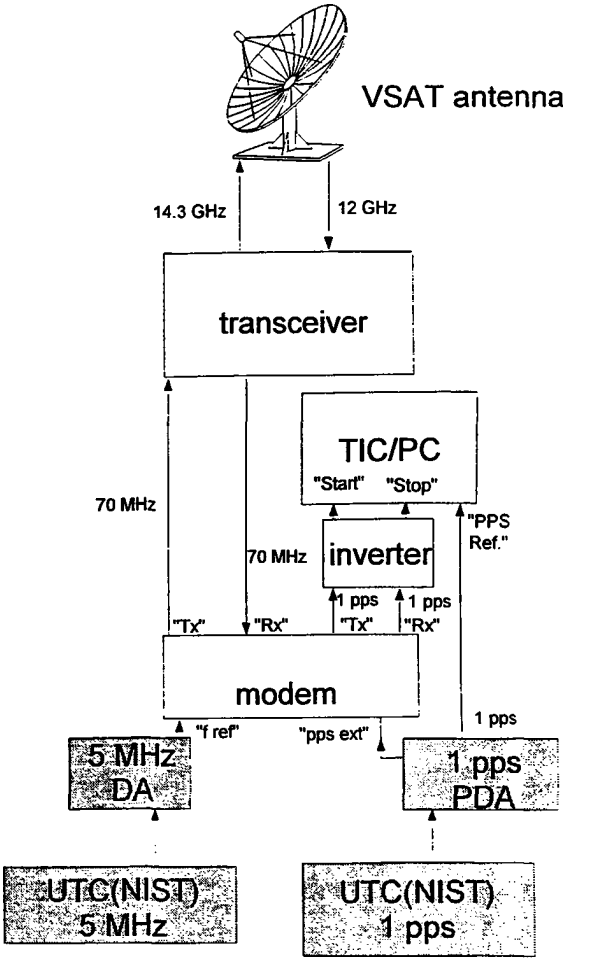


Figure 2. The NIST VSAT earth station used for test purposes.

plotting $\sigma_x(\tau)$ against τ, we could see that the noise type was generally either white phase modulation (PM) or near white PM. This was fortunate since the impact of non-periodic data on white PM noise is easily understood and the dependence on τ is preserved. For a white PM process, where all of the data points are independent, the calculated noise from our data set for τ = 2.3 days would be the same as that for τ = 1 day for a data set taken at a 1 day interval, assuming that the same white PM noise process was present at a 1 day time interval. We don't know this for a fact with the TICs or the modems, but for comparison purposes we will use estimated values for $\sigma_x(\tau$ = 1 day) since data were actually taken at 1 day intervals in measurements 2 and 3. Analyzing non-periodic data for other noise types (flicker PM or white frequency modulation (FM)) is considerably more complicated [6] and will be the subject of a future paper concerning typical TWSTFT data intervals [7].

The noise levels of both individual time-interval counters were estimated to be approximately 50 ps for τ = 1 day, with three sigma confidence limits of approximately +32% and -20%. Note that the noise of both TICs combined in the manner of Eq. 2 gives $\sigma_x(\tau$ = 1 day) = $\frac{1}{2}(50^2 + 50^2)^{1/2}$ = 35 ps (assuming the noise processes in the two TICs are independent). To evaluate the combined noise of the modems and TICs a two-way time transfer was performed by using a direct cable link between the two modems as shown in Figure 3. The individual TIC readings were used in Eq. 2 the same as they would in real TWSTFT. In this case, we obtain a $\sigma_x(\tau$ = 1 day) of 63 ps, again with the noise at larger averaging times decreasing as white PM. The confidence limits are approximately the same as for the TIC results. Clearly, the modems are contributing a significant amount of additional noise to the measurement process.

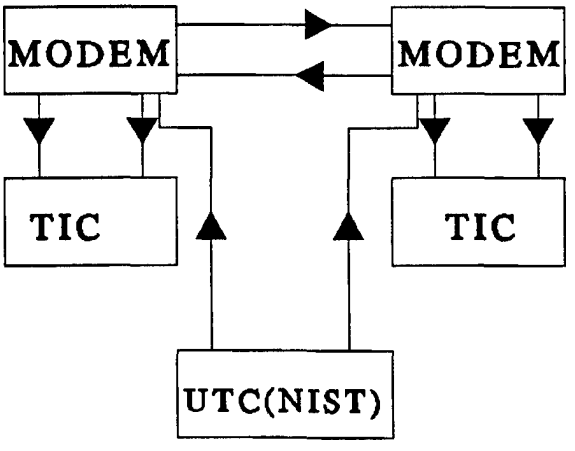


Figure 3. The experimental set-up used to evaluate the stability of a pair of measurement systems.

Measurement 2

Here we again evaluate the noise of the measurement system, but in this case rather than cabling the modems directly together as in Fig. 3, we send the signals up through the satellite using a single earth station. By using this configuration we force the parts of the transmit and receive delays associated with the earth stations in Eqs. 1 and 2 to cancel. The method of doing this is shown in Fig. 4. The 70 MHz outputs of the two modems are fed into a power combiner. The combined 70 MHz signal is then either sent up through the upconversion/transmission equipment of the 3.7 m earth

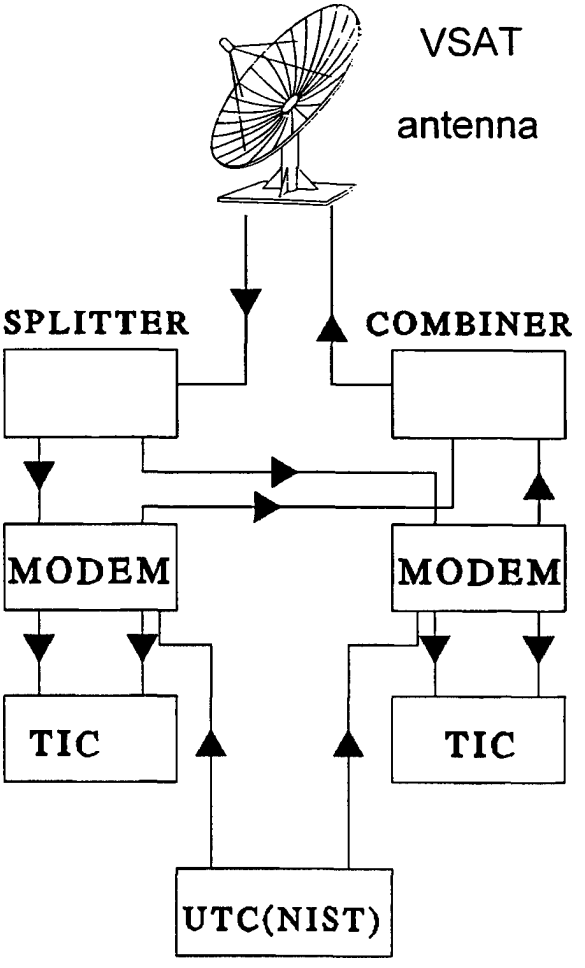


Figure 4. The single-dish common-clock experiment as used with the VSAT earth station. This experiment was also carried out using the 3.7 m earth station.

station or that of the VSAT earth station. The resulting 14.3 GHz signal is transmitted to the satellite, retransmitted by the satellite at 12 GHz, and then received back at the same antenna. This combined signal is then down converted, sent through a power splitter, and then back into the two modems.

Since we are only using one earth station, Tx(A) = Tx(B) and Rx(A) = Rx(B). Therefore, the quantity [Tx(A) - Rx(A)] - [Tx(B) - Rx(B)] will go to zero, and we are once again measuring only the noise contributed by the measurement system. However, in this case, we actually do send the signal up through the satellite. This series of measurements is referred to as "the single-dish common-clock experiment". These measurements were taken with both earth stations three times per week, 300 s at a time, over the course of about 220 days. In addition, measurements were also taken once per day using the 3.7 m earth station over a period of 3 weeks. Since the results were similar on both the 3.7 m and VSAT earth stations, only the results from the former will be discussed.

Figure 5 shows the time difference results obtained from the 3.7 m earth station. Figure 6 shows the $\sigma_x(\tau)$ values derived from the data in Fig. 5. The solid circles in Fig. 6 come from the three weeks of daily data and the x's represent the three-times-per-week data. The calculation of $\sigma_x(\tau)$ for the daily data is straightforward

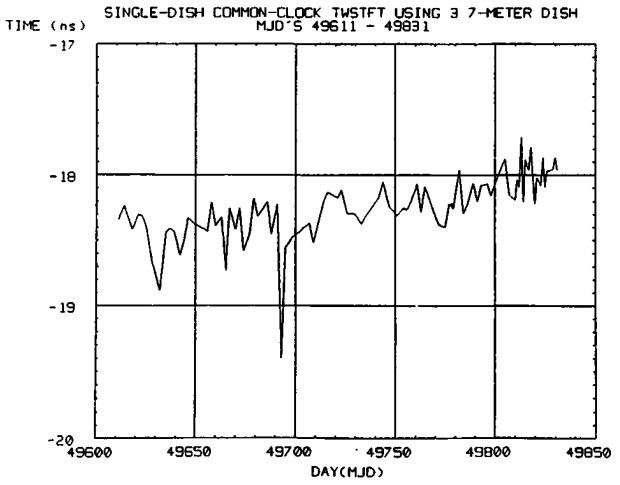


Figure 5. The time difference results obtained from the single-dish common-clock experiment. These results were obtained with the 3.7 m earth station.

since the data are periodic, while the three-times-per-week data were analyzed using the technique discussed earlier. Note that the daily data and the three-times-per-week data are in good agreement at $\tau = 2$ days. The error

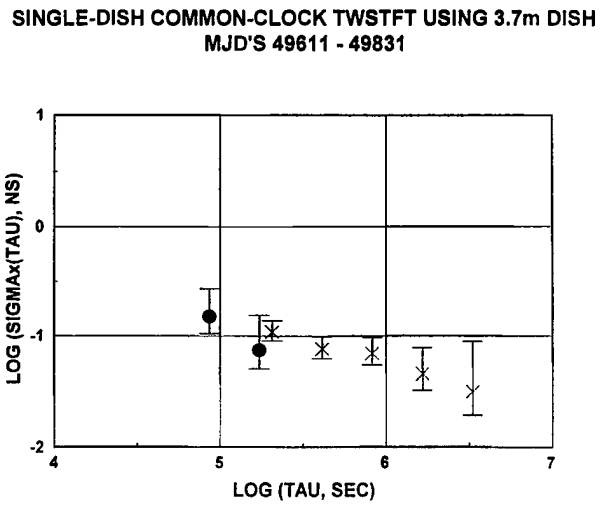


Figure 6. The time stability, $\sigma_x(\tau)$, of the data shown in Fig. 5. The black dots indicate results obtained from performing this experiment every day for 21 days. The x's indicate results obtained from taking data 3 times per week over the course of approximately 220 days. The error bars indicate three-sigma confidence levels.

bars in Fig. 6 represent three-sigma confidence limits.

As Fig. 6 shows, $\sigma_x(\tau = 1 \text{ day})$ is about 150 ps with a noise type between white PM and flicker PM. One might expect to get approximately 60 ps as observed in the experiment of Figure 3 since the earth station delays in the present experiment should cancel. However, there is an important difference. For this measurement, the time it takes for the signal to leave one modem and arrive at the other modem is on the order of 0.25 s, because the signal travels to and from the satellite. In the measurement of Fig. 3, it only takes a few tens of ns for a signal to travel from one modem to another. Flicker noise in the measurement system (most likely the modems) could be the reason that increased signal delay is associated with increased noise.

Another difference between the experimental set-ups shown in Figs. 3 and 4 is the signal-to-noise ratios as received by the modems at 70 MHz. However, measurements were made with the cable connected modems of Fig. 3 in which white noise was injected to create the same signal to noise ratio at 70 MHz as that present in the Fig. 4 configuration. The observed value for $\sigma_x(\tau = 1 \text{ day})$ was the same as that present without the injected noise. Finally, fluctuations in the delays up to and back from the satellite are not likely to cause the higher noise level since the propagation paths are virtually identical. Also, the same satellite transponder is used for both signals.

Figure 7 shows the frequency stability, $\sigma_y(\tau)$, derived

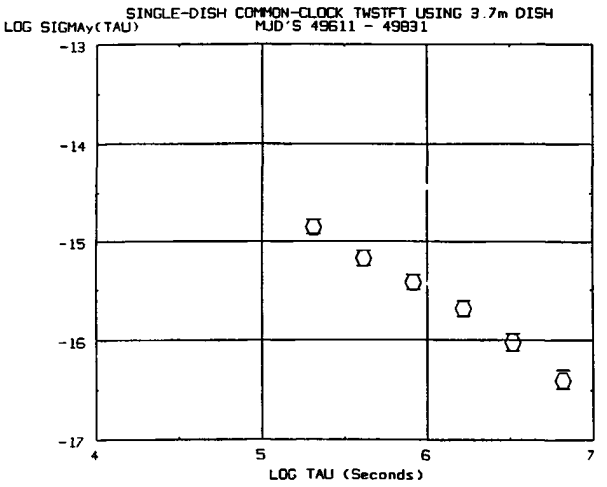


Figure 7. The frequency stability, $\sigma_y(\tau)$, of the results shown in Fig. 5.

from the data in Fig. 5. It is interesting to note that we reach a frequency stability of 10^{-15} at τ between 2.4 and 4.8 days.

To summarize, Figs. 5 through 7 represent the best possible TWSTFT results that could be obtained with the present modems and counters if all the noise from the earth stations, propagation paths and satellite were eliminated.

Measurement 3

In this experiment, we do two-dish common-clock TWSTFT. In other words, we have the two earth stations

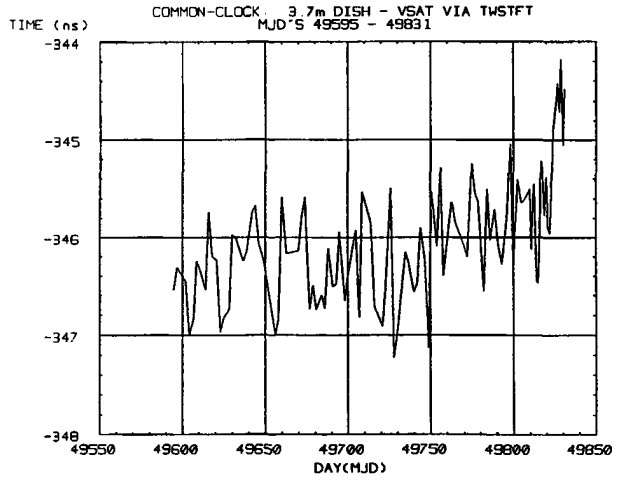


Figure 8. The time difference results obtained from the two-dish common-clock experiment.

shown in Figs. 1 and 2 performing TWSTFT with each other. As in measurement 2, we have data from a 21-day set of once-per-day measurements, and a longer data set consisting of data taken 3 times per week over the course of roughly 230 days. By comparing the results of this series of measurements with those obtained from the single-dish common-clock experiment, we can determine what type of noise the transmitter/antenna/receiver packages contribute to the noise of TWSTFT. However, we cannot determine the magnitude of the noise as we are unable to separate the noise due to one earth station from that due to the other. Furthermore, there may be some noise cancellation or enhancement due to common environmental effects.

Figure 8 shows the results from the two-dish common-clock experiment and Fig. 9 shows the time stability of the data in Fig. 8. Figure 9 shows that $\sigma_x(\tau = 1 \text{ day})$ is about 330 ps, and that the noise type is between white and flicker PM. Because the time stability of this experiment is considerably worse than the time stability obtained from the single-dish common-clock experiment (Fig. 6), we conclude that the earth stations are a significant additional source of noise with a noise type somewhere between white PM and flicker PM.

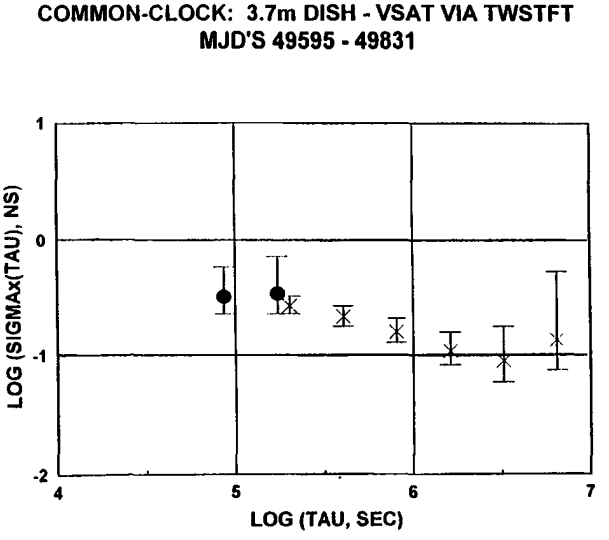


Figure 9. The time stability, $\sigma_x(\tau)$, of the results shown in Fig. 8. The interpretation of the symbols is the same as that in Fig. 6.

Measurement 4

We now report results obtained by doing TWSTFT between maser-based clocks at NIST and USNO. Note that this is no longer common-clock TWSTFT - the earth

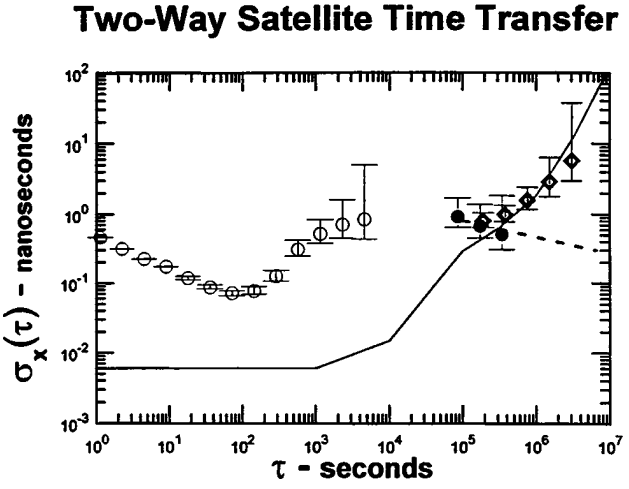


Figure 10. Time stability, $\sigma_x(\tau)$, of TWSTFT performed between NIST(Boulder) and the United States Naval Observatory (Washington D.C.). The open circles were obtained from performing a six-hour time transfer. The closed circles were obtained by performing a 5-minute time transfer once per day, every day, for 21 days. The open diamonds were obtained by performing a 5-minute transfer three times per week over the course of about 129 days. The solid line indicates the approximate level of the combined clock noise which we are attempting to measure.

stations at NIST and USNO are entirely independent and driven by separate clocks.

Figure 10 shows the time stability analysis of three experiments. In one experiment (open circles), a six-hour-long time transfer was performed between NIST and USNO with time interval counter measurements made every second. Note that this experiment has only been performed once and we do not know whether the results obtained are typical. In the second experiment (solid circles), a 5-minute time transfer was performed once per day every day for 21 days. In the third experiment (open diamonds), we did our normal NIST-USNO time transfer routine, i.e., one 5-minute transfer 3 times per week, over the course of 129 days. Figure 10 also shows the approximate stability of the clocks involved (solid line). What the line shows is the stability of the NIST clock, a maser-driven micro-stepper, as compared to a second hydrogen maser. We believe that the NIST clock is comparable in stability to the USNO clock (MC2) for τ less than about 1 day. At τ greater than 1 day maser drift and some steering corrections to the micro-stepper have increased $\sigma_x(\tau)$ for the NIST clock. Therefore, the solid line is a reasonable approximation of the combined stability of the two clocks being measured.

As Fig. 10 shows, at τ = 1 s, the time stability of TWSTFT is 470 ps. The stability of the transfer improves out to $\tau \approx 64$ s ($\sigma_x(\tau \approx 64 \text{ s}) = 70$ ps). However, the transfer stability then becomes worse out to τ = 1.25 hours, at which point $\sigma_x(\tau)$ = 840 ps. From the every-day experiment, we find that the time stability of TWSTFT is about 915 ps at τ = 1 day and that it continues to decrease in a white PM to flicker PM fashion out to τ = 4 days. We have drawn in a dashed line to indicate that the transfer noise very likely continues to go down as the averaging time increases, based on our analysis of the two-dish common-clock experiment. Finally, the long-term transfer results indicate that for $\tau \geq 4$ days, the combined noise of the two clocks under study is larger than the noise of the transfer process.

The transfer noise of TWSTFT is clearly made up of more than one noise process, as is shown in Fig. 10. The white PM noise process at τ = 1 s comes from the modems and counters. However, at least one other process sets in at about 100 s which causes $\sigma_x(\tau)$ to increase again to nearly 1 ns at 10^4 s. As mentioned earlier, the six hour run has been performed only once, and it is thus not known if the rise at 100 s is typical. More measurements are planned to determine the source of this noise process. It is clear, however, from the one day and normal three-times-per-week measurements that the noise consistently increases to near 1 ns by τ = 1 day. The common-clock experiments indicate that at least part of this noise comes from the earth stations. Beyond about four days, clock noise dominates and the transfer noise could not be measured directly. However the common-clock experiments suggest that the transfer noise decreases for increasing τ as in a manner between that of white and flicker PM noise.

Conclusions

The noise contributed to TWSTFT by the two measurement systems (counters, modems, etc.) is about 150 ps at τ = 1 day with a noise type between white and flicker PM. Most of this noise is from the modems. The two-earth station common clock measurements show that the earth stations are a significant source of noise with an observed value of $\sigma_x(\tau$ = 1 day) = 330 ps. The precise magnitude of an individual earth station's noise could not be determined. However, the noise type of the earth station pair's contribution was observed to fall between white and flicker PM for τ greater than 1 day. Finally, a preliminary evaluation of actual TWSTFT between NIST and USNO shows that the transfer noise is rather

complicated. At τ = 1 day it is just slightly smaller than 1 ns and appears to decrease for larger τ at a rate between white and flicker PM. Future experiments are planned to further investigate the characteristics and sources of TWSTFT noise and to achieve improved performance.

Acknowledgments

We would like to thank Angela McKinley, Bill Powell, Jim DeYoung, and Phu Mai of the United States Naval Observatory for their cooperation in the NIST-USNO time transfer experiments. In addition, we would like to thank Wayne Hanson, Al Clements, and Mike Lombardi of NIST for their assistance in various aspects of earth station operations and data analysis.

References

[1] G. de Jong, "Prospects of Two-Way International Time Scale Comparisons," in *Proc. of the 6th European Frequency and Time Forum*, 1992, pp. 69-73.

[2] D. Kirchner, H. Ressler, P. Grudler, F. Baumont, Ch. Veillet, W. Lewandowski, W. Hanson, W. Klepczynski, and P. Uhrich, "Comparison of GPS Common-View and Two-way Satellite Time Transfer Over a Baseline of 800 km," *Metrologia*, vol. 30, pp. 183-192, 1993.

[3] D. W. Hanson, "Fundamentals of Two-Way Time Transfers by Satellite," in *Proc. of the 43rd Annual Symposium on Frequency Control*, 1989, pp. 174-178.

[4] J. Jesperson, "Impact of Atmospheric Non-Reciprocity on Satellite Two-Way Time Transfers," in *Proc. of the 43rd Annual Symposium on Frequency Control*, 1989, pp. 186-192.

[5] The modems used were MITREX 2500's. The commercial equipment used has been identified for technical completeness. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology.

[6] F. Vernotte, G. Zalmansky and E. Lantz, "Noise and Drift Analysis of Non-Equally Spaced Timing Data," in *Proc. of the 25th Annual Precise Time and Time Interval Application and Planning Meeting*, 1993, pp. 379-388.

[7] C. Hackman and T. E. Parker, "Using TVAR with Unevenly Spaced Data," in preparation, to be submitted to *Metrologia*, 1995.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

1/F FLUCTUATIONS SOURCES IN DIRECT DIGITAL FREQUENCY SYNTHESIZERS AND THEIR CONTRIBUTION TO THE OUTPUT OSCILLATIONS POWER SPECTRAL DENSITY

V.N.Kuleshov,H.Y.Liu,and B.E.Leshukov

Moscow Power Engineering Institute,
14 Krasokasarmennaya,Moscow 111250,Russia

Abstract: An analysis of 1/f noise transformation to the output of direct digital frequency synthesizer (DDS) is presented. Basic 1/f noise sources located in the digital-to-analog converter (DAC) are taken into consideration. Using theory of the periodically nonstationary random processes general formulae for power spectral densities (SDs) both noise at the DDS output and 1/f type amplitude and phase fluctuations are obtained. They are simplified for two particular cases: noise sources located in bit current generators and noise sources located in biasing circuit. It is shown that these noise sources usually cause only amplitude fluctuations.

The main causes of 1/f type phase fluctuations brought in by DAC are fluctuations of bit currents switching times. An approximate analysis of these fluctuations is also presented.

Some results of DAC 1/f noise currents measurements are given. Basing on them quantitative estimations of DDS output noise are obtained.

Introduction

An analysis of output DDS noise due to fundamental fluctuation sources is presented in [1]. It seems interesting to expand it taking into consideration an influence of 1/f noise sources. But the approach developed in [1] is not valid to calculate 1/f noises transformation because correlation times of these noises are substantially more than DDS clock cycle. So one needs to develop an approach that doesn't need an assumption of small correlation time. It is presented in this paper.

We are going to use the DDS and DAC models shown in Figures 1,3 and 4 of the paper [1] and designations introduced in [1].

General Formulae for DDS Output Noise due to Bit Currents Noise Components

The approach that is developed in this paper is based on the presentation of the dimensionless unfiltered output DDS oscillation $a_k(t)$ as a sum of weighted two level switching functions $h_{bk}(t)$ of DAC bit currents

$$a_k(t) = \sum_{b=1}^{B} 2^{-(b-1)} h_{b,k}(t) \qquad (1)$$

where

$$h_{bk}(t) = \sum_{r=-\infty}^{\infty} \{ (\sigma[t - rKT_c] - \sigma[t - (1+rK)T_c]) + \sum_{d=1}^{D} (\sigma[t - (x_{2d,bk} + rK)T_c] - \sigma[t - (x_{1d,bk} + rK)T_c] + \sigma[t + (x_{1d,bk} - 1 - rK)T_c] - \sigma[t + (x_{2d,bk} - 1 - rK)T_c]) \qquad (2)$$

is a switching function of the b-th bit current when an oscillation having frequency kf_d/K is generated, $\sigma(t)$ is the unit-step function, $x_{1d,bk}T_c$ and $x_{2d,bk}T_c$ are switching off and on moments at the interval $0 < t < T_d/2$,

D is a number of couple of switchings (off and on) at the interval $0 < t < T_d/2$. Figure 1 shows an oscillation $a_k(t)$ at the DAC output and three first terms of the sum (2) for K=16 and k=1. It is worth to notice that if $a_k(t)$ is defined by eqs.(6),(7),(8) of the paper [1], then both $a_k(t)$ and all $h_{bk}(t)$ are even functions of $(t - T_d/2)$.

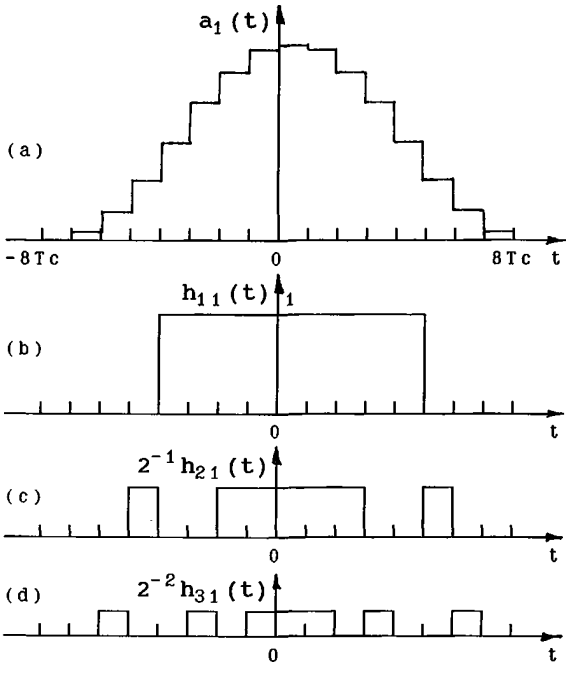


Fig.1 Dimensionless Unfiltered Output DDS Oscillation (a) and Three First Weighted Switching Functions (b,c,d) in the Sum (1)

0-7803-2500-1/95 $4.00 © 1995 IEEE

Introducing Fourier expansions

$$h_{bk}(t) = \sum_{i=-\infty}^{\infty} H_{bk,i} E^i \qquad (3)$$

where E was defined in [1] by eqs. (14-16) and

$$H_{b,k,i} = \frac{\sin(\pi i / K)}{\pi i} + \frac{1}{\pi i} \sum_{d=1}^{D} \left\{ \sin\left(\frac{\pi i(2x_{1d,bk} - 1)}{K}\right) - \sin\left(\frac{\pi i(2x_{2d,bk} - 1)}{K}\right) \right\} \qquad (4)$$

and using eq.(1) we obtain

$$a_k(t) = \sum_{i=-\infty}^{\infty} A_{k,i} E^i = \sum_{b=1}^{B} \sum_{i=-\infty}^{\infty} H_{bk,i} E^i 2^{-(b-1)} \qquad (5)$$

Let us suppose now that b-th bit current of the DAC in accordance with Figure 3 of the paper [1] has a stationary noise component $i_{n,b}(t)$ and its referred to I_0 value is

$$\eta_b(t) = i_{n,b}(t)/I_0 \qquad (6)$$

Than referred to I_0 output DDS noise $\zeta_k(t)$ can be written in the form

$$\zeta_k(t) = \sum_{b=1}^{B} h_{bk}(t)\eta_b(t) \qquad (7)$$

If we substitute eq.(3) in (7), we obtain

$$\zeta_k(t) = \sum_{b=1}^{B} \sum_{i=-\infty}^{\infty} H_{bk,i} E^i \eta_b(t) \qquad (8)$$

Applying the procedure that was used in [1] to come from (23) to (24) and taking into consideration that all $\eta_b(t)$ are stationary random processes, we obtain from (8) a formula for n-th spectral function of the periodically nonstationary process $\zeta_k(t)$

$$S_n(\zeta_k,\zeta_k;\omega) = \sum_{b=1}^{B} \sum_{p=1}^{B} \sum_{i=-\infty}^{\infty} H^*_{bk,i} H_{pk,(n+i)} S_0(\eta_b,\eta_p;\omega - \frac{(n+i)\omega_c}{K}) \qquad (9)$$

Putting in eq.(9) n=0 gives a formula for two sided SD of the output DDS noise

$$S_0(\zeta_k,\zeta_k;\omega) = \sum_{b=1}^{B} \sum_{p=1}^{B} \sum_{i=-\infty}^{\infty} H^*_{bk,i} H_{pk,i} S_0(\eta_b,\eta_p;\omega - \frac{i\omega_c}{K}) \qquad (10)$$

To calculate SDs of amplitude and phase fluctuations one has to apply formulae (35-37) of the paper [1] using eqs.(9) and (10).

It is important to notice that formulae (9),(10) can be applied for any SDs matrix of the processes $h_b(t)$. There is no limitations to correlation times or SDs dependence on frequency. So they can be used for calculation of transformation to the DDS output both 1/f noises and fundamental noises.

Particular cases

Case 1: Noise sources are located in bit current generators. In this case $\eta_b(t)$ and $\eta_p(t)$ are not correlated if $b \neq p$. So $S_0(\eta_b,\eta_p;\omega) = 0$ if $b \neq p$ and from eqs.(9),(10) one obtains

$$S_n(\zeta_k,\zeta_k;\omega) = \sum_{b=1}^{B} \sum_{i=-\infty}^{\infty} H^*_{bk,i} H_{bk,(n+i)} S_0(\eta_b,\eta_b;\omega - \frac{(n+i)\omega_c}{K}) \qquad (11)$$

$$S_0(\zeta_k,\zeta_k;\omega) = \sum_{b=1}^{B} \sum_{i=-\infty}^{\infty} |H_{pk,i}|^2 S_0(\eta_b,\eta_b;\omega - \frac{i\omega_c}{K}) \qquad (12)$$

One can see that SD of the DDS output noise is a superposition of weighted SDs of the bits noise currents having frequency shifts ($i\omega_c/K$) with i= 0, ±1, ±2, ... ±(k-1), ±k, ±(k+1), ... To calculate DDS output noise one has to know SDs of bit noise currents and Fourier coeffitients of switching functions (4).

If the DDS has frequency step more than 5 - 10 kHz one can suppose that SDs of 1/f noises are very small if $\omega > \omega_c/2K$. In this case only the terms of the series (11),(12) that have $i = \pm k$ are not equal to zero in the vicinities of main spectral lines of the oscillation $a_k(t)$. Using this assumption, taking into consideration that $H_{bk,-k}=H_{bk,k}$ (it follows from (4)) and applying formulae (35-37) of the paper [1] together with eqs.(11),(12),one can obtain SDs of amplitude $\alpha_{fk}(t)$ and phase $\psi_{fk}(t)$ fluctuations of the output oscillation

$$S_0(\alpha_{fk},\alpha_{fk};\omega) = \frac{1}{|A_{k,k}|^2} \sum_{b=1}^{B} |H_{bk,i}|^2 S_0(\eta_b,\eta_b;\omega) \qquad (13)$$

$$S_0(\psi_{fk},\psi_{fk};\omega) = 0 \qquad (14)$$

It turns out that under these assumtions 1/f noise components of DAC bit currents cause only amplitude fluctuations.

Case 2: Noise sources are located in the DAC biasing circuit. In this case normalized noise currents $\eta_b(t)$ are fully correlated and they can be expressed in the form

$$\eta_b(t) = 2^{-(b-1)}\eta_u(t) \qquad (15)$$

where $\eta_u(t)$ is referred to I_0 noise current of the most significant bit caused by biasing voltage fluctuations. If we express $\zeta_k(t)$ in terms of $\eta_u(t)$ using (7) and (15) and take into consideration (1) we obtain

$$\zeta_k(t) = a_k(t)\eta_u(t) \qquad (16)$$

It follows from(16) and (5) that

$$S_0(\zeta_k,\zeta_k;\omega) = S_0(\eta_u,\eta_u;\omega) + |A_{k,k}|^2 \left[S_0\left(\eta_u,\eta_u;\omega - \frac{k\omega_c}{K}\right) + S_0\left(\eta_u,\eta_u;\omega + \frac{k\omega_c}{K}\right) \right] \qquad (17)$$

$$S_0(\alpha_{fk},\alpha_{fk};\omega) = S_0(\eta_u,\eta_u;\omega) \qquad (18)$$

$$S_0(\psi_{fk},\psi_{fk};\omega) = 0 \qquad (19)$$

One can see from (18),(19) that this noise source cause only amplitude fluctuations.

Noise due to Switching Times Fluctuations

Real switching functions of bit currents have rise times and fall times that are not equal to zero as it is shown in Figure 1.These switching times depend on the switching transistors parameters and if the parameters have 1/f fluctuations they have to cause switching time fluctuations and additional DDS output noise.

To estimate this noise we use a model with linear changing of the switches output currents during switching times as it is shown in Figure 2.We suppose that rise time of the b-th switch output current is equal to its fall time and designate it $T_{s,b}$.Than output current of the real switch $i_{T,bk}(t)$ can be related to the output current of the ideal switch $i_{bk}(t) = 2^{-(b-1)}I_0 h_{bk}(t)$ by symbolic equation [2]

$$i_{T,bk}(t) = 2^{-(b-1)}I_0 F_b(p) h_{bk}(t) \qquad (20)$$

where p is a symbol of differentiation with respect to time

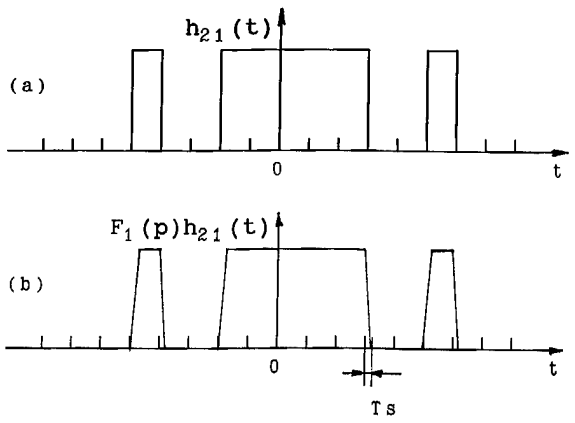


Fig.2 Ideal Switching Function (a) and a Model of Real Function Having Switching Time T_s (b)

and

$$F_b(p) = [\,1 - \exp(-pT_{s,b})]\,/\,pT_{s,b} \qquad (21)$$

is a symbolic transfer function of the switch model. It follows from (1),(5),(20),(21) that after replacing the ideal switch model by the model of real switch an output oscillation of the DAC $a_{t,k}(t)$ can be presented by Fourier series

$$a_{t,k}(t) = \sum_{b=1}^{B} 2^{-(b-1)} \sum_{i=-\infty}^{\infty} F_b(ji\omega_c / K) H_{bk,i} E^i \qquad (22)$$

where

$$F_b(ji\omega_c / K) = \frac{1-\exp(-ji\omega_c T_{s,b} / K)}{ji\omega_c T_{s,b} / K} \qquad (23)$$

Let us suppose that each parameter $T_{s,b}$ has an average value $T_{S,b}$ and small fractional fluctuations $\eta_{t,b}(t)$ that is

$$T_{s,b} = T_{S,b}\,[1 + \eta_{t,b}(t)] \qquad (24)$$

and these fluctuations are slowly varying ones. Than the oscillation (22) can be presented as a sum of its regular part $a_{T,k}(t)$ and random part $\zeta_{t,k}(t)$. Supposing also that switching times of all switches have equal average values $T_{S,b} = T_S$ but independant fluctuations we obtain from (22),(23),(24)

$$a_{T,k}(t) = \sum_{i=-\infty}^{\infty} F(ji\omega_c / K) A_{k,i} E^i \qquad (25)$$

where

$$F(ji\omega_c / K) = \frac{1-\exp(-ji\omega_c T_S / K)}{ji\omega_c T_S / K} \qquad (26)$$

and

$$\zeta_k(t) = \sum_{i=-\infty}^{\infty} \Delta(ji\omega_c / K) \sum_{b=1}^{B} 2^{-(b-1)} H_{bk,i} E^i \eta_{t,b}(t) \qquad (27)$$

where

$$\Delta(ji\omega_c/K) = \exp(-ji\omega_c T_S/K) - F(ji\omega_c/K) \qquad (28)$$

Comparing (27) and (8) we can see that their structures are similar. So if we suppose that $\eta_{t,b}(t)$ are stationary random processes having a matrix of two-sided SDs $S_0(\eta_{t,b},\eta_{t,p};\omega)$ in which only elements with equal indexes b=p are not equal to zero and apply the procedure that was used in [1] to come from (23) to (24), we obtain from (27) a formula for n-th spectral function of the periodically nonstationary process $\zeta_k(t)$

$$S_n(\zeta_k,\zeta_k;\omega) = \sum_{i=-\infty}^{\infty} \Delta^*(ji\omega_c / K)\Delta(j(i+n)\omega_c / K)\ x$$

$$x \sum_{b=1}^{B} 2^{-2(b-1)} H^*_{bk,i} H_{bk,(n+i)} S_0\left(\eta_{t,b},\eta_{t,b};\omega - \frac{(n+i)\omega_c}{K}\right) \qquad (29)$$

If we put in eq.(29) n=0 and introduce a designation

$$S_{t,i}(\omega) = \sum_{b=1}^{B} \left|H_{bk,i}\right|^2 2^{-2(b-1)} S_0(\eta_{t,b},\eta_{t,b};\omega) \qquad (30)$$

we can write a formula for two-sided SD of the output DDS noise due to switching times fluctuations as follows

$$S_0(\zeta_k,\zeta_k;\omega) = \sum_{i=-\infty}^{\infty} \left|\Delta(ji\omega_c / K)\right|^2 S_{t,i}\left(\omega - \frac{i\omega_c}{K}\right) \qquad (31)$$

To derive formulae for SDs of amplitude and phase fluctuations one has to use eqs.(35 - 37) of the paper [1] and eqs. (21 -31).The results look as follows

$$S_0(\alpha_{fk},\alpha_{fk};\omega) = \frac{(k\tau\cos k\tau - \sin k\tau)^2}{2(1-\cos k\tau)\left|A_{k,k}\right|^2} S_{t,k}(\omega) \qquad (32)$$

$$S_0(\psi_{fk},\psi_{fk};\omega) = \frac{(1-\cos k\tau - k\tau\sin k\tau)^2}{2(1-\cos k\tau)\left|A_{k,k}\right|^2} S_{t,k}(\omega) \qquad (33)$$

$$S_0(\alpha_{fk},\psi_{fk};\omega) = \frac{(k\tau\cos k\tau - \sin k\tau)(1-\cos k\tau - k\tau\sin k\tau}{2(1-\cos k\tau)\left|A_{k,k}\right|^2} S_{t,k}(\omega)$$

where

$$\tau = \omega_c T_S / K \qquad (34)$$

is a normalised switching time.

If the switching time is small that is

$$k\tau \gg 1 \text{ when } k \le K/4 \qquad (35)$$

one can use instead of (31 - 34) approximate formulae

$$S_0(\zeta_k,\zeta_k;\omega) = \sum_{i=-K/2}^{K/2} \frac{i^2\tau^2}{4} S_{t,i}\left(\omega - \frac{i\omega_c}{K}\right) \qquad (36)$$

$$S_0(\alpha_{fk},\alpha_{fk};\omega) = \frac{k^4\tau^4}{9\left|A_{k,k}\right|^2} S_{t,k}(\omega) \qquad (37)$$

$$S_0(\psi_{fk},\psi_{fk};\omega) = \frac{k^2\tau^2}{4\left|A_{k,k}\right|^2} S_{t,k}(\omega) \qquad (38)$$

$$S_0(\alpha_{fk},\psi_{fk};\omega) = \frac{k^3\tau^3}{6\left|A_{k,k}\right|^2} S_{t,k}(\omega) \qquad (39)$$

It is easy to see from (36 - 39) that switching time fluctuations give rise to phase fluctuations (38). Under conditition (35) their SD significantly exceeds SD of amplitude fluctuations caused by the same fluctuations source.To estimate fluctuations by means of eqs.(36 -39) one has to know an average switching time T_S, SDs of fractional fluctuations of this time,number k of the output oscillation and its amplitude $|A_{k,k}|$ and Fourier components $H_{bk,i}$ of switching functions (4).

1/f Fluctuations in DAC

In order to calculate noise level at the DDS output due to 1/f fluctuations sources one has to know SDs of referred to I_0 bit currents fluctuations $S_0(\eta_b,\eta_b;\omega)$ caused by internal sources of bit current generators,normalised current fluctuations caused by biasing circuit $S_0(\eta_u,\eta_u;\omega)$, and fractional switching time fluctuations $S_0(\eta_{t,b},\eta_{t,b};\omega)$. It

is useful to express these SDs in terms of 1/f fluctuations characteristics of basic circuit components of the DAC.

We suppose that 1/f fluctuations of resistors give small contribution to total noise and consider only noise sources in bypolar transistors and biasing circuit amplifier. To describe noise properties of bit current generator and biasing circuit we apply the 1/f noise model of transistor introduced in [3] and discussed and used in [4]. It is shown in Figure 3 There is one noise source in this model. It is noise current generator. Its current $i_f(t)$ has two-sided SD $S(i_f,i_f;\omega)$ that has to be defined by measurements and can be written as follows

$$S(i_f,i_f;\omega) = (I_B/I_{BR})^c\,(\omega_R/\omega)^r\,S_R\,(i_f,i_f;\omega_R) \qquad (40)$$

where I_B is an average base current, I_{BR} and ω_R are reference base current and frequency,respectively, c and r are exponents defined by approximation of measurements results, and $S_R(i_f,i_f;\omega_R)$ is a reference value of SD defined from measurements.

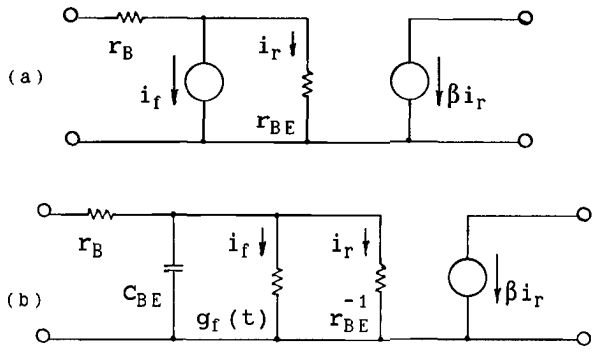


Fig.31/f Noise Models of Bypolar Transistor for Noise Currents Calculation (a) and for Calculation of Switching Time Fluctuations (b)

Using this model to calculate noise component of the b-th bit current generator of the DAC shown in Figure 4 of the paper [1] and supposing that biasing is provided from an ideal voltage source, one obtains

$$S_0(\eta_b,\eta_b;\omega) = I_0^2[(1+m_b)^2 S(i_{f,b},i_{f,b};\omega) + S(i_{fs,b},i_{fs,b};\omega)] \qquad (41)$$

where index b marks current setting transistor of the b-th bit and additional index s marks switching transistor, and $m_b = r_{B,b}/R_{E,b}$. If we suppose for estimations that m_b=m doesn't depend on b and

$$S(i_{f,b},i_{f,b};\omega) = S(i_{fs,b},i_{fs,b};\omega) = 2^{-(b-1)}\,S(i_{f,1},i_{f,1};\omega)$$

than instead of (41) we have

$$S_0(\eta_b,\eta_b;\omega) = I_0^2 2^{-(b-1)}[(1+m)^2+1]\,S(i_{f,1},i_{f,1};\omega) \qquad (42)$$

So to estimate DDS noise components due to internal fluctuations of bit current generators using (12),(13) and (42) we have to find from measurements at least $S(i_{f,1},i_{f,1};\omega)$.

To estimate DDS noise components due to biasing voltage fluctuations by means of (17),(18) we have to find $S_0(\eta_u,\eta_u;\omega)$.It is evident that $\eta_u(t) = v_{no}(t)/(R_{E1}I_0)$, where $v_{no}(t)$ is biasing voltage fluctuation in DAC schematic shown in Figure 4 of the paper [1].Using noise model of the biasing transistor shown in Figure 3,a and supposing that $v_{fA}(t)$ is referred to the amplifier input its 1/f noise voltage we obtain

$$S_0(\eta_u,\eta_u;\omega) = (R_{E1}I_0)^{-2}\{\,(R_E/R_{ref})^2\,S(v_{fA},v_{fA};\omega) + (1+m)^2 R_E^2 S(i_f,i_f;\omega)\,\} \qquad (43)$$

where $S(v_{fA},v_{fA};\omega)$ and $S(i_f,i_f;\omega)$ are two-sided SDs of the amplifier noise voltage and of noise current of the biasing transistor model. They have to be found from measurements.

To find switching time and its fluctuations we used charge control model of the switching transistor [6,7] and modified noise model that was introduced in [5].In this model noise current generator $i_f(t)$ is replaced by fluctuating conductivity $g_f(t)$. This model explains an amplifier phase noise dependence on its parameters and it is convenient to use together with nonlinear transistor model. At the definite operating point $i_f(t)= V_T g_f(t)$, where V_T is thermal voltage. So one can find SD $S(g_f,g_f;\omega)$ using measurements of $S(i_f,i_f;\omega)$.This way we arrived at the formula

$$T_S = (2/\omega_T)\,\{\,(1/\beta) + (V_s + 2V_T)/(r_{B,1}I_0)\,\}^{-1} \qquad (44)$$

where ω_T is a frequency where common emitter current gain extrapolates to unity, β is low frequency common emitter current gain that is assumed equal for all transistors, and V_s is equal to the half of difference between high and low levels at the digital inputs, and

$$S_0(\eta_{t,b},\eta_{t,b};\omega) = 2\left(\frac{1}{2\beta} + \frac{2V_T + r_{B,b}I_b/\beta}{V_s + 2V_T + r_{B,b}I_b/\beta}\right)^2 \frac{S(i_{fs,b},i_{fs,b};\omega)}{(I_b/\beta)^2} \qquad (45)$$

where $r_{B,b}$ is base resistance of the b-th bit switching transistor, I_b is an average b-th bit current.

So basing on the measurements of $S(i_{fs,b},i_{fs,b};\omega)$ and using eqs.(44),(45),(30 - 33) one can estimate fluctuations of this kind.

DAC Noise Measurements

To provide data for quantitative estimations some measurements of low frequency noise currents of the DAC K1118ПА3 were fulfilled. In this DAC I_0= 10 mA, R_{E1}= 170 Ohm, β= 100, R_E = 340 Ohm, R_{ref}= 250 Ohm. The results of three first bits noise currents measurements (b = 1,2,3) are presented in Figure 4. The results of one sided noise currents SDs measurements were divided by I_0^2. The digital inputs were biased using voltage source. So these measurements present a sum of doubled SDs (42) and (43):

$$S_{M1} = 2S_0(\eta_b,\eta_b;\omega) + 2S_0(\eta_u,\eta_u;\omega) \qquad (46)$$

To distinguish between different noise sources we measured also noise currents of the first and second bits with digital inputs biased using current source that was set so that the currents of switching transistors were equal. As one can see from noise model shown in Figure 3,a the measured SD of the first bit current divided by I_0^2 can be presented by the sum

$$S_{M2} = \frac{2S(i_{fs,1},i_{fs,1};\omega)}{(I_0/\beta)^2} + \frac{1}{4}[2S_0(\eta_1,\eta_1;\omega) + 2S_0(\eta_u,\eta_u;\omega)] \qquad (47)$$

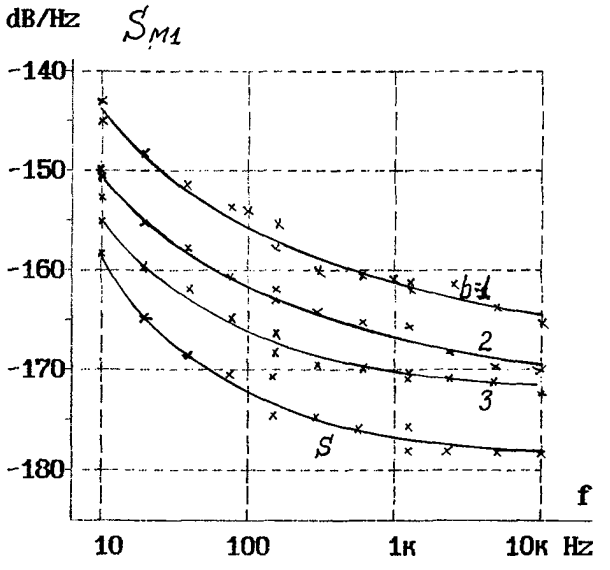


Fig.4 One-sided Spectral Densities of the Noise Currents of First Three Bits (b=1,2,3) with Digital Inputs Biased from Voltage Source, Corresponding to High Logical Level (s - sensitivity Curve)

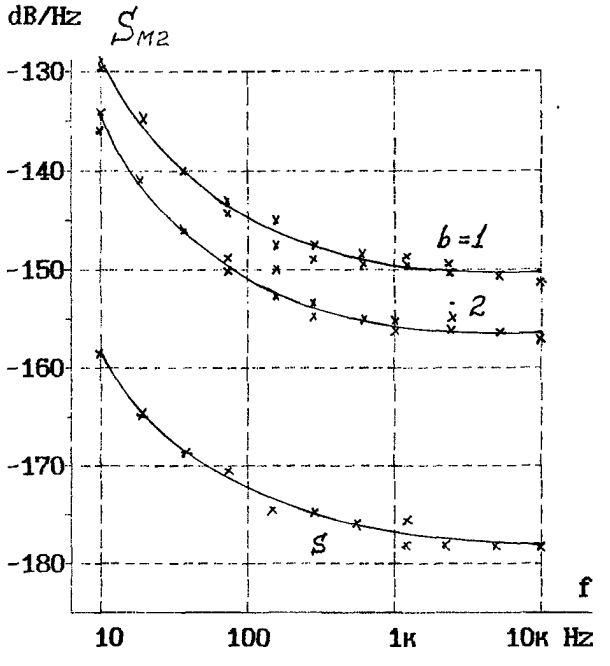


Fig.5 One-sided Spectral Densities of the Noise Currents of First Two Bits (b=1,2) with Digital Inputs Biased from Current Sources, Providing Equal Currents of Both Switching Transistors.

Results of these measurements are shown in Figure 5.

Comparing results shown in Figures 4 and 5, one can see that in the case when b=1 SD shown in Figure 5 in fact presents the first term of the sum (47). So using the data shown in Figure 5 we can obtain

$$\frac{2S(i_{fs,1},i_{fs,1};\omega)}{(I_0/\beta)^2} = 10^{-13}(\omega_R/\omega)^{1.5}+10^{-15} \quad [1/Hz] \qquad (48)$$

where $(\omega_R / 2\pi) = 10$ Hz is chosen as a reference frequency. The flat part of the approximation (48) within 3 dB gives a contribution of fundamental base recombination noise current [1].

Using (48) and (42) with $\beta = 100$ and m=0,3 we can estimate

$$2S_0(\eta_1,\eta_1;\omega) = 2{,}7\ 10^{-17}(\omega_R/\omega)^{1.5} \quad [1/Hz] \qquad (49)$$

(We have not taken into consideration a flat part of the function (48).)

Comparing (49) with the data shown in Figure 4 and using (46) we come to conclusion that the upper curve in Figure 4 can be considered as an estimation of $2S_0(\eta_u,\eta_u;\omega)$. It can be presented by approximation

$$2S_0(\eta_u,\eta_u;\omega) = 4\ 10^{-15}(\omega_R/\omega) + 4\ 10^{-17} \quad [1/Hz] \qquad (50)$$

One can see that the main contribution to this level is provided by the biasing circuit amplifier.

DDS Noise Estimations

To obtain some estimations we consider a DDS with DAC K1118ПА3 supposing that $\omega_c/2\pi = 32$ MHz, K=32, $T_S = 0.005$ μs.

Using (50) and (18) we obtain an estimation of the amplitude fluctuations one sided SD

$$S_0(\alpha_{fk},\alpha_{fk};\omega) = 4\ 10^{-15}(\omega_R/\omega) + 4\ 10^{-17} \quad [1/Hz] \qquad (51)$$

It doesn't depend on oscillation number k and significantly exceeds amplitude noise introduced by internal noise sources of bit current generators. At the flat part of the function (51) it matches with the estimation of fundamental noise presented in [1].

To do estimations of phase fluctuations we have to use eqs. (38),(34),(30),(45) and (48). We need also Fourier coefficients of the switching functions. These coefficients are presented in the Tables 1 and 2.

b \ i	1	2	3	4	5	6	7	8
1	0.317	0.031	0.102	0.031	0.056	0.029	0.035	0.028
2	0.245	0.031	0.081	0.031	0.155	0.029	0.122	0.028
3	0.135	0.031	0.175	0.031	0.017	0.030	0.100	0.028

Table 1. Absolute values of Fourier coefficients $|H_{bk,i}|$ k = 1, K = 32

b \ i	1	2	3	4	5	6	7	8
1	0.026	0.031	0.058	0.030	0.010	0.030	0.293	0.029
2	0.049	0.031	0.158	0.031	0.054	0.029	0.227	0.028
3	0.081	0.031	0.019	0.031	0.090	0.029	0.126	0.028

Table 2. Absolute values of Fourier coefficients $|H_{bk,i}|$ k = 7, K = 32

Using this formulae we obtain for this particular example of DDS an approximate expression

$2S_0(\psi_{fk}, \psi_{fk};\omega) = k^2\, 3.7\, 10^{-17} (\omega_R/\omega)^{-1.5}$ [1/Hz] (52)

The estimations (51) and (52) are shown in Figures 6 and 7. The levels of SDs due to fundamental noise sources that were obtained in the paper [1] and total SDs of amplitude and phase fluctuations are also shown.

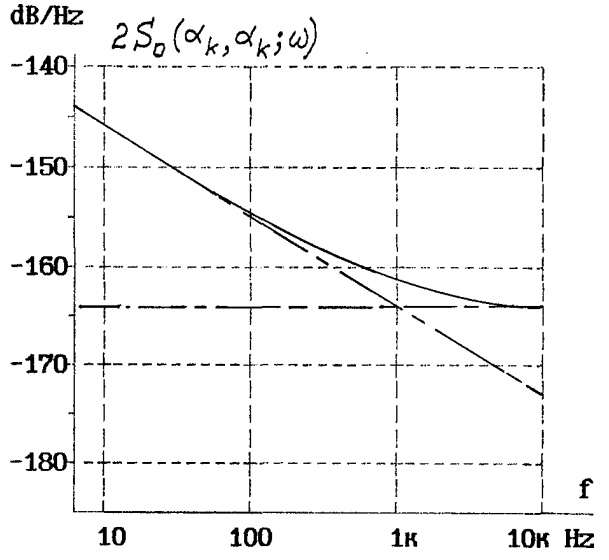


Fig.6 Estimations of One-sided Spectral Densities of DDS Fractional Amplitude Fluctuations

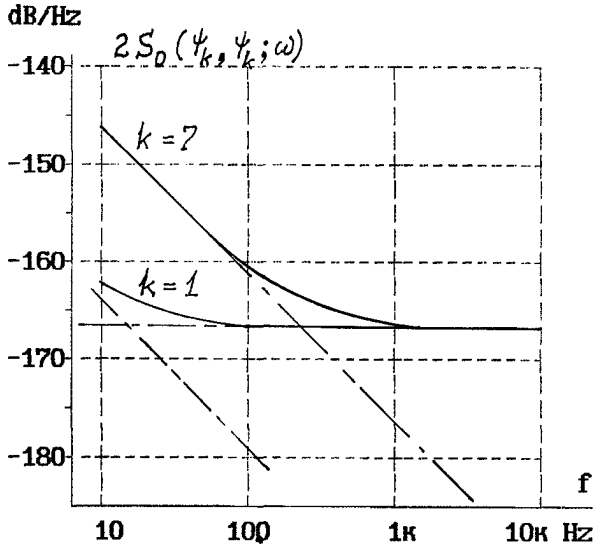


Fig 7 Estimations of One-sided Spectral Densities of DDS Phase Fluctuations for two Frequencies of Output Oscillations (k=1 and k=7)

Discussion

Results of the model analysis, measurements and estimations have shown how one can calculate DAC 1/f noises transformation to the DDS output and where the main fluctuations sources are located.

We have presented estimations only for amplitude and phase fluctuations. One can find spectral densities of the DDS output noise at any offset from the synthesized frequency using the formulae given in the paper.

The estimation of phase fluctuations that has been offered by V.Kroupa [8] lies within the range of the ones shown in Figure 7. But it is clear from our estimation that one has to specify a number of the synthesized frequency k to do such estimations.

Comparison the estimated DDS phase noise level with phase noises of the best quartz crystal oscillators [9] shows that if such oscillator is used as a clock frequency source than its phase noise can be lower than DDS phase noise at the frequencies that are higher than 100 Hz.

Conclusions

An approach to DDS noise analysis based on the presentation of output oscillation as a weighted sum of bit currents switching functions is developed and applied to the analysis of 1/f fluctuations transformation to the DDS output. General formulae that describe transformation of different sources of 1/f noises are obtained. But in contrast to fundamental noise analysis in this case one needs some results of 1/f noises measurements to calculate noise at the DDS output. That is why an approach to experimental definition of DAC main noise sources spectral densities was also developed and tested. Using results experiments with DAC К1118ПА3 some estimations of noise at the DDS output were obtained and discussed.

References

[1] V.N.Kuleshov, and H.Y.Liu, "Fundamental Noise in Direct Digital Frequency Synthesizers", Proc. 1995 IEEE Freq.Control Symp.

[2] М.В.Капранов, В.Н.Кулешов, Г.М.Уткин, "Теория колебаний в радиотехнике", Наука, Москва, 1984

[3] J.L.Plumb, and E.R.Chenette, "Flicker Noise in Transistors", IEEE Trans. ED-10, 1963, No.5, pp.304-308

[4]A.van der Ziel, "Noise: Sources, Characterization, Measurement", Prentice-Hall Inc., Englewood Cliffs, N.J. 1970

[5] В.Н.Кулешов, И.П.Бежняк, "Фликкер-шум в транзисторах и флуктуации амплитуды и фазы в высокочастотных усилителях", Радиотехника и электроника, 1980, т.25,No.11, сс.2393-99

[6] J.J.Sparkes, and R.Beaufoy, "The Junction Transistor as a Charge Control Device", Proc.IRE, 1957, v.45, No.12 pp.1740-42

[7] P.E.Gray, and C.E.Searle,"Electronic Principles: Physics, Models and Circuits", John Wiley & Sons, N.Y., 1969

[8] V.F.Kroupa, "Discrete Spurious Signals and Background Noise in Direct Digital Frequency Synthesizers", Proc.1993 IEEE Int.Freq. Control Symposium, pp.242-250

[9] M.M.Driscoll, "Reduction of Quartz Crystal Oscillator Flicker-of-Frequency and White Noise (Floor) Levels and Acceleration Sensitivity via Use of Multiple Resonators", Proc.1992 IEEE Freq. Control Symposium, pp.334-339

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

FUNDAMENTAL NOISE IN DIRECT DIGITAL FREQUENCY SYNTHESIZERS

V.N.Kuleshov and H.Y.Liu

Moscow Power Engineering Institute,
14 Krasnokasarmennaya, Moscow 111250, Russia

Abstract: An analysis of fundamental noise at the output of direct digital frequency synthesizer (DDS) due to noise sources in the digital-to-analog converter (DAC) is presented. Basing on the periodically nonstationary random processes theory, general formulae for spectral densities (SDs) of both noise floor and amplitude and phase fluctuations are obtained. These formulae are simplified for the case of δ-correlated noises and for the noise sources that have SDs proportional to the output signal value and to the square of the signal value.

For the particular example of DDS, having high speed ECL type DAC, detailed investigation is fulfilled. Using noise models of bypolar transistors and other components, simple formulae for two groups of noise sources are derived. Noise sources of the first group are located in the DAC bit currents generators. Noise sources of the second group are in the circuit of biasing and temperature stabilization of bit currents generators.

Using these formulae and real values of supply voltages, currents, transistors parameters and resistors, some calculations are fulfilled, and quantitative estimations of the output noise SD and both amplitude and phase fluctuations SDs are obtained and discussed.

Introduction

Sine output direct digital frequency synthesizers [1] are now widely used both as output oscillations sources and as parts of microwave frequencies synthesizers for different applications [6,12].Although main problems of the DDSs are connected with spur reduction [1 - 11] there are some applications in which the additional noise inserted by DDS has to be estimated and controlled to meet basic requirements to the system [12]. In some specific structures of DDS [4,5] spurs levels can be much lower than in usual structures [1,2]. If DDS is used as a part of indirect synthesizer the spurs can be put down by phase lock loop filter. This can not be done with respect to low frequency phase noise contributed by DDS. As quartz crystal oscillators phase noise levels were made much lower during last ten years [13], DDS phase noise can be higher than the one of the clock oscillator. So an analysis of noise levels that appear at the DDS output due to its internal fluctuations sources seems rather important problem.

An estimation of DDS phase noise based on the results of experimental investigations of frequency dividers is given in [11].Results of phase noise measurements for microwave "switch and mix" synthesizer having 500 MHz-clocked DDS in its architecture are presented in [12].But we have not found neither DDS noise theoretical analysis nor experimental data for some particular DDS structure.

The aim of this paper is to present theoretical estimation of the noise at the output of the DDS due to fluctuations sources that are located in the DAC and are connected with fundamental properties of charge carriers motion in the DAC components: resistors, transistors and so on. In fact we consider only the sources of thermal noise, shot noise and recombination noise. These sources have two important features: their PSDs can be found from theoretical models [14 -16] and for all of them a model of δ-correlated noise can be used as a good approximation.

DDS Model with Noise Sources

The DDS under consideration block diagram is shown in Figure 1 [1,10,11,12]. The phase accumulator is a linear

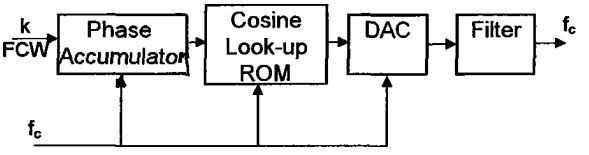


Fig.1. Cosine Output DDS Block Diagram

modulo counter which increments with each clock cycle at a rate dependent on the frequency control word (FCW) k. By means of a ROM algorithm and DAC the linear phase ramp is converted to a stepped cosine oscillation that is shown in Figure 2,a. The DDS output oscillation is obtained at the filter output. Its frequency is

$$f_k = kf_c / K \qquad (1)$$

where $K = 2^R$ is the accumulator capacitance and usually $k \leq K/4$.

Suppose that the DAC has a switched current configuration [18 - 21]. The DAC model is shown in Figure 3. It contains B bit current generators and B

0-7803-2500-1/95 $4.00 © 1995 IEEE

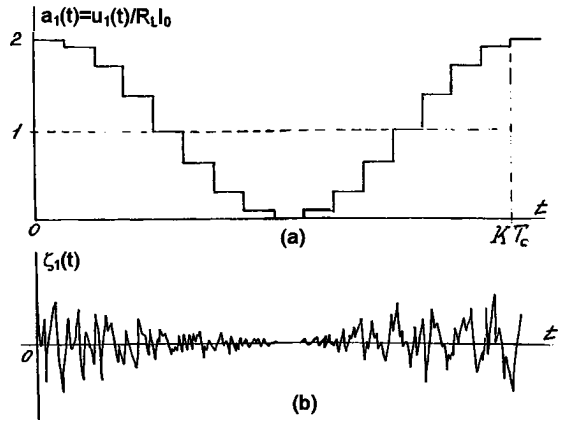


Fig.2. Stepped Cosine Oscillation (a) and Fundamental Noise (b) at the DAC output

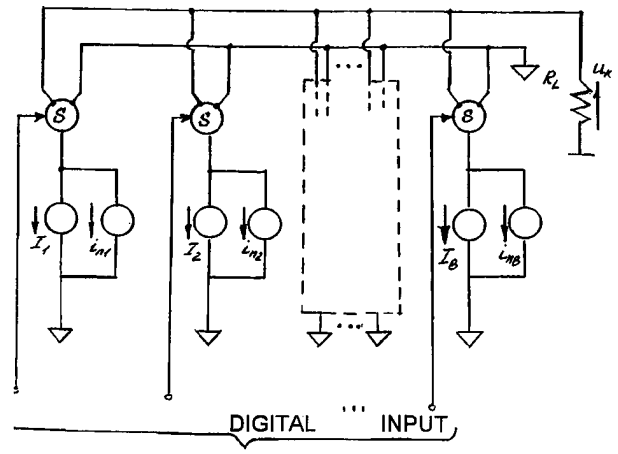


Fig.3. DAC Model with Noise Sources

switches controlled by B-bit input word. We assume that each current generator has an average value I_b and a noise component $i_{n,b}$ (b = 1,2,... B). As usually

$$I_b = 2^{b-1} I_0 \quad (2)$$

In real DAC noise currents can be divided into two groups

$$i_{n,b} = i_{n1,b} + i_{n2,b} \quad (3)$$

Noise currents of the first group $i_{n1,b}$ are caused by internal fluctuation sources of each bit current generator. So there is no correlation between these noise currents of different bits. Noise currents of the second group $i_{n2,b}$ are caused by common biasing voltage fluctuations v_{n0}. So they are fully correlated and can be presented as follows

$$i_{n2,b} = 2^{b-1} G v_{n0} \quad (4)$$

The transconductance G of the most significant bit (MSB) current generator can be calculated for the particular DAC.

The regular part of the DAC output voltage $u_k(t)$ across the load resistance R_L can be written in the form

$$u_k(t) = R_L I_0 a_k(t), \quad (5)$$

where

$$a_k(t) = \sum_{m=0}^{K-1} a_{km} d_m(t) \quad (6)$$

$$a_{km} = 1 + \cos(2\pi km / K), \quad (7)$$

$$d_m(t) = \sum_{w=-\infty}^{\infty} [\sigma[t-(m+wK)T_c - \sigma[t-(m+wK+1)T_c]] \quad (8)$$

$$\sigma = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (9)$$

The noise component of the DAC output voltage $u_{n,k}(t)$ can be presented in the form

$$u_{n,k}(t) = R_L I_0 \zeta_k(t), \quad (10)$$

where

$$\zeta_k(t) = \sum_{m=0}^{K-1} \xi_{km}(t, a_{km}) d_m(t), \quad (11)$$

and $\xi_{km}(t,a_{km})$ - normalised noise voltage corresponding to the normalized regular voltage a_{km} (7).

Now the aim of this paper can be formulated more exactly. It is necessary to find:

1) SD of the normalized noise $\zeta_k(t)$ at the output of DAC,

2) SDs of the output oscillation phase fluctuations and fractional amplitude fluctuations that arise due to addition $u_{n,k}(t)$ (10) to $u_k(t)$ (5).

General Formulae for Oscillation Spectrum and Output Noise SD

It is known that Fourier expansion of the oscillation (6) has the form

$$a_k(t) = \sum_{i=-\infty}^{\infty} A_{k,i} E^i \quad (12)$$

where

$$A_{k,i} = \begin{cases} 1, & if \quad i = 0 \\ \dfrac{\sin(\pi i / K)}{2\pi i / K}, & if \quad i = NK \pm k, N = 0, \pm 1, \pm 2, \ldots \\ 0, & if \quad i \neq 0, i \neq NK \pm k \end{cases} \quad (13)$$

$$E = \exp[j\{(\omega_c t / K) + \Phi_0\}] \quad (14)$$

$$\omega_c = 2\pi f_c = 2\pi / T_c \quad (15)$$

$$\Phi_0 = -\pi / K \quad (16)$$

So the useful component $a_{k,k}(t)$ of the sum (12) that has a frequency f_k (1) is

$$a_{k,k}(t) = A_{k,k} E^k + A_{k,-k} E^{-k} \quad (17)$$

The random process $\zeta_k(t)$ (11) is periodically nonstationary one [16]. Its autocorrelation function

$$B(\zeta_k \zeta_k; t, \tau) = \overline{\zeta_k^*(t) \zeta_k(t+\tau)} \quad (18)$$

(where ζ^* denotes the complex conjugate of ζ and the line over the product means the ensemble average) is a periodic function of t with a period K/f_c. It can be presented by Fourier series

$$B(\zeta_k, \zeta_k; t, \tau) = \sum_{n=-\infty}^{\infty} B_n(\zeta_k, \zeta_k; \tau) E^n \quad (19)$$

It means that one has to know all Fourier components $B_n(\zeta_k,\zeta_k;\tau)$ or their Fourier transforms

$$S_n(\zeta_k, \zeta_k; \omega) = \int_{-\infty}^{\infty} B_n(\zeta_k, \zeta_k; \tau) \exp(j\omega\tau) d\tau \quad (20)$$

to describe the periodically nonstationary process in the

frame of the correlation theory [16]. If n = 0, the formula (20) presents two-sided SD of the process $\zeta_k(t)$.

In order to find formulae for calculation of $S_n(\zeta_k,\zeta_k;\omega)$ let us replace peroidic functions $d_m(t)$ (8) with their Fourier expansions

$$d_m(t) = \sum_{i=-\infty}^{\infty} D_{mi} E^i \tag{21}$$

where

$$D_{mi} = \frac{\sin(\pi i / K)}{\pi i} \exp(-\frac{j2\pi mi}{K}) \tag{22}$$

in the sum (11). We obtain

$$\zeta_k(t) = \sum_{m=0}^{k-l} \sum_{i=-\infty}^{\infty} D_{mi} E^i \xi_{km}(t, a_{km}) \tag{23}$$

As the processes $\xi_{km}(t,a_{km})$ are stationary ones (for any constant value a_{km}) they can be characterised by their SDs matrix $S_0(\xi_{km},\xi_{kp};\omega)$.Using results of the work [17] and formula (23) one can obtain basic general formulae

$$S_n(\zeta_k,\zeta_k;\omega) = \sum_{m=0}^{K-l} \sum_{p=0}^{K-l} \sum_{i=-\infty}^{\infty} D^*_{mi} D_{p(n+i)} S_0(\xi_{km},\xi_{kp};\omega - \frac{(n+i)\omega_c}{K}) \tag{24}$$

General formula for two-sided SD of the normalised noise at the DAC output follows from (24) if we put n=0

$$S_0(\zeta_k,\zeta_k;\omega) = \sum_{m=0}^{K-l} \sum_{p=0}^{K-l} \sum_{i=-\infty}^{\infty} D^*_{mi} D_{pi} S_0(\xi_{km},\xi_{kp};\omega - \frac{i\omega_c}{K}) \tag{25}$$

So the problem of this SD calculation is reduced to the finding of SDs $S_0(\xi_{km},\xi_{kp};\omega)$ and summation.

General Formulae for SDs of Phase and Fractional Amplitude Fluctuations

In order to calculate phase and amplitude fluctuations, let us present the process $\zeta_k(t)$ by a series that is similar to Fourier expansion

$$\zeta_k(t) = \sum_{i=-\infty}^{\infty} Z_{k,i}(t) E^i \tag{26}$$

where $Z_{k,i}(t)$ is a complex random envelope of the process that is obtained from $\zeta_k(t)$ after its passing through an ideal filter having a passband $\{(i-0.5)\omega_c / K, (i+0.5)\omega_c / K\}$. Then a part of the sum (26)

$$\zeta_{k,k}(t) = Z_{k,-k}(t) E^{-k} + Z_{k,-k}(t) E^k \tag{27}$$

contains all spectral components laying near frequencies ($\pm k\omega_c/K$).

If we consider a sum

$$c_{k,k}(t) = a_{k,k}(t) + \zeta_{k,k}(t) \tag{28}$$

as an oscillation that has fractional amplitude fluctuations $\alpha_k(t)$ and phase fluctuations $\psi_k(t)$, suppose that $|\overline{Z_{k,k}(t)}|^2 << A^2_{k,k}$ and substitute (17) and (27) in (28), we obtain

$$\alpha_k(t) = \frac{Z_{k,-k}(t) + Z_{k,k}(t)}{2|A_{k,k}|} \tag{29}$$

$$\psi_k(t) = j\frac{Z_{k,-k}(t) - Z_{k,k}(t)}{2|A_{k,k}|} \tag{30}$$

It was shown in [17] that

$$S_0(Z_{k,k},Z_{k,k};\omega) = S_0(\zeta_k,\zeta_k;\omega + k\omega_c / K) \tag{31}$$

$$S_0(Z_{k,-k},Z_{k,-k};\omega) = S_0(\zeta_k,\zeta_k;\omega - k\omega_c / K) \tag{32}$$

$$S_0(Z_{k,k},Z_{k,-k};\omega) = S_{-2k}(\zeta_k,\zeta_k;\omega - k\omega_c / K) \tag{33}$$

$$S_0(Z_{k,-k},Z_{k,k};\omega) = S_{2k}(\zeta_k,\zeta_k;\omega + k\omega_c / K) \tag{34}$$

where (- 0,5 ω_c / K) $\le \omega \le$ (0,5 ω_c / K). Using (29), (30) and (31 - 34) one can obtain formulae for two-sided SDs of fractional amplitude fluctuations, phase fluctuations and their cross-spectral density

$$S_0(\alpha_k,\alpha_k;\omega) = \frac{1}{4|A_{k,k}|^2}[S_0(\zeta_k,\zeta_k;\omega + k\omega_c / K) + S_0(\zeta_k,\zeta_k;\omega - k\omega_c / K) + S_{-2k}(\zeta_k,\zeta_k;\omega - k\omega_c / K) + S_{2k}(\zeta_k,\zeta_k;\omega + k\omega_c / K)] \tag{35}$$

$$S_0(\psi_k,\psi_k;\omega) = \frac{1}{4|A_{k,k}|^2}[S_0(\zeta_k,\zeta_k;\omega + k\omega_c / K) + S_0(\zeta_k,\zeta_k;\omega - k\omega_c / K) - S_{-2k}(\zeta_k,\zeta_k;\omega - k\omega_c / K) - S_{2k}(\zeta_k,\zeta_k;\omega + k\omega_c / K)] \tag{36}$$

$$S_0(\alpha_k,\psi_k;\omega) = \frac{1}{4|A_{k,k}|^2}[S_0(\zeta_k,\zeta_k;\omega - k\omega_c / K) - S_0(\zeta_k,\zeta_k;\omega + k\omega_c / K) + S_{-2k}(\zeta_k,\zeta_k;\omega - k\omega_c / K) - S_{2k}(\zeta_k,\zeta_k;\omega + k\omega_c / K)] \tag{37}$$

One can notice that to find these spectral densities one has to know not only $S_0(\zeta_k,\zeta_k;\omega)$ (25) but also $S_{2k}(\zeta_k,\zeta_k;\omega)$ and $S_{-2k}(\zeta_k,\zeta_k;\omega)$ (24).

Spectral Densities of Fundamental Noises at the DDS Output

Fundamental noises have usually a correlation time that is much smaller than $T_c = 1/f_c$. So one can put

$$S_0(\xi_{km},\xi_{kp};\omega) = 0, \quad \text{if} \quad m \ne p \tag{38}$$

Using (38), we obtain from (24), (25)

$$S_n(\zeta_k,\zeta_k;\omega) = \sum_{m=0}^{K-l} \sum_{i=-\infty}^{\infty} D^*_{mi} D_{m(n+i)} S_0(\xi_{km},\xi_{km};\omega - \frac{(n+i)\omega_c}{K}) \tag{39}$$

$$S_0(\zeta_k,\zeta_k;\omega) = \sum_{m=0}^{K-l} \sum_{i=-\infty}^{\infty} |D_{mi}|^2 S_0(\xi_{km},\xi_{km};\omega - \frac{i\omega_c}{K}) \tag{40}$$

In some cases one can suppose that $\xi_{km}(t)$ are white noises. Then

$$S_0(\xi_{km},\xi_{km};\omega - \frac{(n+i)\omega_c}{K}) = S_0(\xi_{km},\xi_{km}) \tag{41}$$

Using (41) in (39),(40) and summing the series up, we obtain

$$S_n(\zeta_k,\zeta_k;\omega) = \frac{\sin(\pi n / K)}{\pi n} \sum_{m=0}^{K-l} \exp(\frac{-j2\pi mn}{K}) S_0(\xi_{km},\xi_{km}) \tag{42}$$

$$S_0(\zeta_k,\zeta_k;\omega) = \frac{1}{K} \sum_{m=0}^{K-l} S_0(\xi_{km},\xi_{km}). \tag{43}$$

The last formula shows that in this case SD of the output noise is an average value of the DAC noises Sds,

corresponding to the values a_{km} that are used during the period KT_c. Substituting (42),(43) in (35-7),we arrive to the formulae

$$S_0(\alpha_k,\alpha_k;\omega) =$$

$$\frac{1}{2K\left|A_{k,k}\right|^2}\sum_{m=0}^{K-1}\left[1+\frac{\sin(2\pi k/K)}{2\pi k/K}\cos\left(\frac{4\pi km}{K}\right)\right]S_0(\xi_{km}\xi_{km}) \quad (44)$$

$$S_0(\psi_k,\psi_k;\omega) =$$

$$\frac{1}{2K\left|A_{k,k}\right|^2}\sum_{m=0}^{K-1}\left[1-\frac{\sin(2\pi k/K)}{2\pi k/K}\cos\left(\frac{4\pi km}{K}\right)\right]S_0(\xi_{km}\xi_{km}) \quad (45)$$

$$S_0(\alpha_k,\psi_k;\omega) =$$

$$\frac{1}{2K\left|A_{k,k}\right|^2}\sum_{m=0}^{K-1}\left[\frac{\sin(2\pi k/K)}{2\pi k/K}sin\left(\frac{4\pi km}{K}\right)\right]S_0(\xi_{km}\xi_{km}) \quad (46)$$

There are two particular cases of $S_0(\xi_{km},\xi_{km})$ dependence on a_{km} that are interesting for practical calculations.

Case 1: $S_0(\xi_{km},\xi_{km})$ is proportional to a_{km} . That is

$$S_0(\xi_{km},\xi_{km}) = a_{km}S_0(\xi,\xi) \quad (47)$$

where $S_0(\xi,\xi)$ is SD of some stationary white noise.(Such situation takes place if each noise current in the model shown in Figure 3 is a sum of elementary independent noise currents.) Substituting (47) and (7) in (43 - 46),we obtain after summation

$$S_0(\zeta_k,\zeta_k) = S_0(\xi,\xi) \quad (48)$$

$$S_0(\alpha_k,\alpha_k;\omega) = S_0(\psi_k,\psi_k;\omega) = \frac{2(\pi k/K)^2}{\sin^2(\pi k/K)}S_0(\xi,\xi) \quad (49)$$

$$S_0(\alpha_k,\psi_k;\omega) = 0 \quad (50)$$

So the problem is reduced to calculation of $S_0(\xi,\xi)$. It can be done for particular DAC model.

Case 2 : $S_0(\xi_{km},\xi_{km})$ is proportional to a_{km}^2. It takes place when there is a common noise $\eta(t)$ that modulates all bit currents.Than

$$\zeta_k(t) = \sum_{m=0}^{K-1}a_{km}d_m(t)\eta(t) \quad (51)$$

where $\eta(t)$ is a stationary white noise that has SD $S_0(\eta,\eta)$. Comparing (11) and (51) we can see that $\xi_{km}(t,a_{km}) = a_{km}\eta(t)$ and therefore

$$S_0(\xi_{km},\xi_{km}) = a_{km}^2S_0(\eta,\eta) \quad (52)$$

Substituting (52) in (43 - 46), using (7) and summing up, we obtain

$$S_0(\zeta_k,\zeta_k;\omega) = \frac{3}{2}S_0(\eta,\eta) \quad (53)$$

$$S_0(\alpha_k,\alpha_k;\omega) = \frac{(\pi k/K)^2}{\sin^2(\pi/K)}\left(3+\frac{\sin(2\pi k/K)}{2(2\pi k/K)}\right)S_0(\eta,\eta) \quad (54)$$

$$S_0(\psi_k,\psi_k;\omega) = \frac{(\pi k/K)^2}{\sin^2(\pi/K)}\left(3-\frac{\sin(2\pi k/K)}{2(2\pi k/K)}\right)S_0(\eta,\eta) \quad (55)$$

$$S_0(\alpha_k,\psi_k;\omega) = 0 \quad (56)$$

So in this case the problem is reduced to calculation of $S_0(\eta,\eta)$.

If a model of white noises is not valid, one has to use the formulae (39),(40) and (35 - 37).

Estimations

To obtain quantitative estimations we consider a DDS with DAC K1118ПА3 [18,19].A simplified schematic of the DAC is shown in Figure 4. Noise models of the b-th

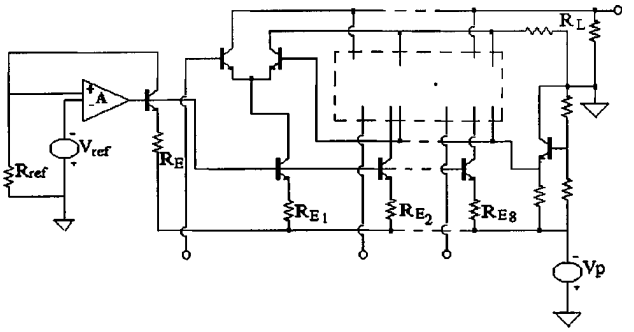


Fig.4. Simplified Schematic of the DAC K1118ПА3

bit current generator with its current switch and of the bias circuit are shown in Figure 5 and Figure 6.

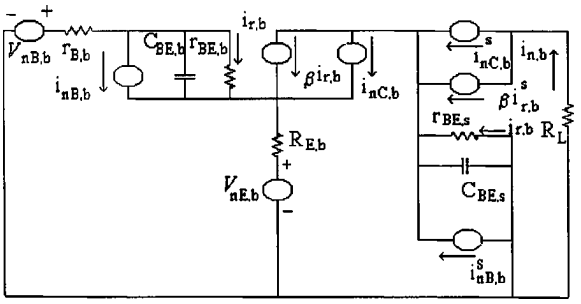


Fig. 5 Noise Model of the b-th Bit Current Gtnerator

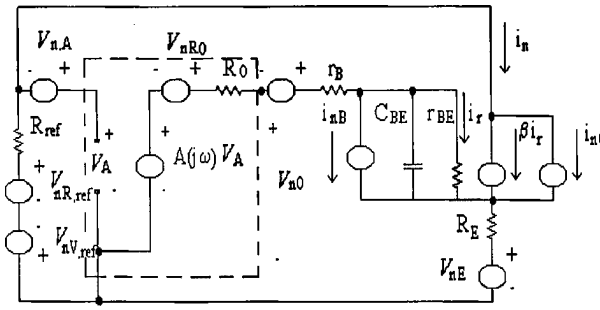


Fig. 6 Noise Model of the DAC Bias Circuit

Using noise model shown in Figure 5, we obtained a simplified formula for SDs of noise currents of the first group $i_{n1,b}(t)$ that were introduced in the sum (3). Deriving this formula, we assumed that $V_E = I_bR_{E,b}$ and $m = r_{B,b}/R_{E,b}$ did not depend on b, all transistors of DAC had identical low frequency common emitter current gains β and $\beta \gg 1$. Under these assumptions we have

$$S_0(i_{n1,b},i_{n1,b};\omega) = qD(\omega)I_b \quad (57)$$

where

$$D(\omega)=\frac{2(1+m)(V_T/V_E)}{1+(1+m)^2(\omega/\varpi_\alpha)^2}+\frac{(1+m)^2[(1/\beta)+(\omega/\omega_\alpha)^2}{1+(1+m)^2(\omega/\omega_\alpha)^2}+\frac{(1/\beta)+(\omega/\omega_\alpha)^2)}{1+(\omega/\omega_\alpha)^2} \quad (58)$$

V_T = k_BT/q, k_B is Boltzmann constant, T is absolute temperature, q is magnitude of the electronic charge, I_b is average b-th bit current (2), ω_α is the α cutoff frequency of the DDS transistors.

If we assume also that $\omega_c \ll \omega_\alpha$,then $D(\omega)$ can be replaced by

$$D = D(0) = 2(1+m)(V_T/V_E)+[(1+m)^2+1]/\beta \quad (59)$$

and instead of (57) we can write

$$S_0(i_{n1,b}, i_{n1,b};\omega) = qDI_b \quad (60)$$

As internal noise sources of different bits are independent it follows from (2),(60),(5),(6),(10),(11) that

$$S_0(\xi_{km},\xi_{km}) = a_{km}(qD/I_0) \quad (61)$$

Comparing (61) and (47) we can see that we deal with the *case 1* and

$$S_0(\xi\xi) = qD/I_0 \quad (62)$$

Using (62),(48),(49) we can estimate noise level at the DDS output due to noise sources of the first group. Supposing that I_0= 10 mA, β = 100, m = 0.3, V_E = 1.7 V, V_T = 25 mV, we obtain an estimation of *one sided* SDs of the output noise

$2S_0(\zeta_k,\zeta_k) = 2.1\,10^{-18}\,1/Hz = -177 dB/Hz$

$2S_0(\alpha_k,\alpha_k;\omega) = 2S_0(\psi_k,\psi_k;\omega) \approx$ - 174dB/Hz (k/K < 1/4)

To estimate a contribution of the noise sources of the second group (4) we have to find SD of the biasing circuit noise voltage $v_{n0}(t)$.Basing on the model shown in Figure 6 we can write a symbolic formula

$$v_{n0} = \frac{A(j\omega)(R_{ref}i_n + v_n) + v_{n,R0}}{1+A(j\omega)R_{ref}y_{21}(j\omega)} \quad (63)$$

where $A(j\omega)$ is a symbolic transfer function of the amplifier, i_n is noise component of the collector current of the biasing circuit transistor, $v_n = v_{n,Rr} - v_{n,Vr} + v_{n,A}$, $v_{n,Rr}$ is thermal noise voltage of R_{ref}, $v_{n,Vr}$ is noise component of the reference voltage, $v_{n,A}$ is the amplifier noise voltage referred to its input, $v_{n,Ro}$ is thermal noise voltage of R_o, $y_{21}(j\omega)$ is symbolic transconductance of the transistor with feedback resistor R_E. Supposing that $|A(j\omega)| \gg 1$ for all DDS output frequencies, $R_o \ll r_B$, $\omega_c \ll \omega_\alpha$ and $v_{n,Vr} = v_{n,A} = 0$, we can reduce an equation (63) to

$$v_{n0} = R_E i_n + v_{n,Rr} \quad (64)$$

Under these assumptions SD of the noise current i_n can be written in the form that is similar to (59)

$$S_0(i_n, i_n) = qI_c[2(1+m)(V_T/V_E)+(1+m)^2/\beta \quad (65)$$

Two sided SD of the noise voltage $v_{n,Rr}$ can be presented in the form

$$S_0(v_{n,Rr}, v_{n,Rr}) = 2k_BTR_{ref} = q2R_{ref}V_T \quad (66)$$

In order to find fractional noise current $i_{n2,b}/I_0$ for the DAC shown in Figure 5 we can use eq.(4) taking into consideration that $G \approx 1/R_{EI}$ and $I_0R_{EI} = V_E$. We then find

$$i_{n2,b}/I_0 = 2^{b-1}(v_{n0}/V_E) \quad (67)$$

On comparing (67) with (2) and taking into consideration (5),(6),(10) we see that we deal with the *case 2* and the common noise in (51) is

$$\eta(t) = v_{n0}(t)/V_E \quad (68)$$

Using (68),(64) and (65),(66) we obtain

$$S(\eta,\eta) = \frac{q}{I_c}\left[\frac{2(1+m)V_T}{V_E}+\frac{(1+m)^2}{\beta}+\frac{2R_EV_T}{R_{ref}V_E}\right] \quad (69)$$

Now we can use eqs.(53 - 55) to estimate DDS output noise level due to biasing circuit fluctuations sources. The DAC K1118ПА3 has $I_c = I_0/2$ = 5 mA, R_E= 340 Ohm, V_E= 1.7 V, R_{ref}= 250 Ohm.We can suppose that m=0.3 and V_T= 25 mV, (k/K=1/32) and using eqs. (53 - 55), (69) arrive at the estimations of one-sided SDs

$2S_0(\zeta_k,\zeta_k) = 9.1\,10^{-18}$ 1/Hz = - 170.4 dB/Hz

$2S_0(\alpha_k,\alpha_k) = 2.12\,10^{-17}$ 1/Hz = - 166.7 dB/Hz

$2S_0(\psi_k,\psi_k) = 1.52\,10^{-17}$ 1/Hz = - 168.2 dB/Hz

It is interesting to estimate also SD of the fractional thermal noise $\zeta_T = v_{n,RL}/(R_LI_0)$ of the load resistance. It is easy to find that

$$S_0(\zeta_T,\zeta_T) = (2q/I_0)(R_{EI}/R_L)(V_T/V_E) \quad (70)$$

If we suppose R_{EI} = 170 Ohm and R_L = 100 Ohm than

$2S_0(\zeta_T,\zeta_T) = 1.6*10^{-18}$ 1/Hz = -178 dB/Hz

Summing up the estimations of all noises that were considered in this paper we obtain

$2S_0(\zeta_k,\zeta_k)_{total}=1.28\,10^{-17}$ 1/Hz = - 169 dB/Hz

$2S_0(\alpha_k,\alpha_k)_{total}=2.86\,10^{-17}$ 1/Hz = - 165.5 dB/Hz

$2S_0(\psi_k,\psi_k)_{total}=2.26\,10^{-17}$ 1/Hz = -166.5 dB/Hz

Discussion

An estimation of DDS white phase noise presented by V.Kroupa [11] is -165 dB/Hz. Our estimation is close to this one. But analysing the formulae that were used one can see that results of calculations depend on the values of I_0,I_c,V_E,β and so on and our estimations could be quite different for other set of the values. It is worth to remind that we assumed that the biasing circuit amplifier did not have an internal noise.

It is necessary to discuss the assumptions under which the models of white noises were used. Comparing (57),(58) and (60),(59) one can see that if the biggest value of ω_k is more than $\omega_\alpha/\sqrt{\beta}$ the dependence of SDs of bit currents noise components on ω has to be taken into consideration and one has to use more general formulae (39),(40) instead of (42),(43). Comparing (63) and (64) one can see that if the amplifier unity gain frequency is less than ω_k the dependence of biasing circuit noise SD on frequency can not be neglected.

We did not pay attention so far to phase noise of clock frequency oscillator. Phase noise floor levels that were achieved in 100-MHz quartz crystal oscillators [13] are less than - 170 dB/Hz. So fundamental phase noise of DDS can be higher than phase noise floor of clock oscillator.

Although we derived all formulae basing on the simplest DDS structure shown in Figure 1 the results of this paper are valid for DDSs based on using Walsch functions [4,5].

Conclusions

A general approach to calculation of noise level at the output of DDS due to internal noise sources of DAC is developed. General formulae (24),(25) and (35 - 37) are valid for all noise sources but they are inconvenient for calculations and analysis. If noise components of bit current generators have correlation time that is much smaller than clock cycle T_c we arrive to the formulae (39),(40).They are simpler but still inconvenient for analysis.

Using white noise model (41) leads to rather simple results (42),(43).The simplest results (48 - 50) and (53-56) are obtained for noise sources located in bit current generators and in biasing circuit.

The formulae show the parameters one has to know to predict fundamental noise level at the DDS output.

Unfortunately this approach can not be applied to calculate an influence of low frequency noises.

Acknoledgements

The authors would like to acknoledge T.Boldyreva for great help in this paper preparation.

References

[1] J.Tirney, C.M.Radar, and B.Gold, "A Digital Frequency Synthesizer", IEEE Trans. on Audio Electroacoust., AU-19, March 1971, pp.43-56

[2] V.F.Kroupa, "Spectra of Pulse Rate Frequency Synthesizers", Proc. IEEE, vol.67,December 1979, pp.1680-82

[3] C.E.Wheatley,III and D.E.Phillips,"Spurious Suppression in Direct Digital Synthesizers,"Proc. 35th Annual Symp.on Freq.Control,1981

[4] О.Т.Матюшин,"Цифровой синтезатор с использованием функций Уолша", Радиотехника и электроника , т.27, N 7, 1982, 66.1301-08

[5] В.В.Шахгильдян, В.Г.Лучков,"Спектральные характеристики цифровых синтезаторов частот",Радиотехника, 1984,N3, сс.62-64

[6] V.Reinhardt, K.Gould, K.McNab and M.Bustamante, "A Short Survey of Frequency Synthesizer Techniques", Proc.40th Annual Freq.Contr.Symposium, May, 1986

[7] H.T.Nicholas,and H.Samueli,"An Analysis of the Output of Direct Digital Frequency Synthesizers in the Presence of Phase Accumulator Truncation", Proc. 41st Annual Symp. on Freq.Control, 1987, pp.495-502

[8] H.T.Nicholas, H.Samueli, and B.Kim, "The Optimization of Direct Digital Frequency Synthesizers in the Presence of Finite Word Length Effects", Proc.42nd Annual Symp.on Freq. Control,1988,pp.357-363

[9] E.M.Mattison,and L.M.Coule,"Phase Noise in Direct Digital Synthesizers", Proc.42nd Annual Symp. on Freq.Control, 1988, pp.352-356

[10]V.S.Reinhardt,"Spur Reduction Techniques in Direct Digital Synthesizers",Proc.1993 IEEE Int.Freq. Control Symposium, pp.230-241

[11]V.F.Kroupa,"Discrete Spurious Signals and Background Noise in Direct Digital Frequency Synthesizers", Proc.1993 IEEE Int.Freq.Control Symposium, pp.242-250

[12]T.J.Endres, R.B.Hall, and A.M.Lopez, "Design and Analysis Methods of a DDS-based Synthesizer for Military Spaceborn Applications", Proc.1994 IEEE Int.Freq.Control Symposium, pp.624-632

[13]M.M.Driscoll, "Reduction of Quartz Crystal Oscillator Flicker-of-Frequency and White Noise (Floor) Levels and Acceleration Sensitivity via Use of Multiple Resonators", Proc.1992 IEEE Freq. Control Symposium,pp.334-339

[14]A.van der Ziel,"Noise:Sources,Characterization, Measurement",Prentice-Hall Inc.,Englewood Cliffs, N.J. 1970

[15] В.Жалуд. В.Кулешов,"Шумы в полупроводниковых устройствах", Сов.радио, Москва, 1977

[16] С.М.Рытов,"Введение в статистическую радиофизику", Наука, Москва, 1966

[17] В.Н.Кулешов,"Разработка и применение системы методов прикладного анализа флуктуаций в источниках колебаний",Дисс.докт.техн.наук,МЭИ, Москва, 1988

[18]А.-Й.К.Марцинкявичус, Э.-А.К.Багдаискис, Р.Л.Пошюиас, Б.В.Драган, Й.М.Кажукаускас, И.Д.Кучинскас, "Быстродействующие интегральные схемы ЦАП и АЦП и измерение их параметров", Радио и связь, Москва, 1988

[19] Б.Г.Федорков , В.А.Телец, "Микросхемы ЦАП и АЦП: функционирование, параметры, применение", Радио и связь, Москва ,1990

[20]P.H.Saul,J.S.Urquhart,"Techniques and Technology for High Speed D-A Conversion", IEEE Journ. of Solid State Circuits,vol.SC-19, No.1,1984,pp.62-68

[21]T.S.Guy, L.M.Trythall, and A.J.Brodersen, "A Sixteen-Bit Monolithic Bipolar DAC",IEEE Journ.of Solid State Circuits, vol.SC-17,No.6, pp.1127-32

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THE ORIGIN OF 1/f PM AND AM NOISE IN BIPOLAR JUNCTION TRANSISTOR AMPLIFIERS*

F. L. Walls*, E.S. Ferre-Pikal+ and S.R. Jefferts*

*National Institute of Standards and Technology
325 Broadway, Boulder CO, 80303

+National Institute of Standards and Technology and University of Colorado at Boulder
Boulder, CO, 80303

Abstract - In this paper we report the results of extensive research on phase modulation (PM) and amplitude modulation (AM) noise in linear bipolar junction transistor (BJT) amplifiers. BJT amplifiers exhibit 1/f PM and AM noise about a carrier signal that is much larger than the amplifier's thermal noise at those frequencies in the absence of the carrier signal. Our work shows that the 1/f PM noise of a BJT based amplifier is accompanied by 1/f AM noise which can be higher, lower, or nearly equal depending on the circuit implementation. The 1/f AM and PM noise in BJTs is primarily the result of 1/f fluctuations in transistor current, transistor capacitance, circuit supply voltages, circuit impedances, and circuit configuration. We discuss the theory and present experimental data in reference to common emitter amplifiers, but the analysis can be applied to other configurations as well. This study provides the functional dependence of 1/f AM and PM noise on transistor parameters, circuit parameters, and signal frequency, thereby laying the groundwork for a comprehensive theory of 1/f AM and PM noise in BJT amplifiers. We show that in many cases the 1/f PM and AM noise can be reduced below the thermal noise of the amplifier.

Introduction

The focus of this paper is to present a theory which explains the up-conversion of the baseband (near dc) noise of amplifier circuits to produce phase modulation (PM) and amplitude modulation (AM) noise about a coherent high frequency signal in linear amplifiers which use silicon bipolar junction transistors (BJTs) as the active gain element. This theory is based on a large number of experimental measurements and theoretical modeling of BJTs. The theory provides a starting point for designing amplifiers with low PM and AM noise and for estimating how the noise scales with operating point and signal frequency.

Equation 1 shows the model of a perfect signal of frequency ν_0 and amplitude V_0 with amplitude fluctuations characterized by $\varepsilon(t)$ and phase fluctuations characterized by $\phi(t)$.

$$V(t) = [V_0 + \varepsilon(t)]\cos(2\pi\nu_0 t + \phi(t)). \qquad (1)$$

PM noise is characterized by

$$S_\phi(f) = [\Delta\phi(f)]^2 \frac{1}{BW}, \qquad 0 < f < \infty, \quad (2)$$

where BW is the bandwidth of the measurement, and f is the Fourier frequency offset from ν_0. The AM noise is characterized by

$$S_a(f) = \frac{[\Delta\varepsilon(f)]^2}{V_0^2} \frac{1}{BW}, \qquad 0 < f < \infty, \quad (3)$$

where BW is the bandwidth of the measurement. These definitions show that one cannot define the added phase noise or amplitude noise of an amplifier except in the presence of a carrier signal. In much of the literature the single sideband AM noise, 1/2 $S_a(f)$, and PM noise, $\mathcal{L}(f)$, are used and we will also for the remainder of the paper.

U.S. Government work not protected by U.S. copyright

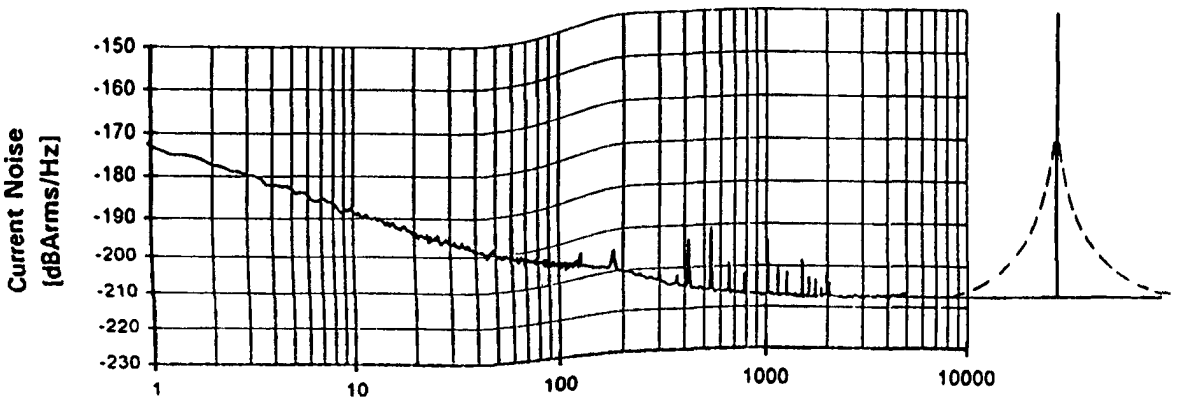


Fig. 1. Current noise of an amplifier as a function of frequency.

The solid line in Fig. 1 shows the output noise power per Hz bandwidth of an amplifier as a function of frequency. Near zero frequency the noise power typically varies as approximately 1/f where f is frequency. At higher frequencies (approximately 300 Hz for silicon BJTs, the noise power/Hz is independent of frequency and given by

$$P_N = 4kTFG, \qquad (4)$$

where k is Boltzmann's constant, T is temperature in kelvin, F is the noise figure, and G is the power gain. At 300 K this thermal noise is equal to -171 dBm/Hz for F = G = 1. When a coherent signal is introduced this thermal noise is divided equally between AM and PM noise and scales as $1/P_0$, where P_0 is the output signal power:

$$1/2\ S_a(f) = \mathcal{L}(f) = 2kTFG/P_0. \qquad (5)$$

If the noise on the input signal is low enough, it is possible to show that the amplifier adds a small amount of noise power, in addition to the thermal noise given in Eq. 4, to the output signal. This excess noise, which is proportional to the input signal power, is illustrated by the dashed line in Fig. 1. When referenced to the center of the signal, this excess noise has the same 1/f characteristics as that observed near dc and is therefore commonly referred to up-converted dc or baseband noise. *In contrast to the thermal noise, circuit parameters determine whether the 1/f portion of the AM noise is higher, equal to, or lower than the 1/f portion of the PM noise.*

In the following we present a theory and experimental data to explain the creation of excess 1/f PM and AM noise in amplifiers which use silicon BJTs as the active gain element. We find that the 1/f PM and AM noise is primarily the result of 1/f fluctuations in transistor current, transistor capacitance, circuit supply voltages, and circuit impedances. This theory shows the first order dependence of 1/f PM and AM noise in linear amplifiers on circuit parameters, transistor parameters, and signal frequency and therefore can be used to select transistors and choose the circuit parameters that minimize added 1/f noise in amplifiers. Although the theory is discussed relative to a common emitter (CE) configuration it can also be used to analyze other configurations. In many cases the 1/f noise can be reduced to such an extent that it is much lower than the thermal noise for frequency offsets from the signal larger than a few Hz [1].

Early Work on 1/f Noise in Amplifiers And Frequency Multipliers

Halford et al. [2] and Andressen [3] were the first to show that the 1/f PM noise of CE amplifiers and frequency multipliers can be reduced by adding an unbypassed resistance R_E in the emitter leg of the transistor[2,3]. See Fig. 2. The mechanisms by which the up-conversion occurred and the reason that only feedback to stabilize the rf gain provided significant reduction of the flicker PM noise were not understood. At the time of this early work the excess AM noise was ignored by virtually everyone except Healy [4]. Because no theory existed to explain the reduction of the PM noise with

unbypassed emitter resistor or the effect on the AM noise it was not possible to design amplifiers with low 1/f noise except by trial and error.

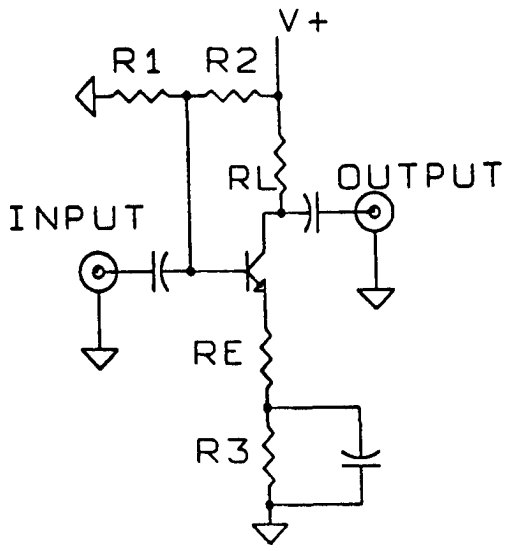


Fig. 2. CE amplifier with unbypassed emitter resistor.

Our work in this area started several years ago with the publication of references [5,6]. This work showed that the AM and PM noise in amplifiers is often very similar and raised the possibility that the noise might be due to the same mechanisms.

1/f Noise in Linear Amplifiers

We first derive the complete general expression for added AM and PM noise in an amplifier. The first order approximations to these expressions are then used to derive the added AM and PM noise for a CE amplifier with a single dominant pole. We then show in a simplified example how fluctuations in the emitter current cause fluctuations in the magnitude of the gain thereby inducing AM noise. We then include the current and voltage dependence of various transistor and circuit elements to develop a more complete theory of both AM and PM noise in CE amplifiers.

A General Expression of Added AM and PM Noise in Amplifiers

The general transfer function for an amplifier, G(s), can be written as

$$G(s) = G_o F_L(s) F_H(s), \qquad (6)$$

where $s = j\omega$ ($\omega = 2\pi\nu_o$ = angular frequency), G_O is the mid-band gain, F_L is the low frequency dependence of the transfer function, and F_H describes the high frequency dependence of gain on frequency. For the purposes of the present discussion, we can concentrate on the behavior of F_H, as we are assuming that the amplifier is operating at a frequency well above its low frequency cut-off. F_H can be written as

$$F_H(s) = \frac{(1+s/\omega_{z1})(1+s/\omega_{z2}) \ldots (1+s/\omega_{znh})}{(1+s/\omega_{p1})(1+s/\omega_{p2}) \ldots (1+s/\omega_{pnh})}, \qquad (7)$$

where $\omega_{p1}, \omega_{p2}, \ldots, \omega_{pnh}$ are positive numbers representing the angular frequencies of the n real poles and $\omega_{z1}, \omega_{z2}, \ldots, \omega_{znh}$ are real numbers representing the frequencies of the n zeros. (Note that these can be positive, negative, or infinite.) In many real-world cases the zeros are at such high frequencies that they are of little significance in determining the upper cutoff frequency of G(s). In this case the function $F_H(s)$ can be approximated by

$$F_H(s) = \frac{1}{(1+s/\omega_{p1})(1+s/\omega_{p2}) \ldots (1+s/\omega_{pnh})}. \qquad (8)$$

An alternative representation which will prove useful in the following discussion is to analyze G(s) in terms of the gain and phase. That representation leads to G(s) being described by

$$G(s) = G_O |F_H(s)| e^{j\vartheta}, \quad \vartheta = \tan^{-1} \frac{\mathrm{Im}(F_H(s))}{\mathrm{Re}(F_H(s))}. \qquad (9)$$

It should be noted that the two representations are equivalent.

Finally in the case which we will analyze in detail, the CE amplifier, one pole is dominant, at which point we can further approximate $F_H(s)$ as

$$F_H(s) = \frac{1}{1+s/\omega_{pl}}, \quad (10)$$

Equation 9 can now be written as

$$G(s) = G_o\left(\frac{1}{1+\delta^2}\right)^{1/2} e^{j\vartheta}, \vartheta = \mathrm{Tan}^{-1}\delta, \delta = \left(\frac{\omega}{\omega_{pl}}\right). \quad (11)$$

If we consider the input signal to be perfect, that is

$$V_{in} = V_o \cos\omega t, \quad (12)$$

then the output signal is given by

$$V_{out} = V_{in}G = V_oG_o\left(\frac{1}{1+\delta^2}\right)^{1/2} e^{j\vartheta}\cos(\omega t) = V_oG_o\left(\frac{1}{1+\delta^2}\right)^{1/2}\cos(\omega t+\vartheta). \quad (13)$$

The AM noise added by an amplifier can be written as

$$\frac{1}{2}S_a(f) \cong \frac{1}{2}\left(\frac{\Delta G_v}{G_v}\right)^2 \frac{1}{BW} + \frac{2kTFG}{P_o}, \quad (14)$$

where G_V is the magnitude of the voltage gain. The PM noise added by an amplifier can be written as

$$\frac{1}{2}S_\phi(f) \cong \frac{1}{2}\Delta\vartheta^2 \frac{1}{BW} + \frac{2kTFG}{P_o}. \quad (15)$$

Equations 14 and 15 also include the contribution of the thermal noise. These expressions are completely general and can be used on any two port element such as an amplifier, attenuator, or level shifter, as long as the output frequency is the same as the input by expressing G_v as a function of the transistor and circuit parameters at the operating carrier frequency. Equations 14 and 15 can also be extended to cover two port elements where the frequency is changed by multiplying by the square of the ratio of output to input frequency.

Single Pole Approximation for Gain of Common-Emitter Amplifier

In this section we use the above results to develop a physical model which explains the up-conversion of baseband noise to create excess AM and PM noise in CE amplifiers. The equation describing gain in a practical CE amplifier is very complicated; therefore, several approximations will be used to simplify the equations so that the underlying principles are more easily seen.

In the dominant pole case at frequencies ω much smaller than the 3 dB bandwidth (ω_{pl}), that is $\delta = \left(\frac{\omega}{\omega_{pl}}\right) < 1$, Eq. 11 can be further approximated and simplified to

$$G = G_o\left(\frac{1}{1+\delta^2}\right)^{1/2} e^{j\vartheta} = G_o\left(1-\frac{\delta^2}{2}+\frac{3\delta^4}{8}\ldots\right)e^{j(\delta+\delta^3\ldots)} = G_o\left(1-\frac{\delta^2}{2}\right)e^{j\delta}. \quad (16)$$

This result can be used to simplify Eq. 13 to

$$V_{out} = V_{in}G = V_oG_o\left(1-\frac{\delta^2}{2}\right)\cos(\omega t+\delta). \quad (17)$$

Equation 14 can be expanded to yield

$$\frac{1}{2}S_a(f) \cong \frac{1}{2}\left(\frac{\Delta G_o}{G_o}\right)^2 \frac{1}{BW} + \frac{1}{2}\delta^2\Delta\delta^2 \frac{1}{BW} + \frac{2kTFG}{P_o}. \quad (18)$$

Equation 18 can also be expressed as

$$\frac{1}{2}S_a(f) \cong \frac{1}{2}\left(\frac{\Delta G_o}{G_o}\right)^2 \frac{1}{BW} + \frac{1}{2}\delta^2\left(S_\phi(f) - \frac{2kTFG}{P_o}\right) + \frac{2kTFG}{P_o}, \quad (19)$$

where

$$\frac{1}{2}S_\phi(f) \cong \frac{1}{2}\Delta\delta^2 \frac{1}{BW} + \frac{2kTFG}{P_o}. \quad (20)$$

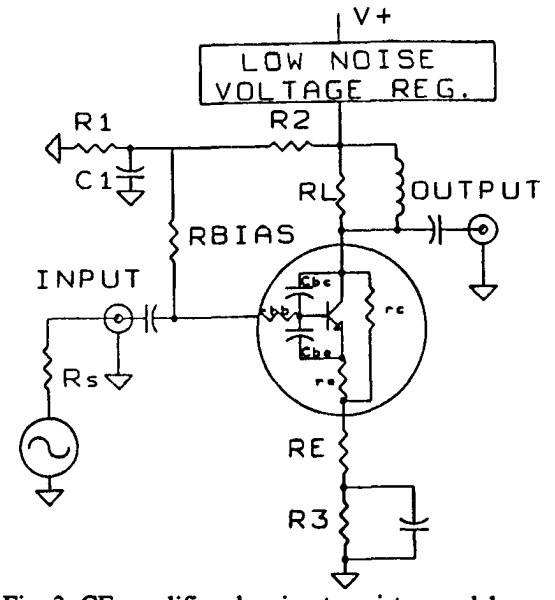


Fig. 3. CE amplifier showing transistor model.

Fig. 3 shows the schematic of a simple CE amplifier. In the approximation that the load impedance can be represented by R_L, and ω much smaller than ω_{pl} (that is the phase shift δ is small) we can write the gain as

$$G = G_o\left(\frac{1}{1+\delta^2}\right)^{1/2} e^{-j\vartheta} = G_o(1-\frac{\delta^2}{2})e^{-j\delta}, \qquad (21)$$

$$G_o = -\frac{R_L}{r_e + R_E + r_g/\beta},$$

$$r_g = \left(r_{bb} + \frac{R_S R_{BIAS}}{R_S + R_{BIAS}}\right)K, \qquad (22)$$

$$\delta \cong \omega C_{bc} G_o\left(r_e + R_E + r_g\right) + \omega C_{be} G_o r_g r_e\left(1 - \frac{R_E}{R_E + r_e}\right)M/R_L + .. \qquad (23)$$

$$M = \frac{1}{1 + (\omega C_{be}\beta)^2\left(\frac{r_e R_E}{r_e + R_E}\right)^2}, \qquad (24)$$

$$K = \frac{1 + (\omega C_{bc}\beta)^2 r_e\left(\frac{r_e R_E}{r_e + R_E}\right)}{1 + (\omega C_{bc}\beta)^2\left(\frac{r_e R_E}{r_e + R_E}\right)^2}, \qquad (25)$$

where the R's are defined in Fig. 3, r_g is the total effective input impedance, and δ is the phase shift due to C_{bc}, and C_{be}. A further simplification can be made for CE amplifiers with large gain or $C_{be}\frac{r_e}{R_L}\left(1 - \frac{R_E}{R_E + r_e}\right) << C_{bc}$. In this case and for $K \approx 1$, and $M \approx 1$, the gain becomes

$$G_o \cong -\frac{R_L}{r_e + R_E + r_g/\beta},$$

$$r_g = \left(r_{bb} + \frac{R_S R_{BIAS}}{R_S + R_{BIAS}}\right), \qquad (26)$$

$$\delta \cong \omega C_{bc} G_o\left(r_e + R_E + r_g\right).$$

When the emitter current is noise modulated as

$$I_E = I_{dc} + \Delta I_E \cos(\Omega t + \theta), \qquad (27)$$

the intrinsic emitter resistance r_e is also modulated. For $\Omega << \nu_o$ and $\gamma = \frac{\Delta I_E}{I_E} << 1$, the average value of r_e is given by

$$r_e' \cong \frac{K}{I_{dc}[1 + \gamma\cos(\Omega t + \theta)]} \cong r_e[1 - \gamma\cos(\Omega t + \theta)], \qquad (28)$$

with K = 26 ohms/mA emitter current. We can now express the gain as

$$G_o \cong \frac{-R_L}{r_e + R_E + r_g/\beta}\left[1 + \frac{\gamma r_e \cos(\Omega t + \theta))}{r_e + R_E + r_g/\beta}\right]. \qquad (29)$$

Averaging over the phase of the modulation yields

$$\overline{G_o^2} \cong \left(\frac{R_L}{r_e + R_E + r_g/\beta}\right)^2 \left[1 + \frac{\gamma^2}{2}\left(\frac{r_e}{r_e + R_E + r_g/\beta}\right)^2\right]. \quad (30)$$

The single sideband AM noise spectrum added by the amplifier due to current modulation follows the spectrum of the current modulation and is given by

$$1/2\, S_a(f) \cong 1/4 \left(\frac{r_e}{r_e + R_E + r_g/\beta}\right)^2 \gamma^2(f). \quad (31)$$

Considering the possible variation of the other parameters affecting the gain and phase of this simplified CE amplifier, we obtain a more complete expression for the AM noise given by

$$\begin{aligned}\frac{1}{2}S_a(f) \cong\ & \frac{1}{4}\left(\frac{r_e}{r_e + R_E + r_g/\beta}\right)^2 \gamma^2 + \frac{1}{4}\left(\frac{r_g/\beta}{r_e + R_E + r_g/\beta}\right)^2 \left(\frac{\Delta r_g}{r_g}\right)^2 \\ & + \frac{1}{4}\left(\frac{R_E}{r_e + R_E + r_g/\beta}\right)^2 \left(\frac{\Delta R_E}{R_E}\right)^2 + \frac{1}{4}\left(\frac{\Delta R_L}{R_L}\right)^2 \\ & + \frac{\delta^2}{4}\left(\frac{r_e}{r_e + R_E + r_g/\beta}\right)^2 \left(\omega C_{bc} G_o (r_e + R_E + r_g)\right)^2 \gamma^2 + \frac{2kTFG}{P_o} \\ & + \frac{\delta^2}{4}\left[\left(\omega G_o (r_e + R_E + r_g)\right)^2 \Delta C_{bc}^2 + \left(\omega C_{bc} G_o\right)^2 (\Delta r_e^2 + \Delta R_E^2 + \Delta r_g^2)\right]. \end{aligned} \quad (32)$$

Eq. 32 can be considerably simplified if we make use of the following approximations,

$$r_g > (r_e + R_E), \quad \Delta r_e = \gamma\, r_e. \quad (33)$$

$$\begin{aligned}\frac{1}{2}S_a(f) \cong\ & + \frac{G_o^2}{4}\left[\left(\frac{\gamma r_e}{R_L}\right)^2 + \left(\frac{\Delta r_g}{\beta R_L}\right)^2 + \left(\frac{\Delta R_E}{R_L}\right)^2\right] \\ & + \frac{1}{4}\left(\frac{\Delta R_L}{R_L}\right)^2 + \frac{2kTFG}{P_o} \\ & + \frac{(\omega G_o)^4 (C_{bc} r_g)^2}{4}\left((r_g \Delta C_{bc})^2 + C_{bc}^2\left(\Delta R_E^2 + \Delta r_g^2\right)\right). \end{aligned} \quad (34)$$

The first term in the above equation scales as $\left(\frac{r_e}{r_e + R_E + r_g/\beta}\right)^2$, while the contribution of the terms associated with ΔC_{bc} scale as $\left(\frac{r_g}{r_e + R_E + r_g/\beta}\right)^4$. Both sets of terms are reduced by increasing R_E, although the AM noise originating from capacitance modulation is reduced much faster than that originating from current modulation. The first term is also decreased by increasing the dc emitter current I_E. The limit to which I_E can be increased depends on the maximum ratings of the transistor and its environment.

The PM noise added by the CE amplifier in the same approximation used above for the AM noise yields

$$\begin{aligned}1/2\, S_\phi(f) = \mathcal{L}(f) = & \frac{1}{4}\left(\omega G_o (r_e + R_E + r_g)\right)^2 \Delta C_{bc}^2 + \\ & \frac{1}{4}\left(\omega C_{bc} G_o\right)^2 \left(\Delta r_e^2 + \Delta R_E^2 + \Delta r_g^2\right) + \frac{2KTFG}{P_o} \\ & + \frac{1}{4}\left(\frac{r_e}{r_e + R_E + r_g/\beta}\right)^2 \left(\omega C_{bc} G_o (r_e + R_E + r_g)\right)^2 \gamma^2. \end{aligned} \quad (35)$$

For $\Delta r_e = \gamma r_e$, $r_g > (r_e + R_E)$, Eq. 35 can be reduced to

$$\begin{aligned}\mathcal{L}(f) = & \frac{1}{4}(\omega G_o)^2 \left[\left(r_g \Delta C_{bc}\right)^2 + (C_{bc})^2 \left(\frac{\gamma\, r_g r_e G_o}{R_L}\right)^2\right] \\ & + \frac{1}{4}(\omega G_o)^2 (C_{bc})^2 \left(\Delta R_E^2 + \Delta r_g^2\right) + \frac{2kTFG}{P_o}. \end{aligned} \quad (36)$$

The PM noise due to capacitance modulation is reduced as $\left(\frac{R_L r_g}{r_e + R_E + r_g/\beta}\right)^2$, while the contribution of the terms associated with current modulation scale as $\left(r_e r_g R_L\right)^2 \left(\frac{1}{r_e + R_E + r_g/\beta}\right)^4$. Both sets of terms are reduced by increasing R_E and reducing r_g, although the PM noise originating from current modulation is reduced much faster with R_E than those originating from capacitance modulation. Increasing the dc emitter current I_E primarily affects the term originating from current modulation.

Current modulation originates from the intrinsic baseband modulation generated within the transistor and by noise in the power supply. The modulation of C_{bc} is induced by baseband modulation of the base collector voltage. This

baseband voltage modulation can originate from the current modulation and finite dc gain or from the power supply.

Measurement of AM and PM Noise Induced by Current Modulation in a CE Amplifier

Current noise was injected in the emitter of the transistor shown in Fig. 3 to test the functional form of Eq. 31. Current noise of order 22 X 10^{-12} A_{rms}^2/Hz was used, which was large enough that the induced AM and PM noise was usually above the noise floor of our measurement system. Table 1 shows the measured AM noise sensitivity to γ ($\Delta I_B/I_B$) and the predictions calculated using Eq. 31 for a 2N2222A and a microwave transistor (with much lower C_{bc}) for $R_E = 0\ \Omega$ and 10 Ω. The characteristics of the transistors are given in Table 2. Column A in Table 1 shows the measured AM noise sensitivity to γ as a function of Fourier frequency offset from a 5 MHz carrier for the two transistors with $R_E = 0$, while column B shows the calculated value for $r_g/\beta \cong 1\ \Omega$ and $r_e \cong 1.5\ \Omega$. Column C shows the measured AM noise sensitivity with $R_E = 10\ \Omega$ and Column D shows the calculated values. The calculated and measured improvements in the AM noise sensitivity to γ due to increasing R_E from 0 to 10 Ω agree very well for both transistors.

	A	B	C	D
	2N2222A			
Fourier frequency	AM sensitivity $R_E = 0\ \Omega$ [dBc/Hz rel to γ=1]	AM theory $R_E = 0\ \Omega$ [dBc/Hz rel to γ=1]	AM sensitivity $R_E = 10\ \Omega$ [dBc/Hz rel to γ=1]	AM theory $R_E = 10\ \Omega$ [dBc/Hz rel to γ=1]
100 Hz	-8.5	-10.5	-23.4	-25.1
50 Hz	-8.6	-10.5	-23.3	-25.1
20 Hz	-8.6	-10.5	-22.9	-25.1
10 Hz	-8.6	-10.5	-22.7	-25.1
5 Hz	-8.9	-10.5	-23.3	-25.1
	microwave transistor			
Fourier frequency	AM sensitivity $R_E = 0\ \Omega$	AM theory $R_E = 0\ \Omega$	AM sensitivity $R_E = 10\ \Omega$	AM theory $R_E = 10\ \Omega$
100 Hz	-8.5	-10.5	-23.1	-25.1
50 Hz	-8.5	-10.5	-23.4	-25.1
20 Hz	-8.5	-10.5	-23.3	-25.1
10 Hz	-8.5	-10.5	-22.9	-25.1
5 Hz	-8.6	-10.5	-22.5	-25.1

Table 1. AM and PM noise sensitivities to current noise for a CE amplifier when $\gamma = \Delta I_B/I_B = 1.9 \times 10^{-5}$, $I_B = 25$ mA, and $V_{CB} \approx 9$ V.

	E	F	G	H
	2N2222A			
Fourier frequency	PM sensitivity $R_E = 0\ \Omega$ [dBc/Hz rel to γ=1]	PM theory $R_E = 0\ \Omega$ [dBc/Hz rel to γ=1]	PM sensitivity $R_E = 10\ \Omega$ [dBc/Hz rel to γ=1]	PM theory $R_E = 10\ \Omega$ [dBc/Hz rel to γ=1]
100 Hz	-16.3	-15.6	-45.6	-43.7
50 Hz	-16.2	-15.6	-45.2	-43.7
20 Hz	-15.6	-15.6	-45.8	-43.7
10 Hz	-16.3	-15.6	-45.4	-43.7
5 Hz	-16.3	-15.6	-45.6	-43.7
	microwave transistor			
Fourier frequency	PM sensitivity $R_E = 0\ \Omega$ [dBc/Hz rel to γ=1]	PM theory $R_E = 0\ \Omega$ [dBc/Hz rel to γ=1]	PM sensitivity $R_E = 10\ \Omega$ [dBc/Hz rel to γ=1]	PM theory $R_E = 10\ \Omega$ [dBc/Hz rel to γ=1]
100 Hz	-43.2	-37.4	limited by	-65.5
50 Hz	-42.6	-37.4	system	-65.5
20 Hz	-42.1	-37.4	floor	-65.5
10 Hz	-42.4	-37.4	<-48.6	-65.5
5 Hz	-41.8	-37.4		-65.5

Table 1 (cont.)

Column E shows the measured PM noise sensitivity to γ for $R_E = 0\ \Omega$ while column G shows the PM noise sensitivity to γ for $R_E = 10\ \Omega$. The calculated values using the last term in Eq. 35 are shown in columns F and H. For the 2N2222A, the calculated values are in agreement with the measured values. Eq. 35 predicts that the PM noise (due to current noise) at $R_E = 10\ \Omega$ is reduced relative to the PM noise at $R_E = 0\ \Omega$ by the square of the reductions observed for the AM noise. The measured reduction of PM noise follows this relatively closely for the 2N2222A. The measured PM noise sensitivity to γ for the microwave transistor is shown in column E. This PM sensitivity should be reduced relative to that of the 2N2222A by the ratio of $(C_{bc1}/C_{bc2})^2$. For these two transistors that ratio is approximately 22 dB. The measured reduction is about 5 dB higher. Measurements at $R_E = 10\ \Omega$ for the microwave transistor were limited by the noise floor of the measurement system.

Table 1 along with Eqs. 34 and 35 can be used to predict the AM and PM levels due to current fluctuations in a CE amplifier. For a given current noise (ΔI_E), dc emitter current (I_E) and unbypassed emitter resistor (R_E), one can estimate the resulting level of the AM noise. One can also calculate the maximum allowed value of γ ($\Delta I_E/I_E$) for a desired AM noise level. In a similar way, one can predict the resulting PM noise due to current fluctuations, but the values of C_{cb} and r_g are required.

	2N2222A	Microwave transistor
f_T	300 MHz	8 GHz
C_{be}	25 pF	1.6 pF
C_{bc} (V_{CB} = 9V)	8 pF	.65 pF
β	$75 < \beta < 375$	$50 < \beta < 300$
F	$\leq$ 4 dB	1.5 dB

Table 2. Transistor parameters for Table 1.

Measurements of AM and PM Noise Induced by Collector-Base Voltage Modulation in a CE Amplifier

The collector-base capacitance is a junction capacitance that varies with the dc voltage between the collector and base terminals (V_{CB}) according to Eq. 37. (In this equation, K_0 is a constant that includes the permittivity of the material and the doping concentration, V_{bi} is the built-in potential of the p-n junction, and n is a parameter that depends on the doping profile of the junction.) Collector-base voltage modulation will therefore modulate C_{bc}.

$$C_{bc} = \frac{K_0}{(V_{bi} + V_{CB})^n}. \quad (37)$$

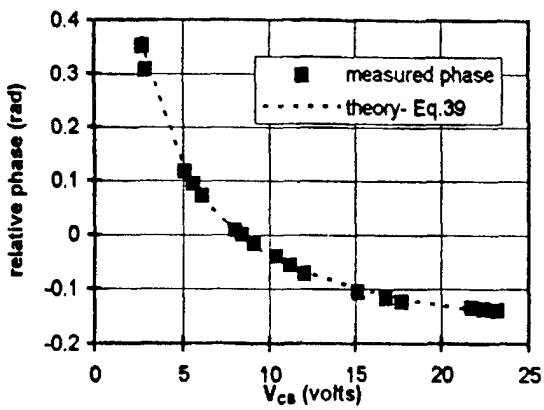


Fig. 4. Relative phase shift of a 20 MHz CE amplifier as a function of V_{CB}.

Fig. 4 shows the relative phase shift across a 20 MHz CE amplifier (shown in Fig. 3) as a function of collector-base voltage (V_{CB}) when the active element was a 2N2222A transistor. This graph shows the dependence of C_{bc} on V_{CB}, and can be used to obtain the exponent n in the phase equation of the amplifier (Eq. 38). The optimum fit, shown as a dotted line in Fig. 4, occurred when $n \cong 1$. As shown in Fig. 4, the measured phase shift follows Eq. 38 fairly well for V_{BC} from 3 to 24 V.

$$\text{phase} = \vartheta = K_1\omega + K_2\omega C_{bc} = K_1\omega + \frac{K_3\omega}{(V_{bi} + V_{CB})^n}. \quad (38)$$

The PM sensitivity to voltage fluctuations can be obtained by squaring the partial derivative of the phase with respect to V_{CB}:

$$\left(\frac{\partial\vartheta}{\partial V_{CB}}\right)^2 = \left(\frac{nK_3\omega}{(V_{bi} + V_{CB})^{n+1}}\right)^2. \quad (39)$$

To measure the effect of ΔV_{CB} on the AM and PM noise, voltage noise was added to the collector terminal of the CE amplifier. Filtering on the base reduced any direct current modulation to an insignificant level. Fig. 5 shows the PM noise sensitivity to V_{CB} fluctuations measured with a 2N2222A transistor in the CE circuit (for R_E = 30 Ω) as a function of V_{CB} at 5, 10, and 20 MHz carrier frequencies. The dotted lines are predictions on how the PM should vary with V_{CB} based on Eq. 39 for n = 1. The form of the PM noise follows closely the predictions from 6 to 17 V and scales as ω^2 as expected.

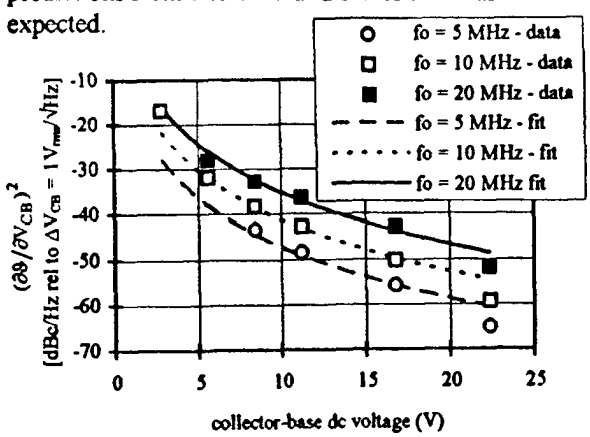


Fig. 5. PM noise sensitivity to ΔV_{CB} in a CE amplifier.

The change of PM sensitivity to V_{CB} fluctuations was also measured for different R_Es (for a V_{CB} = 8V). Table 3 shows the PM sensitivity measurements and predicted values (from Eq. 35) as a function of Fourier frequency offset from a 5 MHz carrier for the CE amplifier in Fig. 3 when voltage noise was injected at the collector terminal. Columns A and D show the measured PM sensitivity to V_{CB} fluctuations for R_E = 0 Ω and R_E = 10 Ω respectively. Columns B and E show the PM sensitivity values predicted by the first term in Eq. 35. For the calculations it was assumed that half of the output capacitance in the transistor was due to parasitics and the other half due to the junction capacitance (Eq. 37).

	A	B	C	D	E	F
	2N2222A					
Fourier frequency	PM sensitivity $R_E = 0\ \Omega$ [dBc/Hz rel to $\Delta V_{CB} = 1 V_{rms}/\sqrt{Hz}$]	PM theory $R_E = 0\ \Omega$ [dBc/Hz rel to $\Delta V_{CB} = 1 V_{rms}/\sqrt{Hz}$]	AM sensitivity $R_E = 0\ \Omega$ [dBc/Hz rel to $\Delta V_{CB} = 1 V_{rms}/\sqrt{Hz}$]	PM sensitivity $R_E = 10\ \Omega$ [dBc/Hz rel to $\Delta V_{CB} = 1 V_{rms}/\sqrt{Hz}$]	PM theory $R_E = 10\ \Omega$ [dBc/Hz rel to $\Delta V_{CB} = 1 V_{rms}/\sqrt{Hz}$]	AM sensitivity $R_E = 10\ \Omega$ [dBc/Hz rel to $\Delta V_{CB} = 1 V_{rms}/\sqrt{Hz}$]
100 Hz	-40.7	-36.2	-42.4	-52.3	-48.6	-63
50 Hz	-40.7	-36.2	-42.6	-52.4	-48.6	-62.8
20 Hz	-40.6	-36.2	-42.8	-52.4	-48.6	limited by
10 Hz	-40.6	-36.2	-43.1	-52.4	-48.6	system
5 Hz	-40.6	-36.2	-43.1	-52.4	-48.6	noise floor
	microwave transistor					
100 Hz	-73.6	-63.9	limited by	limited by	-76.3	limited by
50 Hz	-74.3	-63.9	system	system	-76.3	system
20 Hz	-73.2	-63.9	noise floor	noise floor	-76.3	noise floor
10 Hz	limited by	-63.9	<-63	<-77	-76.3	<-69
5 Hz	sys. floor	-63.9			-76.3	

Table 3. AM and PM noise sensitivities to collector-base voltage fluctuations for a CE amplifier with $V_{CB} = 8$ V and $I_E = 25$ mA. For these measurements the output power level was adjusted to be approximately 6 dBm and $\Delta V_{CB} \cong 2.8 \times 10^{-4}\ V_{rms}/\sqrt{Hz}$.

The measured and predicted values for the PM sensitivity when a 2N2222A was used differ by approximately 4 dB. This is probably due to the approximations made in Eq. 35 and the uncertainty in some of the values used (r_g, C_{bc}). From theory, the reduction in PM when R_E is added should be proportional to the reduction in gain. Both the measured and calculated sensitivities show approximately a 12 dB reduction for R_E changed from 0 Ω to10 Ω. The measured gain reduction was 13.5 dB. When a microwave transistor was used, the PM sensitivity to V_{CB} fluctuations for $R_E = 0\ \Omega$ was reduced by 33 dB, compared to the predicted reduction of 28 dB. The predicted value was obtained from Eqs. 35 and 39 when using $n = 0.55$ and $C_{bc} = 0.3$ pF. These values where obtained from the C_{bc} vs. V_{CB} plot in the specification sheet of the transistor. For $R_E = 10\ \Omega$ the noise was limited by the noise floor of the measurement system. The AM sensitivity to V_{CB} fluctuations is shown in columns C ($R_E = 0\ \Omega$) and F ($R_E = 10\ \Omega$) of Table 3. The AM sensitivity when using a 2N2222A was reduced by approximately twice the amount of reduction in PM sensitivity when an unbypassed emitter resistor was added ($R_E = 10\ \Omega$), as predicted by theory. The PM noise is thus a first order effect in the up-conversion of V_{CB} fluctuations, while the AM noise is a second order effect.

Reduction of AM and PM Noise Using Active Feedback

The AM and PM noise due to current fluctuations can be reduced considerably by adding a servo system that reduces the current noise. Fig. 6 illustrates such a system. A circuit that reads the current at the emitter, amplifies and inverts it, and injects the signal back into the base was added to the CE amplifier. Noise was added to the base of the transistor causing fluctuations in the emitter current. Emitter current fluctuations are up-converted to AM and PM noise, and voltage fluctuations at the base

also up-convert to PM noise. The servo reduces both the current fluctuations and the voltage fluctuations at the base; therefore, it should also reduce the AM and PM noise.

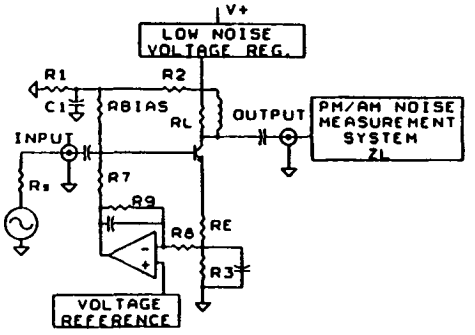


Fig. 6. CE amplifier with active feedback to stabilize the emitter current.

Fig. 7 shows the measured emitter voltage noise (proportional to current noise), and the AM and PM noise of the CE amplifier for open loop and close loop configurations. When the loop was closed the voltage noise, the AM noise, and the PM noise were reduced by the amount of the gain of the servo amplifier, as expected.

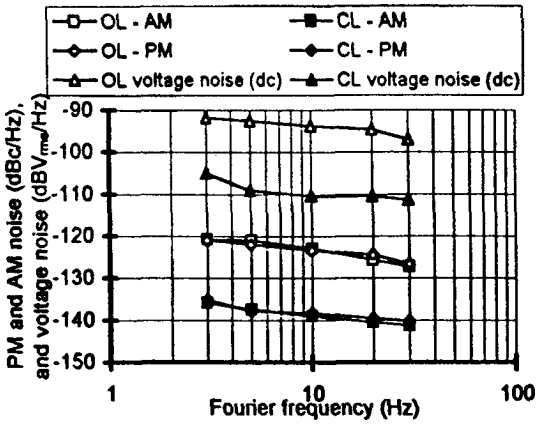


Fig. 7. Voltage noise, AM noise and PM noise of CE amplifier for open and close loop configurations.

Conclusion

We have presented a theory that explains the upconversion of baseband noise into AM and PM noise about a carrier in BJT amplifiers. From the theory, AM and PM noise equations were developed for CE amplifiers, but the analysis can also be applied to any BJT amplifier configuration. By injecting current noise and voltage noise in a CE amplifier we were able to verify to a few dB the leading term in both AM and PM noise due to baseband current modulation and collector-base voltage modulation. Based on this study we were able to confirm the functional dependence of AM and PM noise on transistor parameters, circuit parameters, and carrier frequency given by the derived noise equations. The developed theory lays the groundwork for a comprehensive theory of AM and PM noise in BJT amplifiers. In addition, the derived noise equations that describe the origin of AM and PM noise provide a basis for the design of low noise amplifiers.

Acknowledgments

The authors thank T.E. Parker, F.G. Ascarrunz, C.W. Nelson, and M. Sicardi for helpful discussions on the topic. This work was supported in part by PM Milstar (Army), Ft. Monmouth, NJ, and the Calibration Coordinaton Group.

References

[1] E.S. Ferre-Pikal, F.L. Walls and C.W. Nelson, "Design criteria for BJT amplifiers with low 1/f AM and PM noise", these proceedings.

[2] D. Halford, A.E. Wainwright, and J.A. Barnes, "Flicker noise of phase in rf amplifiers and frequency multipliers: characterization, cause, and cure", Proc. 22nd Ann. Frequency Control Symp., 1968, pp. 340-341.

[3] S.G. Andressen and J.K. Nesheim, "Phase noise of various frequency doublers", IEEE Trans. Instrum. Meas., Vol. IM-22, June 1973, pp. 185-188.

[4] D.J. Healy III, "Flicker of frequency and phase and white frequency and phase fluctuations in frequency sources", Proc. 26th Ann. Frequency Control Symp., 1972, pp. 29-42.

[5] L.M. Nelson, C.W. Nelson, and F.L. Walls, "Relationship of AM to PM noise in selected rf oscillators", Proc. 47th Ann. Frequency Control Symp., 1993, pp. 298-302.

[6] F.G. Ascarrunz, E.S. Ferre, and F.L. Walls, "Investigations of AM and PM noise in X-band devices", Proc. 47th Ann. Frequency Control Symp., 1993, pp. 303-311.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DESIGN CRITERIA FOR BJT AMPLIFIERS WITH LOW 1/f AM AND PM NOISE*

E.S. Ferre-Pikal*, F.L. Walls+, and C.W. Nelson#

*National Institute of Standards and Technology and University of Colorado at Boulder
Boulder, CO 80303

+National Institute of Standards and Technology
325 Broadway, Boulder, CO 80303

#Spectradynamics, Inc.
445 South 43rd st., Boulder, CO 80303

Abstract - In this paper we discuss guidelines for designing linear bipolar junction transistor (BJT) amplifiers with low 1/f amplitude modulation (AM) and phase modulation (PM) noise. These guidelines are derived from a new theory that relates AM and PM noise to transconductance fluctuations, junction capacitance fluctuations, and circuit architecture. We analyze the noise equations of each process for a common emitter (CE) amplifier and use the results to suggest amplifier designs that minimize the 1/f noise while providing other required attributes such as high gain. Although we use a CE amplifier as an example, the procedure applies to other configurations as well. Experimental noise results for several amplifier configurations are presented.

Introduction

In this paper we discuss the design criteria for low 1/f amplitude modulation (AM) and phase modulation (PM) noise in linear bipolar junction transistor (BJT) amplifiers using the theory discussed in [1]. This theory relates AM and PM noise to current fluctuations, voltage fluctuations, impedance fluctuations, and circuit architecture. We analyze the noise equations of each process for a common emitter (CE) amplifier and use the results to suggest amplifier designs that minimize the 1/f noise while providing other required attributes such as high gain. Although this is done specifically for a CE amplifier, the same procedure can be applied to any amplifier configuration. Guidelines for the design of low 1/f noise CE amplifiers are discussed in detail. We show low 1/f noise results for high gain CE amplifiers at carrier frequencies of 5 MHz and 100 MHz. Experimental results for other configurations (common base [CB], common collector [CC]) are also presented.

Noise Equations

As derived in [1], for an amplifier with voltage gain = $G_V e^{j\delta}$ (where G_V is the magnitude and δ is the phase shift), the AM noise is given by

$$\frac{1}{2}S_a(f) = \frac{1}{2}\left[\frac{\Delta G_V(i,V,\nu,f)}{G_V}\right]^2 \frac{1}{BW} + \frac{2kTFG}{P_o}, \quad (1)$$

and the PM noise by

$$\frac{1}{2}S_\phi(f) = \frac{1}{2}\frac{\Delta\delta(i,V,\nu,f)^2}{BW} + \frac{2kTFG}{P_o}. \quad (2)$$

In the first term of Eq. 1, $\Delta G_V(i,V,\nu,f)$ refers to the fluctuations in the voltage gain due to current noise, voltage noise, and impedance fluctuations, and BW is the bandwidth of the measurement. $\Delta G_V(i,V,\nu,f)$, which depends on dc current, dc voltage, circuit parameters, carrier frequency, and Fourier frequency, is the result of baseband 1/f noise up-converted to the carrier frequency. Similarly, $\Delta\delta(i,V,\nu,f)$ in Eq. 2 refers to the fluctuations in the phase shift of the amplifier due to current noise, voltage noise, and impedance fluctuations, and is also a function of dc current, dc voltages, circuit parameters, carrier frequency, and Fourier frequency. The second term in Eqs. 1 and

0-7803-2500-1/95 $4.00 © 1995 IEEE

2 is the thermal noise of the amplifier. In this term, k is Boltzmann's constant, T is the temperature in kelvin, F is the noise figure, G is the power gain of the amplifier, and P_o is the output power of the amplifier. Applying these equations (1 & 2) to the CE amplifier shown in Fig. 1, the resulting noise equations are

$$\frac{1}{2}S_a(f) = \frac{1}{4}\left(\frac{r_e}{r_e + R_E + r_g/\beta}\right)^2 \gamma^2 + \frac{1}{4}\left(\frac{r_g/\beta}{r_e + R_E + r_g/\beta}\right)^2\left(\frac{\Delta r_g}{r_g}\right)^2 + \frac{1}{4}\left(\frac{R_E}{r_e + R_E + r_g/\beta}\right)^2\left(\frac{\Delta R_E}{R_E}\right)^2 + \frac{1}{4}\left(\frac{\Delta R_L}{R_L}\right)^2 + \frac{1}{2}\delta^2\left(S_\phi(f) - \frac{2kTFG}{P_o}\right) + \frac{2kTFG}{P_o}, \quad (3)$$

$$\frac{1}{2}S_\phi(f) = \frac{1}{4}\left(\omega G_o[r_e + R_E + r_g]\right)^2 \Delta C_{bc}^2 + \frac{1}{4}(\omega C_{bc} G_o)^2\left(\gamma^2 r_e^2 + \Delta R_E^2 + \Delta r_g^2\right) + \frac{1}{4}\left(\frac{r_e}{r_e + R_E + r_g/\beta}\right)^2 \left(\omega C_{bc} G_o[r_e + R_E + r_g]\right)^2 \gamma^2 + \frac{2kTFG}{P_o}, \quad (4)$$

where $r_g = R_s \| R_{BIAS} + r_{bb}$, $R_L = R_L' R_{BIAS}/(R_s + R_{BIAS})$, $\gamma = \Delta I_E/I_E$, δ is the phase shift across the amplifier, and $G_o = \frac{R_L}{r_e + R_E + r_g/\beta}$. For details on the derivation of Eqs. 3 and 4 see [1].

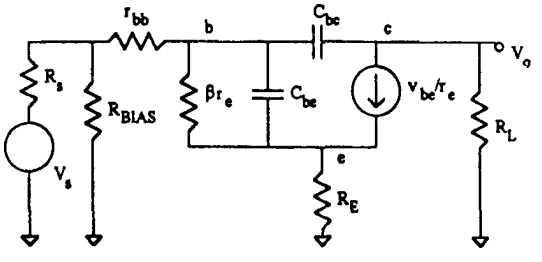


Fig. 1. Hybrid π model of CE amplifier.

Design Guidelines

From the noise equations one can develop guidelines for designing low 1/f AM and PM noise BJT amplifiers. In general, the effect of current fluctuations on AM noise can be reduced by increasing the standing current. For PM noise, the guidelines depend on the dominant poles of the phase equation. Therefore, it is difficult to come up with a unique set of rules for low 1/f PM noise design. The general rule that always applies is to design for the largest possible bandwidth (i.e., maximize frequency of the dominant poles). For this reason we have chosen a specific configuration, a CE amplifier, to develop guidelines for low 1/f AM and PM noise design. The same procedure for deriving guidelines can be applied to any amplifier configuration.

Looking at the AM noise equation (3), it is obvious that an unbypassed emitter resistance decreases the contribution from current noise and input impedance noise (first two terms). Similarly, all the terms in the PM noise equation have a factor G_o^2, suggesting that an increase in the unbypassed emitter resistance (gain reduction) will result in a reduction of PM noise in the amplifier. This agrees with experimental data showing that an increase in the unbypassed emitter resistance does reduce the PM noise [2,3]. The problem with this approach is that the gain decreases at the same rate that the AM and PM noise decrease[1]. Using the noise equations one can derive guidelines for the design of low 1/f noise amplifiers without having to sacrifice the gain. These guidelines are listed below.

1. Large dc current.

The contribution of the current noise in both the AM and PM noise can be reduced by increasing the dc current through the transistor. For example, let's assume that the current noise term dominates the AM noise; that is

[1] The rate at which the PM noise decreases will depend on which term dominates the noise. Looking at Eq. 4, all terms have the factor G_o^2 except the current noise term which goes as G_o^4. Therefore if the current noise term dominates, the PM noise will be reduced at twice the rate that the gain is reduced.

$$\frac{S_a(f)}{2} = \frac{1}{4}\left(\frac{r_e}{r_e + R_E + r_g/\beta}\right)^2 \gamma^2. \qquad (5)$$

As I_{dc} increases, r_e ($\propto I_{dc}^{-1}$) decreases, and so does the term inside the bracket in Eq. 5 for a specific unbypassed emitter resistance. The smaller r_e, the more effective is the unbypassed emitter resistor in reducing the AM and PM noise due to current fluctuations. It is also worth mentioning that γ ($\Delta I_B/I_B$) is not necessarily a constant, but depends on the noise process from which it originates, and can be a function of operating point and transistor type.

2. Use a servo to reduce current noise.

For amplifiers in which the current noise is a problem (even for large I_B) a servo loop can be used to reduce the dc current noise. For the most part, the transistors used in our circuits had very low current noise; thus transconductance fluctuations did not limit the AM (or PM flicker noise). Other transistors might have larger current noise and the addition of a servo to reduce the current noise could be needed to achieve low levels of AM and PM flicker noise.

3. Use transistors with low input and output capacitances.

As mentioned in [1], flicker PM noise is caused by fluctuations in the phase shift of the amplifier. In general, these fluctuations are proportional to the input and output capacitances of the transistor, so it is to be expected that transistors with small capacitances (high f_T) result in better PM noise. This might not be true for all cases, since the PM noise depends on other factors such as current noise and impedance fluctuations. If the current noise in high f_T transistors is large this might turn out to be a problem for both AM and PM noise. It is therefore important to know the current noise level of the transistor to be used, in addition to the input and output capacitances and the f_T.

4. Large collector-base voltage.

The dominant term in the PM noise equation (for a CE amplifier) is the collector-base capacitance fluctuations term. The collector-base capacitance is a junction capacitance which depends on the dc voltage at the collector-base terminals (V_{CB}). This capacitance has the form

$$C_{bc} = \frac{K}{(V_{bi} + V_{CB})^n}, \quad 0.1 \le n \le 4, \qquad (6)$$

where V_{bi} is the built in potential of the p-n junction, K is a constant that includes the permittivity of the semiconductor and the doping concentration, and n is a parameter that depends on the doping profile of the junction. From Eq. 6 it is obvious that fluctuations in V_{CB} will result in fluctuations of the collector-base capacitance. The sensitivity of PM noise to V_{CB} fluctuations is easily obtained from:

$$\left(\frac{\partial\phi}{\partial V_{cb}}\right)^2 = \left(\frac{n\omega G_o r_g K}{(V_{bi} + V_{CB})^{n+1}}\right)^2. \qquad (7)$$

Equation 7 demonstrates that higher V_{CB} results in smaller sensitivity of PM noise to collector-base voltage fluctuations (C_{bc} fluctuations).

5. Use filters to reduce voltage noise at transistor terminals.

In many cases fluctuations at the transistor terminals are the result of noise in the power supply. While a large collector-base voltage (V_{CB}) will reduce the effect of these fluctuations in the PM noise, it is also useful to have some kind of filtering of the power supply noise to reduce the effective collector-base voltage noise.

6. Small dc gain.

In addition the amplifier should be designed to have a small dc gain. The reason is that the fluctuations at the base terminal show up at the collector amplified by the dc gain; therefore a small dc gain will help minimize the effect of ΔC_{bc} in the PM noise.

7. Work at small source impedance.

From [1], the phase shift of a CE amplifier (when neglecting C_{be}) is given by

$$\delta = \frac{\omega C_{bc} R_L (r_e + R_E + r_g)}{r_e + R_E + r_g/\beta}, \qquad (8)$$

where $r_g = R_S \| R_{BIAS} + r_{bb}$. For a high gain CE amplifier with $r_g > r_e + R_E$, the phase shift is proportional to r_g. The use of sources with small impedance will help reduce the phase shift and thus the PM noise performance. Both the ΔC_{bc}

term and the γ ($\Delta I_B/I_B$) term are proportional to r_g, and a reduction in R_S will reduce both these terms.

8. Be aware of the noise in other components used (resistor, capacitors, inductors).

The flicker noise in resistive, capacitive and inductive components that make up the rest of the amplifier can also result in PM and AM noise if their levels are high. Several terms in the AM and PM noise equations are the result of impedance fluctuations. By carefully choosing low noise components these terms can be reduced considerably. Inductive components, which might be very convenient to use since they provide large impedances without the power dissipation of resistors, have flicker noise that depend on the type of core used. To prevent noisy inductors use air cores whenever possible.

As mentioned earlier, some of these guidelines might not apply to other configurations. In general, a large standing current helps reduce AM noise due to current fluctuations, but in some cases, this might jeopardize the PM noise. For example, in cases in which the base-emitter diffusion capacitance (C_{be}) appears in the dominant pole, a large current might degrade the PM noise since C_{be} depends on the dc collector current. An increase in the current will then cause a decrease in the pole frequency (increase in the phase shift). In addition, current fluctuations will be up-converted to PM noise through C_{be}.

One must also consider thermal noise, shot noise, and heat dissipation in the transistor when designing the amplifier.

Measurement Systems

The low noise levels of our amplifiers required the use of two-channel cross-correlation measurement systems [4,5]. The PM noise was measured using the system shown in Fig. 2. This setup has two similar phase noise detectors that are fed into a two-channel cross-correlation Fast Fourier Transform (FFT) signal analyzer. The phase shifter in each detector is adjusted so that the input signals to the mixer are in quadrature. The power spectral density (PSD) of the output voltage of channels 1 and 2 includes PM noise of the test amplifier, noise due to the components in the channel (mixer, amplifier, and FFT analyzer), and a fraction of the AM noise in the source that was "converted" to PM noise in the mixer. The AM to PM conversion factor in a double balance mixer is between -15 dB and -25 dB. The cross-correlation of channels 1 and 2 includes the PM noise in the amplifier and the converted AM noise from the source (see Eqs. 9-11). The uncorrelated noise (from the independent channels) is averaged away as $N^{-1/2}$, where N is the number of averages.

$$PSD(V_{n1}) = PM_{amp} + NOISE_{channel\ 1} + \beta_a AM_{source}, \quad (9)$$

$$PSD(V_{n2}) = PM_{amp} + NOISE_{channel\ 2} + \beta_a AM_{source}, \quad (10)$$

$$PSD(V_{n1}XV_{n2}) = PM_{amp} + \frac{NOISE_{channels\ 1,2}}{\sqrt{N}} + \beta_a AM_{source}. \quad (11)$$

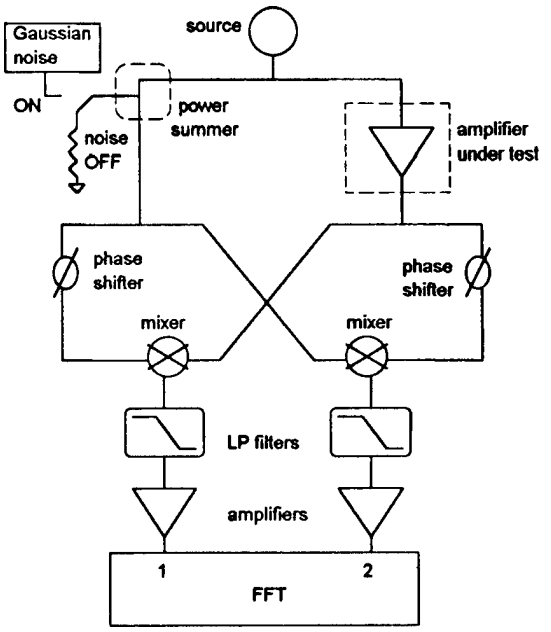


Fig. 2. PM noise measurement system for amplifiers.

In our case the noise floor of the system was limited by the AM noise of the source. At a carrier frequency of 5 MHz the noise floor was $\mathcal{L}(10\ Hz)_{floor} \cong$ -170 dBc/Hz. At a carrier frequency of 100 MHz, the noise floor was $\mathcal{L}(10\ Hz)_{floor} \cong$ -162 dBc/Hz.

The AM noise was measured using the system shown in Fig. 3. In this system a low noise source drives the amplifier under test. After the amplifier, the power is split and each leg is fed into an AM detector. The AM detector consists of a power splitter, a phase shifter, and a mixer. The phase shifter is adjusted so that the two inputs to the mixer are in phase. The PSD of the output voltage of channels 1 and 2 includes AM noise from the source and the amplifier, and noise from the channel. The cross-correlation of the voltages of channels 1 and 2 includes only AM noise from the source and the amplifier since the uncorrelated noise (from the individual channels) is averaged away as $N^{-1/2}$ (see Eqs. 12-14). Since in many cases the noise of the source was larger than the amplifier noise, a limiter was used to reduce its AM noise. The approximate noise floor achieved with this setup was $1/2S_\alpha(10\ \text{Hz})_{\text{floor}} \cong -162$ dBc/Hz for a carrier frequency of 5 MHz.

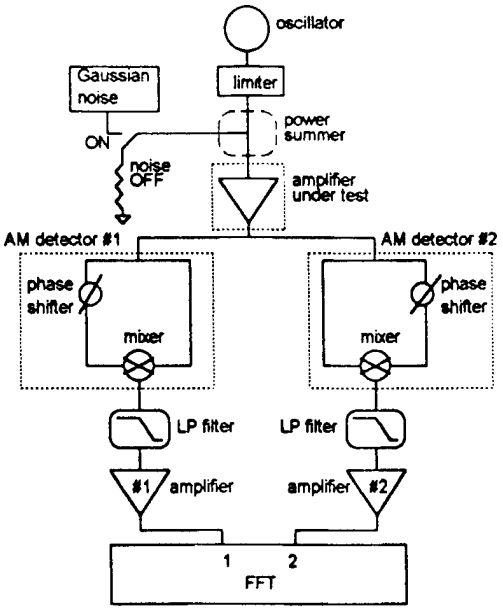


Fig. 3. AM noise measurement system for amplifiers.

$$\text{PSD}(V_{n1}) = \text{AM}_{\text{amp}} + \text{AM}_{\text{source}} + \text{NOISE}_{\text{channel 1}}, \quad (12)$$

$$\text{PSD}(V_{n2}) = \text{AM}_{\text{amp}} + \text{AM}_{\text{source}} + \text{NOISE}_{\text{channel 2}}, \quad (13)$$

$$\text{PSD}(V_{n1}XV_{n2}) = \text{AM}_{\text{amp}} + \text{AM}_{\text{source}} + \frac{\text{NOISE}_{\text{channels 1,2}}}{\sqrt{N}}. \quad (14)$$

Experimental Results on Amplifier Flicker Noise

This section shows experimental noise results on different amplifier designs and configurations. Measurements were done on amplifiers at 5 MHz and 100 MHz carrier frequencies.

The first design considered was a 5 MHz CE amplifier. To emphasize the importance of the biasing scheme and the components used in the amplifier, we started with a poor design for flicker noise. Figure 4 shows the initial CE design. In this amplifier the dc current is 6 mA and the collector-base voltage is 2.8 V. No filtering of power supply noise is included. The large collector resistor combined with the small unbypassed emitter resistance (5Ω) result in a high rf gain of 22 dB. The dc gain is approximately 0 dB. The 4:1 transformer at the output is used to reduce the output impedance of the amplifier.

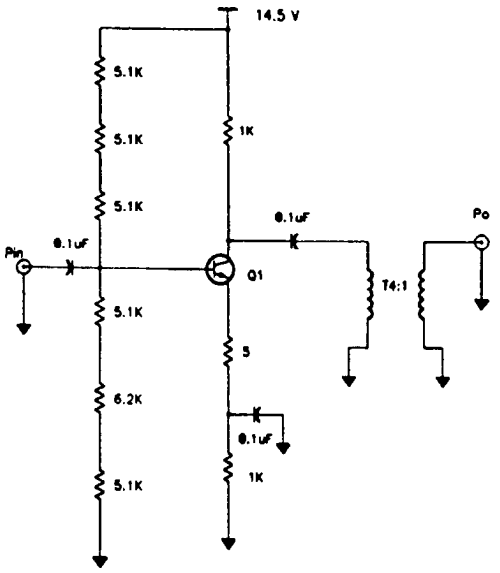


Fig. 4. Circuit diagram of CE amplifier #1.

The first column of Table 1 shows some parameters for the transistor used in this amplifier (2N2222A). Notice the low level of flicker current noise and the low noise figure (F). [Current noise

values were obtained from a small number of samples. Values might vary from transistor to transistor.]

	Transistor in CE #1 (2N2222A)	Transistor in CE #2
f_T	300 MHz	8 GHz
C_{be}	25 pF	1.6 pF
C_{bc} [V_{CB} = 9 V]	8 pF	0.65 pF
F	≤ 4 dB	1.5 dB
current noise @ 1 Hz ($I_E \cong$ 22 mA)	-170 dBA_{rms}/Hz	-164 dBA_{rms}/Hz

Table 1. Parameters for transistors used in the CE amplifiers.

The AM and PM noise results for this CE amplifier are shown in Fig. 5. The thermal noise level is approximately -154 dBc/Hz. The reason for such a high level was the low output power level (2 dBm) needed to avoid compression. (Due to the small V_{CB} (collector-base dc voltage) and the small dc current, the output power of the amplifier started to compress at 2 dBm.) The flicker noise level is similar for AM and PM noise; at 2 Hz it is approximately -138 dBc/Hz.

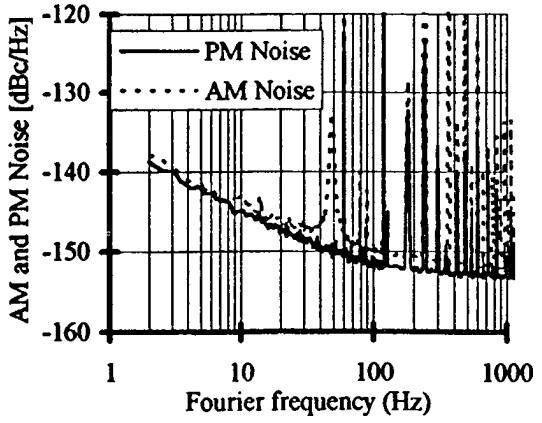


Fig. 5. AM and PM noise for CE amplifier #1.

A separate CE amplifier was designed and built according to the guidelines discussed earlier. The circuit diagram of this CE amplifier is shown in Fig. 6. The dc collector current is 23 mA and V_{CB} = 9 V. Several filters for the supply noise were added, including a darlington pair configuration at the output of the power supply, and several RC filters at the base bias resistive chain. In addition, a transistor with very small input and output capacitances was used (see Table 1, column 2 for transistor parameters). A large inductor at the collector was added in parallel to the 1 kΩ resistor. This achieves a low dc gain and does not affect the gain at the rf frequency of operation (5 MHz) due to the inductor's large impedance at that frequency. We expect a gain similar to the gain of the CE amplifier #1 since R_C, R_E (unbypassed emitter resistor), and the output transformer are the same. The resulting rf gain is 24.7 dB, somewhat higher than the gain of the CE amplifier #1, primarily due to the larger dc current.

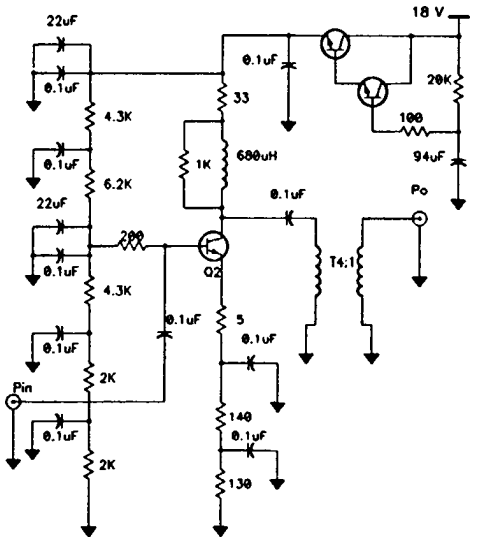


Fig. 6. Circuit diagram for CE amplifier #2.

Figure 7 shows the AM and PM noise results for the CE amplifier #2. The thermal level can be calculated using Eq. 15:

$$\mathcal{L}(f)_{thermal} = \frac{1}{2} S_\phi(f)_{thermal} = \frac{1}{2} S_a(f)_{thermal} = \frac{2kTFG}{P_o}. \quad (15)$$

For this amplifier G = 24.7 dB, P_o = 13 dBm , so the thermal noise for a F = 0 dB is

$$1/2\ S_a(f)_{thermal} = \mathcal{L}(f)_{thermal} = -174.1\ dBm + 24.7\ dB - 13\ dBm = -162.4\ dBc/Hz.$$

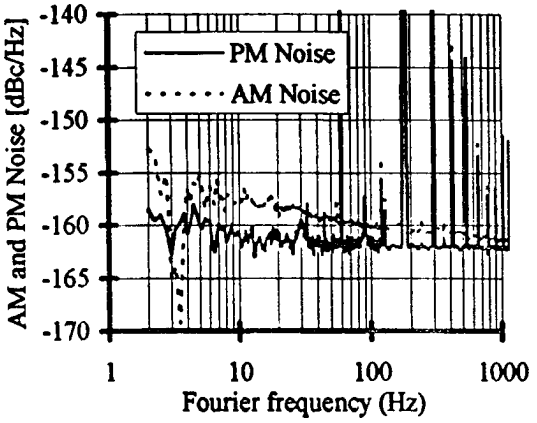


Fig. 7. AM and PM noise for CE amplifier #2.

From the PM noise trace, the thermal level is approximately -162 dBc/Hz; therefore the F of the transistor is very small, probably less than 1 dB. At 10 Hz from the carrier, the PM noise level is $\mathcal{L}$(10 Hz) = -161 dBc/Hz, only 1 dB above thermal. At 2 Hz from the carrier, the PM level is approximately $\mathcal{L}$(2 Hz) = -159 dBc/Hz. This number includes both flicker and thermal noise; the flicker level can be calculated by subtracting the thermal noise.

$$\mathcal{L}_{flicker}(2\text{ Hz}) \cong 10 \log (10^{-15.9} - 10^{-16.2}) \cong -162 \text{ dBc/Hz.}$$

The AM level at 10 Hz from the carrier is about $(1/2)S_a(10\text{ Hz}) \cong -158$ dBc/Hz. This value is limited by the noise floor of the measurement system (-162 dBc/Hz) and the thermal noise level (-162 dBc/Hz). The actual flicker level can be calculated by

$$(1/2)S_{a,flicker}(10\text{ Hz}) \cong 10 \log (10^{-15.8} - 10^{-16.2} - 10^{-16.2}) \cong -165 \text{ dBc/Hz.}$$

Comparing results for the two different CE amplifiers, we see that at 2 Hz from the carrier, the PM noise was reduced by approximately 20 dB, and the AM was reduced by about 15 dB. These numbers include the contributions of the thermal noise and the measurement system noise; thus the actual reduction of 1/f noise is probably larger.

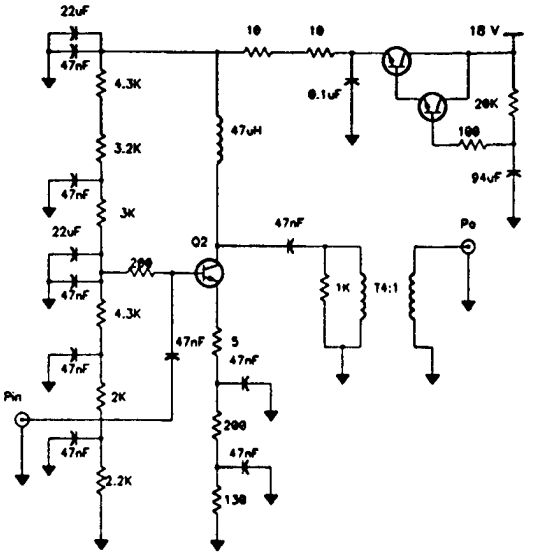


Fig. 8. Circuit diagram for low noise 100 MHz CE amplifier.

A similar CE amplifier design was tried at a carrier frequency of 100 MHz (Fig. 8). This amplifier has a dc current of 19 mA, a large collector-base voltage (9 V), a small dc gain (less than .05), and an ac gain of 18.5 dB. As in the 5 MHz CE amplifier #2, a high f_T transistor is used (Table 1, column 2); in addition, a darlington pair voltage regulator and RC filters are used to reduce noise voltage from the supply.

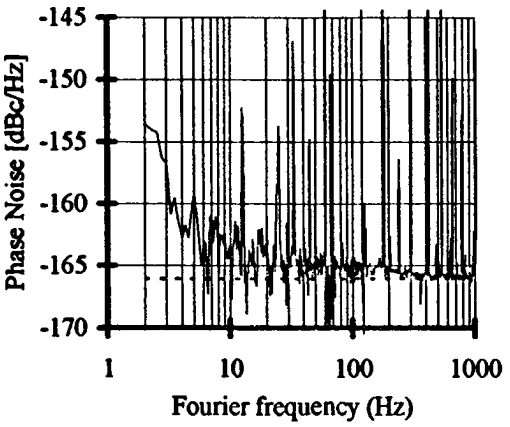


Fig. 9. PM noise for 100 MHz CE amplifier.

Figure 9 shows the measured PM noise for the 100 MHz CE amplifier just discussed. At Fourier frequencies above 100 Hz the PM noise is dominated by the thermal noise

(-166 dBc/Hz). This agrees with the theoretical value calculated from Eq. 15 for a gain of 18.5 dB, P_o = 11.5 dBm, and F = 1 dB:

$$\mathcal{L}(f)_{thermal} = -174.1\text{dBm} + 18.5\text{dB} - 11.5\text{dBm} + 1\text{dB} = -166.1\ \text{dBc/Hz}.$$

This calculated thermal noise is shown in the graph as a dotted line. At Fourier frequencies below 20 Hz the PM noise is limited by the measurement system noise. At 10 Hz, the measured PM noise is -162 dBc/Hz, and at 2 Hz the level is -154 dBc/Hz. These numbers are very close to the measurement system noise floor.

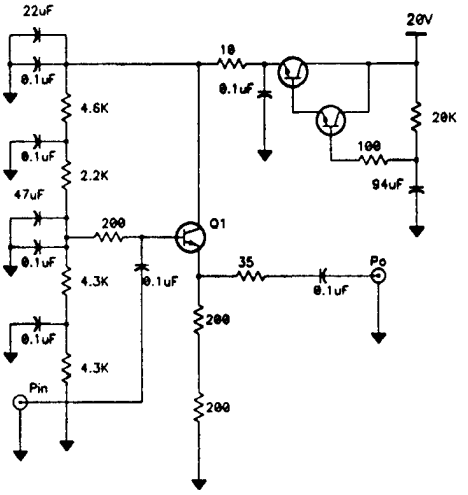


Fig. 10. Common collector amplifier.

Another BJT configuration studied was the CC amplifier. Figure 10 illustrates the circuit diagram of the 5 MHz common collector amplifier tested. In this CC amplifier V_{CB} = 8.5 V, and the dc collector current is 24 mA. Parameters for the transistor used are shown in column 1 of Table 1. Due to its small ac gain (-1.5 dB) and small phase shift, we expect (from theory) both the AM and PM noise of this amplifier to have very low levels. Figure 11 shows a plot of the measured PM and AM noise. As expected the levels are very low; at 10 Hz from the carrier $\mathcal{L}$(10 Hz) = -169 dBc/Hz and 1/2S_a(10 Hz) = -162 dBc/Hz. These values are approximately equal to the measurement system noise floor levels.

Due to its excellent noise characteristics, this stage can be used as a buffer stage either at the input or output of multiple stage amplifiers. One benefit that this stage could add to a CE amplifier if placed as the input stage is that it would decrease the input impedance to the CE stage, reducing the total phase shift of the CE. As discussed earlier, this should improve the PM noise of the CE stage.

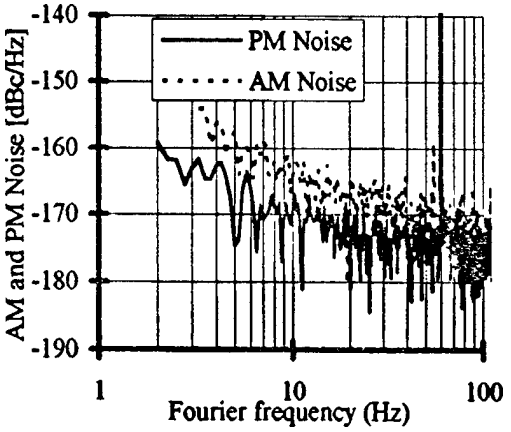


Fig. 11. AM and PM noise for 5 MHz CC amplifier.

Another amplifier to be discussed is the 5 MHz isolation amplifier described in [6]. This amplifier, shown in Fig. 12, consists of three common base stages that provide a high isolation between the input and output ports. The dc current through the collectors is approximately 28 mA and the ac gain of a single channel is approximately 7 dB.

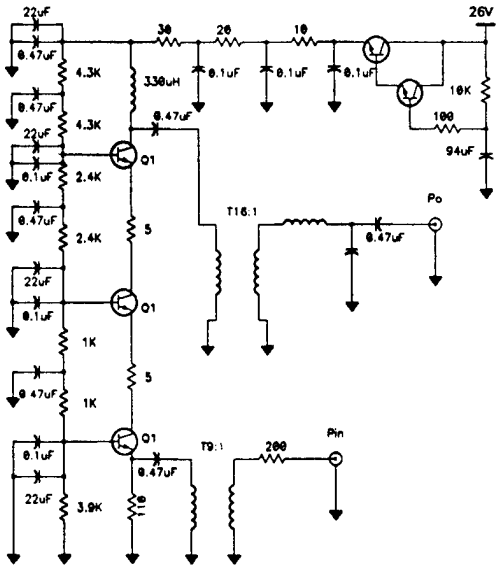


Fig. 12. 5 MHz isolation amplifier.

The details of the design are discussed in [6] and will not be repeated here. Additional filtering

of the power supply noise was included to further reduce the already low PM noise of this amplifier.

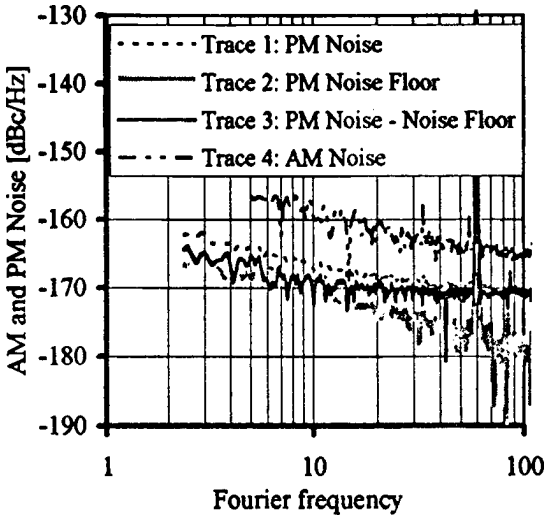


Fig. 13. PM and AM noise for 5 MHz isolation amplifier.

PM and AM noise results for this amplifier are shown in Fig. 13. Trace 1 (in Fig. 13) shows the measured PM noise for this amplifier. Notice that this trace is close to the noise floor of the measurement system (trace 2). Trace 3 was obtained by subtracting the noise floor of the measurement system from the measured PM noise and includes only the PM noise of the amplifier. The PM noise level at a Fourier frequency of 10 Hz is approximately $\mathcal{L}(10\text{ Hz}) = -168$ dBc/Hz and the thermal noise $\mathcal{L}(f)_{thermal} = -171$ dBc/Hz. The AM noise seems to be limited by the noise floor of the AM measurement system. At 10 Hz from the carrier $1/2\ S_a(10\text{ Hz}) = -160$ dBc/Hz, compared to the noise floor level of -162 dBc/Hz.

Conclusion

We have developed guidelines for designing CE amplifiers with high gain and low 1/f AM and PM noise. These guidelines minimize the up-conversion of baseband voltage noise, current noise, and impedance noise to noise about the carrier. According to the theory, the larger the gain and the phase shift of an amplifier, the higher the levels of AM and PM noise. This was found to be true experimentally. CE amplifiers, which have the potential for high gains and large phase shifts had the highest noise levels, but by following the design criteria we were able to reduce their 1/f noise considerably with only a modest reduction in the gain. Our 5 MHz CE amplifier, with a 25 dB gain, had AM and PM noise levels better than -160 dBc/Hz at 10 Hz from the carrier. We also had a 100 MHz CE amplifier (18.5 dB gain) with PM noise of $\mathcal{L}(10\text{ Hz}) \leq -162$ dBc/Hz. AM and PM noise in CC and CB configurations was also studied. Due to their small phase shift and small gain these stages have very low 1/f AM and PM noise. Our 5 MHz CC amplifier had noise levels similar to the noise floor of the measurement systems [$\mathcal{L}(10\text{ Hz}) \leq -169$ dBc/Hz, $1/2S_a(10\text{ Hz}) \leq -162$ dBc/Hz]. In addition, our 5 MHz distribution amplifier (composed of 3 CB stages) had noise levels of $\mathcal{L}(10\text{ Hz}) \leq -168$ dBc/Hz, and $1/2S_a(10\text{ Hz}) \leq -160$ dBc/Hz.

Acknowledgments

The authors thank S.R. Jefferts, T.E. Parker, F.G. Ascarrunz and M. Sicardi for helpful discussions on the topic. This project was supported in part by PM Milstar (Army), Ft. Monmouth, NJ, and by the Calibration Coordination Group.

References

[1] F.L. Walls, E.S. Ferre-Pikal and S.R. Jefferts, "The origin of 1/f PM and AM noise in bipolar junction transistor amplifiers", these proceedings.
[2] D. Halford, A.E. Wainwright, and J.A. Barnes, "Flicker noise of phase in rf amplifiers and frequency multipliers: characterization, cause, and cure", Proc. 22nd Ann. Frequency Control Symp., 1968, pp. 340-341.
[3] S.G. Andressen and S.K. Neheim, "Phase noise of various frequency doublers", IEEE Trans. Instrum. Meas., Vol. IM-22, June 1973, pp. 185-188.
[4] F.L. Walls, S.R. Stein, J.E. Gray and D.J. Glaze, "Design considerations in state-of-the-art signal processing and phase noise measurement systems", Proc. 30th Ann. Frequency Control Symp., 1976, pp. 269-274.
[5] W.F. Walls, "Cross-correlation phase noise measurements", Proc. 46th Ann. Frequency Control Symp., 1992, pp. 257-261.
[6] C.W. Nelson, F.L. Walls, M. Sicardi and A. De Marchi, "High isolation distribution amplifier at 5 MHz and 10 MHz", Proc. 48th IEEE Freq. Control Symp., 1994, pp. 567-571.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ADVANCED PHASE NOISE SUPPRESSION TECHNIQUE FOR NEXT GENERATION OF ULTRA LOW-NOISE MICROWAVE OSCILLATORS

E. N. Ivanov, M. E. Tobar and R. A. Woode

Department of Physics, University of Western Australia, Nedlands, WA, 6907, Australia

Abstract

In this paper we report an advanced phase noise reduction technique developed to improve the short term frequency stability of the microwave oscillators. The technique is based upon the ultra sensitive microwave frequency discriminator with the effective noise temperature close to its physical temperature. The frequency discriminator comprises a room temperature sapphire loaded cavity operating on a "whispering gallery" mode with an unloaded Q factor of 185 000 at *9 GHz*. The phase noise spectral density of the microwave loop oscillator, incorporating such a discriminator as a sensor for the frequency servo, has been measured to be equal to $S_{\varphi}^{osc}(F) \approx -60 - 10\log_{10}\left(F^3\right)$ *dBc/Hz* at frequencies below a few *kHz*. This corresponds to a level of phase noise as low as *−120 dBc/Hz* and *−150 dBc/Hz* at offset frequencies *100 Hz* and *1 kHz* respectively.

Introduction

High-Q sapphire loaded cavities (SLC) excited in "whispering gallery" modes and operating at room temperature have previously allowed the design of very low-noise microwave oscillators with phase noise of the order of *−125...−128 dBc/Hz* at *1 kHz* offset from the carrier [1]. This performance is achieved by using a phase noise reduction technique where the SLC is used both as a band-pass filter in the loop oscillator and a dispersive element of a frequency discriminator (FD) [2]. The phase noise of such oscillators within the bandwidth of the frequency control system is limited mainly by the intrinsic noise sources in the FD, provided that the gain of the frequency servo is high enough.

Expanding the ideas of phase noise suppression discussed in [3-5] we have now constructed two room temperature *9 GHz* oscillators with an advanced FD and the exceptional phase noise performance corresponding to the oscillator phase noise spectral density S_{φ}^{osc} of *−150 dBc / Hz* at *1 kHz* offset frequency. This is probably the first time that the oscillator phase noise was reduced below the level of its amplitude noise for the room temperature devices.

The proposed phase noise reduction technique is quite general and can have valuable implications for any types of microwave oscillators.

Phase Noise Reduction Technique

The general configuration of the microwave loop oscillator with additional phase noise reduction system is shown in Figure 1. The suppression of the oscillator

0-7803-2500-1/95 $4.00 © 1995 IEEE

phase noise is provided by frequency locking the oscillator to the SLC resonance. This is practically achieved by applying the filtered signal from the FD to the voltage controlled phase shifter (VCP) in the microwave loop oscillator [1]. The FD and VCP represent the sensor and the frequency tuning element of the microwave frequency control system (FCS) respectively. The FD shown in Figure 1 comprises the SLC and the phase detector (PD) based on the microwave double balance mixer (DBM) and the phase shifter. The latter is required to set the two ports of the DBM in quadrature.

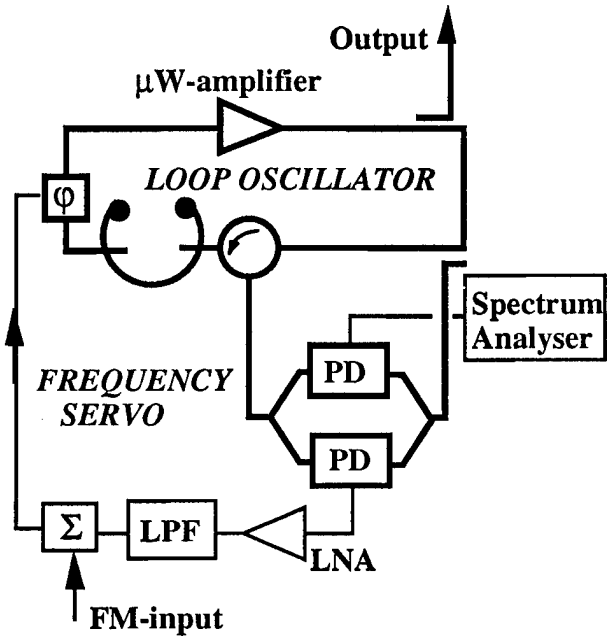


Figure 1. Loop oscillator with frequency servo and additional microwave readout system

The analytical expression for the spectral density of the oscillator phase noise at offset frequencies within the bandwidth of the high gain FCS is given by

$$S_{\varphi}^{osc}(F) = S_{\varphi}^{n/f}(F) + \frac{S_f^{res}(F)}{F^2}, \qquad (1)$$

where $S_f^{res}(F)$ is the spectral density of the SLC resonant frequency fluctuations and $S_{\varphi}^{n/f}(F)$ is the phase noise floor of the FD. Assuming that most of the noise sources in the FD are uncorrelated, the $S_{\varphi}^{n/f}(F)$ can be expressed as a sum of different components

$$S_{\varphi}^{n/f}(F) = \sum_i S_{\varphi}^{(i)}(F), \qquad (2)$$

representing the contributions from the particular noise sources.

Fluctuations of the phase difference between the ports of the PD $\Delta\varphi$ limit the noise floor of the FD at the level

$$S_{\varphi}^{(1)}(F) = \left\{\frac{1-\beta^2}{2\beta}\frac{f_{res}}{2Q_0F}\right\}^2 S_{\varphi}^{\Delta}(F), \qquad (3)$$

where f_{res} and Q_0 are the SLC resonant frequency and the unloaded Q-factor respectively, $S_{\varphi}^{\Delta}(F)$ is the spectral density of $\Delta\varphi$ fluctuations and β is the SLC effective electrical coupling coefficient given by $\beta = \beta_1/(1+\beta_2)$, where β_1 and β_2 are the SLC partial coupling coefficients.

Fluctuations of the $\Delta\varphi$ are caused either by the vibration sensitivity of the phase shifter for the mechanical device or by the intrinsic $1/F$ phase fluctuations for the voltage controlled device [6]. Assuming the following set of the experimentally measured parameters $Q_o = 185000$, $\beta_1 = 0.75$, $\beta_2 = 0.17$, $f_{res} = 9$ GHz and $S_{\varphi}^{\Delta}(1kHz) \approx -164$ dBc/Hz for the typical commercial VCP, yields the FD phase noise floor $S_{\varphi}^{(1)}(F)$ of the order of $-143dBc/Hz$ at 1 kHz offset frequency.

The finite effective noise temperature of the frequency discriminator T_{FD} imposes another limit on the FD noise floor

$$S_{\varphi}^{(2)}(F) = \frac{k_B T_{FD}}{P_{inc}}\frac{(1+\beta)^4}{4\beta^2}\left\{1+\left(\frac{\Delta f_{0.5}}{F}\right)^2\right\}, \qquad (4)$$

where P_{inc} is the power of the microwave signal incident on the resonator (SLC), k_B is the Boltzman constant,

$$\Delta f_{0.5} = \frac{f_{res}}{2Q_o}(1+\beta)$$

The effective noise temperature T_{FD} of the conventional FD can be expressed as

$$T_{FD} = T_o + T_{PD}, \qquad (5)$$

where the first term is due to the Nyquist noise in the transmission lines ($T_o \approx 300K$) and the second term is due to the voltage fluctuations at the phase detector output. For the PD based on the Watkins-Johnson M14A mixer the effective noise temperature T_{FD} was measured to be a function of the offset frequency F and power at

the RF port of the phase detector P_{RF}:

$$T_{FD} = T_o + \frac{A}{F} \quad \text{at} \quad P_{RF} \le -13dBm \qquad (6)$$

$$T_{FD} = T_o + \frac{B * P_{RF}^{\alpha}}{F} \quad \text{at} \quad -13dBm \le P_{RF} \le 3dBm$$

where $A = 3*10^7\ K*Hz$, $B = 5.6*10^8\ K*Hz$, $\alpha = 0.84$ and P_{RF} is in mW. Substituting $P_{RF} = 0.1\ mW$ in (6) yields $T_{FD} \approx 8*10^4\ K$ at $F = 1\ kHz$ which corresponds to $S_{\varphi}^{(1)}(1\ kHz) \approx -123\ dBc/Hz$ for the SLC with $Q_o = 185000$, $\beta = 0.85$ and $P_{inc} = 20\ mW$. This estimate for the phase noise is consistent with the experimental results obtained for the room temperature SLC oscillator of the same type as shown in Figure 1 [1].

To further reduce the oscillator phase noise, the FD with the lower effective noise temperature is required. This is achieved by introducing a low-noise microwave amplifier into the FD in a similar way suggested in [3]. The effective noise temperature of the advanced FD with microwave amplifier is given by

$$T_{AFD} = T_o + T_{AMP} + T_{PD}/G_{AMP}, \qquad (7)$$

where T_{AMP} and G_{AMP} are the microwave amplifier effective noise temperature and gain respectively.

The small signal operation of the microwave amplifier is essential for reduction of the FD effective noise temperature. This is because in the small signal regime, the gain of the amplifier is maximum, and its effective noise temperature is much less than T_{PD}. For low-noise high gain microwave amplifiers this approach allows the T_{AFD} to be effectively reduced from the level determined by the PD noise (5) to much lower limit $T_{AFD}^{min} = T_o + T_{AMP}$ set by the Nyquist noise in the transmission lines and the finite effective noise temperature of the microwave amplifier. For instance, the AFS4-08500960 Miteq amplifier enables the $T_{AFD}^{min} \approx 360\ K$ at $9\ GHz$ which is $23\ dB$ less than for the conventional FD.

Presence of the microwave carrier at the input of the amplifier results in $1/F^n$ terms in the dependence of T_{AFD} on offset frequency. The simplified analytical expression for the effective noise temperature of the advanced FD comprising the AFS4-08500960 microwave amplifier is given by

$$T_{FD} = T_{FD}^{min}\left\{1 + \frac{F_c}{F}\right\}^2, \qquad (8)$$

where the corner frequency F_c is proportional to the carrier power at the amplifier input $F_c(Hz) = 1.5*10^6 * P_{inc}^{amp}(mW)$.

The experimental dependence of the FD phase noise floor $S_{\varphi}^{n/f}(F)$ as a function of the offset frequency at $P_{inc} = 50\ mW$ is shown in Figure 2. For offset frequencies from $30Hz$ to $20kHz$ the noise floor

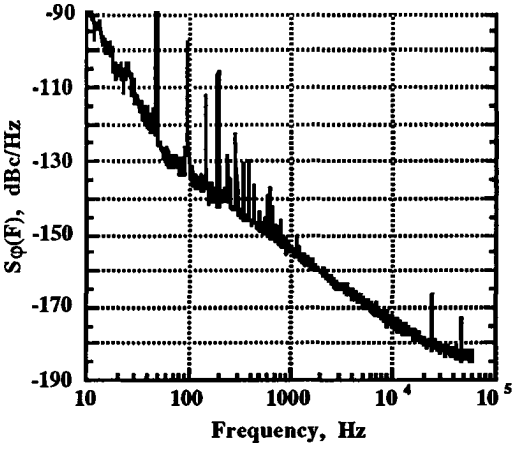


Figure 2. 9 GHz advanced frequency discriminator noise floor at $P_{inc} = 50\ mW$

varies as $S_{\varphi}^{osc}(F) \approx -95 - 10log_{10}(F^2)\ dBc/Hz$ resulting in $-155\ dBc/Hz$ at $F = 1\ kHz$. At low offset frequencies below $30Hz$ the slope of the $S_{\varphi}^{n/f}(F)$ becomes $-30\ dB/decade$. This indicates that the amplifier incident power P_{inc}^{amp} is of the order of $-50\ dBm$ (8).

It is worth mentioning that the performance of the advanced FD (Figure 2) is similar to that reported in [3].

All the details concerning the design and tuning of the advanced FD will be published shortly.

Oscillator Phase Noise Performance

The expected phase noise performance of the microwave loop oscillator with the advanced FD is shown in Figure 3. Curve 1 in Figure 3 shows the phase noise of the free-running loop oscillator. Curves 2 and 3 correspond to the predicted level of oscillator phase noise when the signal is taken either from the loop amplifier output or from the filtered port of the SLC respectively. According to these simulations the phase noise of the oscillator S_φ^{osc} at offset frequencies $100Hz$ and $1kHz$ is of the order of $-125dBc/Hz$ and $-147dBc/Hz$, respectively. This performance is achieved at $P_{inc} = 20\ mW$.

Increasing the microwave power P_{inc} enables further cancellation of the oscillator phase noise, provided that the microwave amplifier in the advanced FD is kept in the small signal regime. Assuming that the oscillator phase noise is primarily limited by the finite noise temperature of the advanced FD, values of $S_\varphi^{osc}(F)$ as low as $-160\ dBc/Hz$ appear to be attainable at $1kHz$ with $P_{inc} = 500\ mW$.

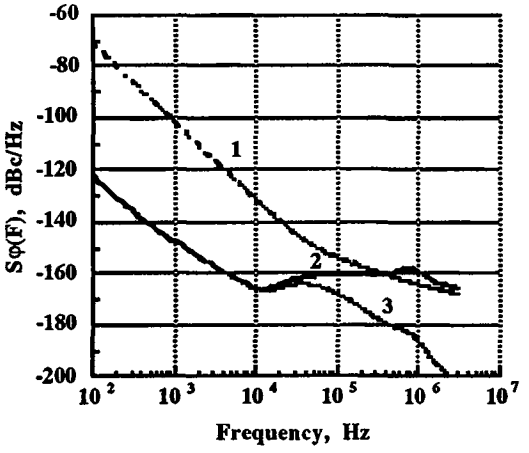


Figure 3. Predicted phase noise performance of the 9 GHz RT SLC oscillators: free running oscillator (1), frequency stabilised oscillator, unfiltered signal (2), frequency stabilised oscillator, SLC filtered signal (3)

The range of offset frequencies, within which the phase noise of the free-running oscillator is effectively suppressed, is limited by the bandwidth of the low-noise amplifier (LNA) in the feedback loop (Figure 1). Results of simulations shown in Figure 3 are obtained for the typical LNA (LT1028) with the gain-bandwidth product $G*B$ of the order of $60\ MHz$. The peak of the order of a few dB in the phase noise spectra of the locked oscillator, at frequencies near $1\ MHz$, (curve 2) is due to the excessive phase lag in the servo loop low-pass filter. Reducing the gain of the servo loop or the bandwidth of the loop filter minimises the magnitude of this peak. However, the oscillator phase noise degrades at offset frequencies above $10\ kHz$ because of these changes. For this reason if the low noise performance at offset frequencies above $10\ kHz$ is important, a more broad band LNA, like for instance AD829 with $G*B \approx 500\ MHz$, is required.

To measure the "electronic" component of the oscillator phase noise $S_\varphi^{(2)}(F)$, an additional FD based on the same SLC was used (Figure 1). This technique allows the "optimistic" estimate of the oscillator phase noise giving the minimal level of S_φ^{osc} which can be potentially achieved, provided that the SLC resonant frequency does not fluctuate and other components of the FD noise floor are negligible as compared with $S_\varphi^{(2)}(F)$. The results of these measurements are shown in Figure 4.

Curve 1 corresponds to the free running oscillator. Curves 2,3 and 4 correspond to the frequency stabilised oscillator at different levels of the DC servo gain: $38\ dB, 48\ dB$ and $58\ dB$ respectively. In the latter case the actual level of the "electronic" component of the phase noise at frequencies below $3...5\ kHz$ is $3\ dB$ better than that given by curve 4. This is because the results in Figure 4 represent the raw data without the correction on the finite noise floor of the additional FD. Subtracting $3\ dB$ from the measured raw data yields the low limit on the oscillator phase noise $\left\{S_\varphi^{osc}\right\}_{min}$ equal to $-98\ dbc/Hz$, $-128\ dbc/Hz$ and $-153\ dbc/Hz$ at offset frequencies $10\ Hz$, $100\ Hz$ and $1\ kHz$ respectively and with a power incident on the SLC of $P_{inc} \approx 50\ mW$.

To perform adequate measurements of oscillator phase noise, an external FD based on a second similar SLC was used. By changing the physical temperature of the second SLC, tuning of the external FD on the oscillator

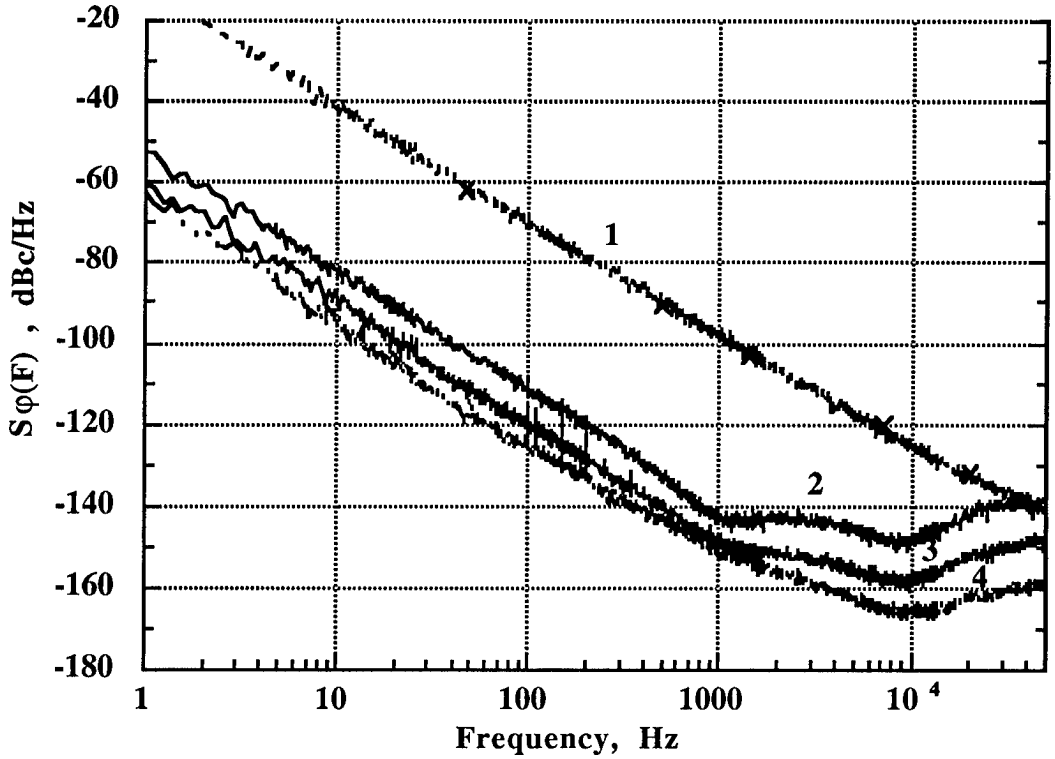


Figure 4. Electronic component of oscillator phase noise: free running oscillator (1), frequency stabilised oscillator with DC loop gain 38 dB (2), 48 dB (3) and 58 dB (4)

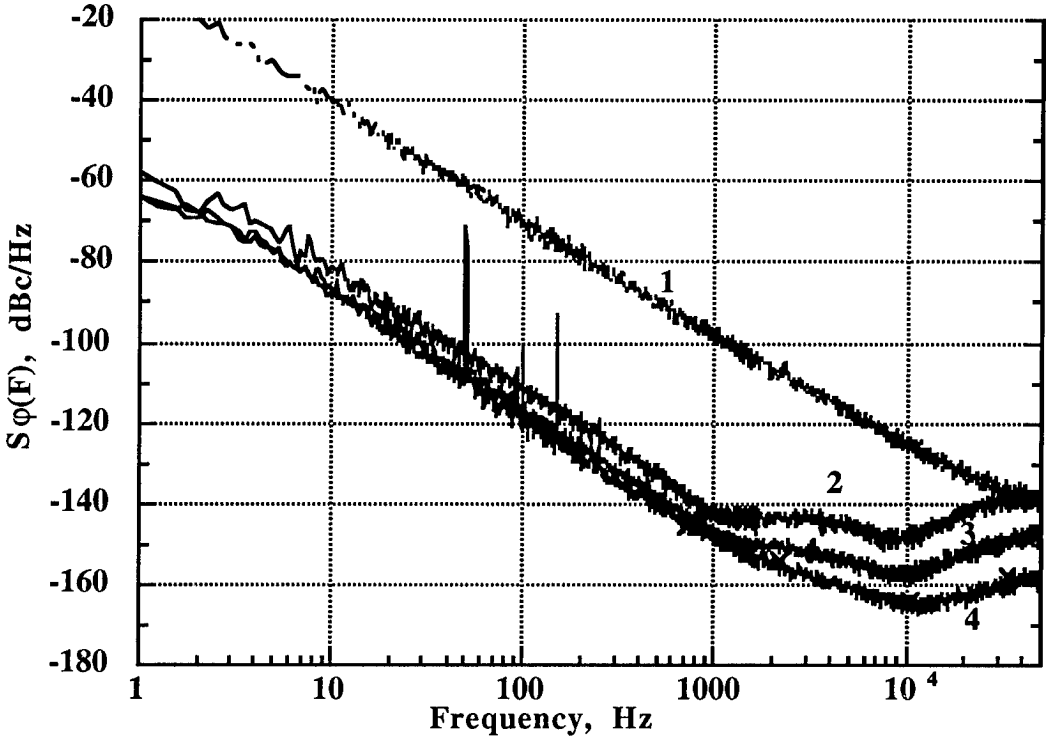


Figure 5. Oscillator phase noise spectral density: free running oscillator (1), frequency stabilised oscillator with DC loop gain 38 dB (2), 48 dB (3) and 58 dB (4)

operating frequency was achieved. The excellent temperature stability of both commercial SLC's available from Poseidon Scientific Instruments $\left|\frac{1}{f_{res}}\frac{df_{res}}{dT}\right| \approx 0.03\,ppm/^{o}C$ enabled optimal tuning of the readout system to be easily maintained during the whole period of the noise measurements.

The results of the oscillator phase noise measurements obtained with the external FD are shown in Figure 5. Curve 1 shows the phase noise of the free running oscillator, curves 2,3 and 4 correspond to the frequency stabilised oscillator at *38 dB, 48 dB* and *58 dB* levels of the DC loop gain.

A few bright lines at the noise spectra occur at frequencies 50, 100 and 150 Hz which correspond to mains harmonics. These bright lines can be in principle eliminated by using better quality power supplies and getting rid of the spurious ground loops.

It is worth mentioning that no traces of noise degradation caused by the vibration sensitivity of the different elements in the oscillator-discriminator system are visible in Figure 5. This is primarily due to the compact arrangement of the oscillator-discriminator system and improved vibration isolation.

A close resemblance between the noise spectra shown in Figure 4 and 5 justifies the previously described method of measuring the "electronic" component of the phase noise as an efficient tool for the study of the oscillator phase noise performance.

Like in the previous case the phase noise spectra in Figure 5 corresponding to high loop gains ought to be corrected to compensate for the measuring system noise floor. This is because the measured phase noise spectra $\tilde{S}_{\varphi}^{osc}(F)$ are essentially the following combinations

$$\tilde{S}_{\varphi}^{osc}(F) = S_{\varphi}^{osc}(F) + \left\{ S_{\varphi}^{n/f}(F) + \frac{S_f^{res}(F)}{F^2} \right\}_{ext}, \quad (9)$$

where the first term is the phase noise of the oscillator under test and the second term in brackets is the measuring system noise floor. The latter is a sum of the FD phase noise floor $S_{\varphi}^{n/f}(F)$ and the phase noise due to the SLC resonant frequency fluctuations $S_f^{res}(F)/F^2$.

Substituting in (9) the experimental values $\tilde{S}_{\varphi}^{osc}(1\ kHz) \approx -149\,dBc/Hz$, and $S_{\varphi}^{n/f}(1\ kHz) \approx -155\,dBc/Hz$ and neglecting $S_f^{res}(F)$ yields the estimate for the oscillator phase noise $S_{\varphi}^{osc}(1\ kHz) \approx -150\,dBc/Hz$.

Introducing a VCP into the advanced FD allows the fast electronic tuning of the oscillator with the rate of the order of $3kHz/V$. This experiments were performed in order to phase lock two oscillators and measure the phase noise by using the standard two oscillator technique. However, because of the intrinsic *1/F* phase noise in the VCP (see above) the phase noise of the electronically tuneable oscillator degraded to the level of $-144\,dBc/Hz$ at $1\ kHz$.For this reason the two oscillator method of the phase noise measurements did not allow us to see the full potential of the developed oscillator.

Another version of the two oscillator method where the phase locking is achieved by changing the SLC operating temperature should let us confirm our current results. In this case the cross-correlation measuring technique suggested in [7] has to be invoked in order to improve the phase noise floor of the single channel measuring system which is currently of the order of $-147\,dBc/Hz$.

Conclusion

We have demonstrated that a new phase noise reduction technique allows a real breakthrough in the filed of low-noise microwave oscillators. Using this technique we have designed the 9 GHz oscillator operating at room temperature with the SSB phase noise equal to $-150\ dBc/Hz$ at $1\ kHz$ offset frequency. This is *20 dB* better than presently achieved levels.

We have also reduced the vibration sensitivity of the oscillator which allowed us to minimise the degradation of the oscillator phase noise at low offset frequencies and to achieve S_{φ}^{osc} equal to $-89dBc/Hz$ and $-120\ dBc/Hz$ at $10Hz$ and $100Hz$ respectively.

Acknowledgments

This research was supported by the Australian Research Council in the framework of scientific collaboration between the University of WA and Poseidon Scientific Instruments. The authors would like to thank Mr. Andre Luiten from the UWA for the useful ideas regarding the

design of the low-noise oscillators

References

[1] M. E. Tobar, E. N. Ivanov, R. A. Woode and J. H. Searls, "Low noise microwave oscillators based on high-Q temperature stabilised sapphire resonators", Proc. 1994 IEEE International Frequency Control Symposium, 1-3 June 1994, pp.433-440.

[2] M. J. Bianchini, J. B. Cole, R. DiBiase., Z. Galani et all, "A single resonator Ga As FET oscillator with noise degeneration", in 1984 IEEE MTT-S Digest, pp. 270-273.

[3] D. G. Santiago and G. J. Dick, "Microwave frequency discriminator with a cryogenic sapphire resonator for ultra-low phase noise", Proc. 1992 IEEE Frequency Control Symposium, pp. 176-182.

[4] D. G. Santiago and G. J. Dick, "Closed loop tests of the NASA sapphire phase stabiliser", Proc. 1993 IEEE International Frequency Control Symposium, 2-4 June 1993, pp. 774-778

[5] D. P. Tsarapkin, "Low phase noise sapphire disk dielectric resonator oscillator with combined stabilisation", Proc. 1994 IEEE International Frequency Control Symposium, 1-3 June 1994, pp. 451-458.

[6] M. E. Tobar and D. G. Blair, "Phase noise of a tuneable and fixed frequency sapphire loaded superconducting cavity oscillator", in 1992 IEEE MTT-S Digest, pp.447-480.

[7] F. L. Walls, "Reducing errors, complexity, and measurement time of PM noise measurements", Proc. 1993 IEEE International Frequency Control Symposium, 2-4 June 1993, pp.289-297.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

AN EXTENSION OF THE ALLAN VARIANCE WITH INCREASED CONFIDENCE AT LONG TERM

D. A. Howe
Time and Frequency Division, National Institute of Standards and Technology
Boulder, CO 80303

ABSTRACT - I build a new statistic for the characterization of frequency stability of clocks and oscillators. It improves on the traditional Allan variance statistic especially at long averaging times.

INTRODUCTION AND SUMMARY

The impulse response of the commonly used technique of differencing of independent random variables overlaps (correlates) "odd" and "even" deviates except the first and last one if the data are not wrapped. This implies there are about twice as many deviates as there are actual degrees of freedom in the mean-square, and the resulting sample Allan variance is appropriately divided by 2(M-1). We obtain

$$\hat{\sigma}^2_{\bar{y}} = \frac{1}{2(M-1)} \sum_{k=1}^{M-1} (\bar{y}_{k+1} - \bar{y}_k)^2 = \frac{1}{2(M-1)} \sum_{k=1}^{M-1} (\nabla \bar{y}_{k+1})^2,$$

where $\{\bar{y}_{k'}\}$, k' = 1,2,3,...,N-1 are fractional frequency differences averaged over τ_0 and $\{\bar{y}_k\}$, k = 1,2,3,...,M are fractional frequency differences averaged over interval $m\tau_0$. Hence $\hat{\sigma}^2_{\bar{y}}$ is implicitly dependent on dimensionless quantity m, a scale parameter which for efficiency can be limited to rational powers of 2, i.e., 2^i=m, i=0,1,2,3,...

The sample Allan variance is useful as a power-law (octave band) spectral estimator but is time-shift (phase) sensitive and depends on where we start the calculation with respect to data in process. For large data sets and small scale values of m, the odd and even deviates average together in the equation for a fairly accurate estimation of a broadband spectral distribution or variance of first differences. The division by 2(M-1) is arguably due to overlapping two sets of deviates but ought to approach M-1 for an accurate estimate in the statistic as $m\tau_0 \rightarrow T/2$ since the first and last deviates do not overlap (unless the data are wrapped). Under most conditions, wrapping the data improves the estimate. However, at the largest scale, the estimate degrades for another reason, namely the sample variance has no response to an evenly symmetric function at this scale. These reasons cause an estimation error or what can be misinterpreted as a "bias" at long intervals in virtually all cases even though the estimation is actually unbiased.

Realizing this, we can re-express deviates in terms of "in-phase" and "phase-shifted" versions of sample variances. Examining this result reveals that the argument in the sample Allan variance above is essentially twice the in-phase variance only. Shifting the observation or sampling window by τ_0 and wrapping the data with an overall frequency difference removed will conveniently yield N-1 sample variances which can be averaged with the usual in-phase sample Allan variance. For $\{\bar{y}_{k'}\} = \bar{y}_1,\ldots,\bar{y}_{N-1}$, this statistic is given by

$$\hat{\sigma}^2_{total}(\tau) = \frac{1}{N-1} \sum_{j}^{N-1} \left[\frac{1}{2(M-1)} \sum_{k=1}^{M-1} (\nabla \bar{y}_{k+1,j})^2 \right]$$

where $\{\bar{y}_{k',j}\} = \bar{y}_{j+1}, \bar{y}_{j+2},\ldots, \bar{y}_{N-1}, \bar{y}_1, \bar{y}_2,\ldots,\bar{y}_j$ are spaced by τ_0 and $\{\bar{y}_{k'}\}$ is therefore wrapped and reindexed by j. $\{\bar{y}_{k,j}\}$ are averages implied over $\tau = m\tau_0$. This variance reduces estimation errors universally seen in previous treatments, thereby providing a much better estimate of frequency stability for measurement times longer than say 20% of the data length.

DISCUSSION

This paper presents a very simple example which shows a common and heretofore undiscussed pitfall in the calculation of the sample version of the Allan variance for long averaging times τ. The sample Allan variance is preferred over the modified variance for long-term frequency stability estimation since τ can be as long as T/2 whereas 50% more data length is required overall for the modified version

U.S. Government work not protected by U.S. copyright

using the same τ. The concepts of this paper can easily be extended to the modified variance however.

In frequency-standards metrology, measurements of average relative frequencies are passed through a first-order high-pass filter which is used for removing nonstationary behavior [Box and Jenkins, 1970]. A digital (discrete) version of such a filter (known as finite impulse response or FIR) is routine for creating stationary variates from original nonstationary variables with a "red" PSD having a power-law noise process no steeper than $\propto f^{-2}$ [Rabiner and Gold, 1975]. Furthermore it necessarily correlates otherwise random uncorrelated measurements assumed to be independent [Beran, 1992] [Box, Hunter, Hunter, 1978]. The sample variance of overlapping (not necessarily maximally overlapped) first-differences quantifies spectral features of the data using a smooth broadband, constant-Q, equivalent frequency-domain response [Allan, Weiss, Jespersen, 1991] [Howe and Percival, 1994]. This filter unfortunately has deep nulls at the reciprocal of the averaging time and its harmonics. For broadband noise, a sequence of many overlapping deviates is averaged; hence there is an average of as many arbitrary phases associated with an equivalent frequency-domain filter. However, for long scales, the nulls in the filter response can dramatically affect the result because averaging cannot be claimed as an advantage since only one or two deviates may be involved. The problem is that particular stochastic processes having even functional symmetry over the finite observation can put virtually all of the noise power precisely in the filter's nullpoints at the longest scale.

FIR filter concepts are used in this paper because they relate directly to the nth-order differences used in the statistics quantifying frequency stability. They also help clarify the problem addressed in this paper in using the sample Allan variance.

To begin, the N-sample standard variance is not convergent for the often encountered case of red noise processes of the measured relative phase of an oscillator $\{x_{k'}\}$; it is highly variable with averaging time $\tau = m\tau_0$ or correspondingly half of the reciprocal sampling frequency and its aliases. We assume that differencing techniques allow probability theory to be applied to such non-convergent time series. Differencing creates the new series $\{x'_1,\ldots,x'_{N-1}\}$ formed from the original series $\{x_1,\ldots,x_N\}$ by the "∇" operator as [Brockwell and Davis, 1987]

$$x'_k = x_{k+1} - x_k = \nabla x_{k+1}. \qquad (1)$$

Essentially, differences of high enough order "pre-whiten" measured data which is then subject to classical statistical treatment. A first-difference of average frequency has proven to be a simple and adequate whitening filter for oscillators [Barnes, et al., 1971]. Thus the two-sample (first difference) standard variance of frequency was one of the first suggested frequency stability measures [Allan, 1966]. Briefly, it is (using the notation in [Allan, 1966]):

$$\sigma_{\bar{y}}^2(\tau) = \sigma_{std}^2(2, \tau = T). \qquad (2)$$

All measurements are discrete, not continuous. Interval Δk is taken to be minimum (denoted in a time series by τ_0) with other longer intervals given by $m\Delta k$ (denoted by τ), where m = 1,2,3,....for efficiency however m is often limited to increments of rational powers of 2, i.e., $m=2^i$, i = 0,1,2,3,... More specifically, we start with assumed independent discrete variables $\{x_{k'}\}$, sampling (spatial or temporal) interval Δk (separating time marks $t_{k'}$ if a time series), and m which designates the shift or "stride" in an unprimed index k derived from k′ such that k is incremented once for every mth increment of k′. Hence k = k′/m, and $m\Delta k$ (i.e., τ) is the physical spacing or scaling of measurements within total observation M or T for k = 1,2,3,...,M. Recall that M and T are actually dimensions of space or time which are proportioned to dimensionless integers for mathematical convenience. Although confusing, averaging-time dimension "τ" is often used in the context of statistics having dimensionless Δk, k, and M such that $\frac{m\Delta k}{M} = \frac{m\tau_o}{T}$. Most commonly, the unprimed index k has an implied scale m and $\{x_{k,m}\}$ is a sequence with spacing $m\Delta k$ of measured phase differences between two oscillators; $\{\bar{y}_{k,m}\}$ are average frequency differences. That is

$$\bar{y}_{k,m} = \bar{y}_{k,m}(t_{k'} - m\tau_o, t_{k'}) = \frac{1}{m\tau_o} \int_{t_{k'}-m\tau_o}^{t_{k'}} y(t')dt' = \frac{x_{k'-m}(t_{k'}-m\tau_o) - x_{k'}(t_{k'})}{m\tau_o} = \frac{x_{k-1,m} - x_{k,m}}{m\tau_o}. \qquad (3)$$

Additionally, an important procedure assumed in this paper is that <u>overall</u> frequency difference ($\Delta\bar{y}_{1\to N}$) is removed. This means that $x_N = x_1$ which without loss could be set to 0. From eq (1) and eq (3) it

follows that $\{\bar{y}_{k'}\} = \bar{y}_1,\ldots,\bar{y}_{N-1}$ is such that $\bar{y}_{k'} = \frac{1}{\tau_o} x'_{k'}$, and generally

$$\bar{y}_{k,m} = \frac{1}{m\tau_o} x'_{k,m}. \quad (4)$$

By removing an initial time offset, we can construct $\{\bar{y}_{k'}\}$ from $\{x_{k'}\}$ and vice-versa.

If process $\{\bar{y}_{k,m}\}$ is white, eq (2) is equivalent to

$$\sigma^2_{\bar{y}}(m\tau_o) = \tfrac{1}{2}\langle(\bar{y}_{k+1,m} - \bar{y}_{k,m})^2\rangle, \quad (5)$$

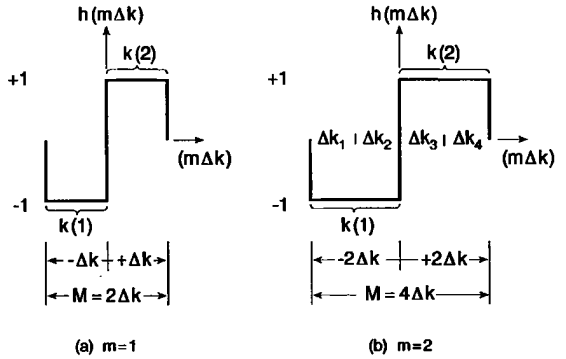
where $\{\bar{y}_{k,m}\}$ is an infinitely long series of nonoverlapping averaged discrete measurements of process y(t); the averaging time is given by $m\tau_0 = \tau$; "< >" denotes an infinite-time ensemble average; and k(origin) = t_0. We of course never can calculate this variance but can calculate a statistic which serves as an estimate. We often refer to a calculation of "the variance" when we usually mean "the statistic" or sample variance.

If series $\{\bar{y}_k\}$ is derived from a continuous function y(t), the fractional frequency difference of two oscillators, then we must infer that each sampled random variable is an average over $\tau = m\tau_0$. The recommended sample variance of first-differences of $\{\bar{y}_k\}$ (denoted hereafter as AVAR) is

$$\hat{\sigma}^2_{\bar{y}}(\tau) = \frac{1}{2(M-1)} \sum_{k=1}^{M-1} (\bar{y}_{k+1} - \bar{y}_k)^2 = \frac{1}{2(M-1)} \sum_{k=1}^{M-1} (\nabla\bar{y}_{k+1})^2. \quad (6)$$

Equation (3) can be combined with eq (6) for AVAR in terms of $\{x_{k'}\}$. The elements $\bar{y}_{k+1} - \bar{y}_k$ (called "deviates") corresponding to a particular value of τ derive from changes in the data having a corresponding frequency of $\frac{1}{2\tau}$ (and its aliases) and a corresponding phase given by the starting point at k=k(origin) [Howe, Allan, Barnes, 1981]. The sample variance is therefore not only dependent on τ but is additionally always phase-sensitive, i.e., dependent on where we start the calculation with respect to data in process.

Although this paper is not directly about an inconsistent statistic due to an improper number of degrees of freedom, the improper normalization or scaling by degrees of freedom for short data lengths also create an obvious inaccuracy or bias in its corresponding statistic. Briefly, inaccuracy or bias can be quantified using simulated data by an inconsistent trend (given by the slope) of the calculated statistic decomposed as a function of τ or equivalently what is called the power-law characterization (again the slope) of an estimated power spectral density (PSD). The division by 2(M-1) in eq (6) is justified by the argument that the two-sample variance in the infinite limit is the same as the standard variance (which divides by M-1 only) for a white Gaussian power-law process since the cross product of $(\bar{y}_{k+1}-\bar{y}_k)^2$ goes to zero yielding $(\bar{y}^2_{k+1}+\bar{y}^2_k)$, namely, the sum of two standard variances shifted by mΔk. For M = 2Δk and M = 4Δk, eq (6) has finite impulse response (FIR) or "convolving function" shown in figure 1(a) and 1(b).



Typical first-order high-pass impulse responses; mΔk=k.
Deviates ($x_{k+1} - x_k$) are from adjacent random variables.
The usual variance overlaps adjacent deviates.

Fig. 1

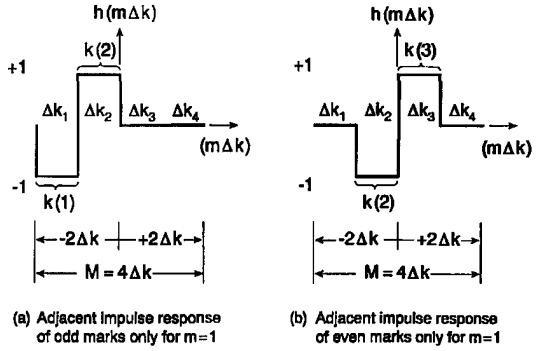
The sample variance of first-difference deviates is an average of these squared deviates at spacings of mΔk within the total observation M as shown in eq (6). Adjacent deviates are not independent. Each element $(\bar{y}_{k+1}-\bar{y}_k)$ is half of the preceding and following elements (except for the first and last). For example, element $(\bar{y}_3-\bar{y}_2)$ is not independent of preceding element $(\bar{y}_2-\bar{y}_1)$ since both elements equally include $\bar{y}_2$. Similarly, the following element $(\bar{y}_4-\bar{y}_3)$ contains $\bar{y}_3$. As a result the variance will correlate each element even though the original data is uncorrelated (as is the case with a white Gaussian process) [Yoshimura, 1976]. This correlation vanishes at the longest scale because deviates no longer overlap, that is, as mΔk → M, there are no overlapped deviates; in this case therefore the division (or normalization) should be M-1 in eq (6) rather than 2(M-1). Wrapping the $\{x_{k'}\}$ data allows 2(M-1) to be used even at the longest scale and hence is a way to get around this particular long-scale bias. Wrapping means that $x_\varsigma = x_{\varsigma \bmod N}$ for $\varsigma < 1$ and

$\varsigma > N$, i.e., $x_1 = x_{N+1}$. An assumption of stationarity implies that the wrap can be applied. The procedure also must include matching the endpoints (i.e., $x_N = x_1$) of $\{x_{k'}\}$ to avoid a step in the wrapped data. This, as mentioned, is easily done when an overall frequency difference is removed. We must eliminate the increment x_N to x_{N+1} in the wrap to avoid a potential bias since they are made the same value. Furthermore we must pay attention that N is even (N-1 is odd) because AVAR (in terms of $\{x_{k'}\}$) is a second difference of distinct phase values and could inadvertently respond with 0 at the longest possible scale if N is odd. Using $\{\bar{y}_{k'}\}$ derived from $\{x_{k'}\}$ eliminates the concern regarding whether the total should end up being odd or even.

Yet another pathological analysis error can occur. We first separate the variance into two nonoverlapped versions with odd and even indexes to emphasize that it is a sum of variances [see, for example, Jenkins and Watts, 1968]:

$$\sigma_{\bar{y}}^2(\tau) = \tfrac{1}{2}\langle(\bar{y}_{k+1} - \bar{y}_k)^2\rangle$$
$$= \tfrac{1}{2}[\langle(\bar{y}_{2k} - \bar{y}_{2k-1})^2\rangle + \langle(\bar{y}_{2k+1} - \bar{y}_{2k})^2\rangle] \quad (7)$$

where the first term on the right is the odd-indexed deviates and the second term is the even-indexed deviates.



(a) Adjacent impulse response of odd marks only for m=1

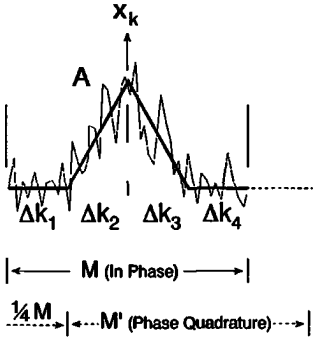
(b) Adjacent impulse response of even marks only for m=1

Fig. 2

To go further, note that deviate $\bar{y}_{k+1} - \bar{y}_k$ requires a total interval of $2m\Delta k$ and is centered at k. That is, $\bar{y}_k$ is derived from the interval spanning k-1 to k, and $\bar{y}_{k+1}$ is derived from span k to k+1. So the location of $\nabla\bar{y}_{k+1}$ is centered at k and its information content spans $2m\Delta k$ with odd indices at k = 1,3,5... and even indices at k = 2,4,6... as shown in figure 2 where m = 1. From this view, the localization of the odd and even marks of span $2\Delta k$ have a fixed phase relationship or phase difference of Δk corresponding to a half-period of a fundamental periodic change to which the deviate identifies and assigns an amplitude. Now consider AVAR as a power-law (constant-Q) spectral estimator [Allan, Weiss, Jespersen, 1991]. For a frequency component in $\{\bar{y}_k\}$ at $f = 1/(2\tau)$, the first term of the bottom expression of eq (7) detects an in-phase or 0° reference phase, and the second term detects an out-of-phase or 180° (relative to 0°) phase. By inspection, there is a fundamental change at the same frequency $f = 1/(2\tau)$ which can go undetected in the interval, namely, one-quarter and three-quarter period (or 90° and 270° phase-shifted) changes.

Deviations of a particular kind can go undetected in a $2m\Delta k$ (or 2τ) interval using AVAR. From a spectral point of view, the usual finite estimation of the original Allan definition is problematic because of this. From a time domain viewpoint the problem is that deviates are zero if the average frequency of the first interval equals the average frequency of the second which is true for an infinite set of even functions. The Allan statistic therefore has no response to any even function over the whole interval. As an example of the problem, hypothetical data are shown in figure 3 by the noisy plot over observation M. The statistic as described by eq (6) for $M = 4\Delta k$ will have deviates made from mean values $\bar{y}_k$ for segments k=1,...,4. A linear fit is indicated by the solid-line segments plotted with slopes 0, $+A/\Delta k$, $-A/\Delta k$, and 0 respectively. The resultant plot has even functional symmetry over M with

$$\bar{y}_1 = \bar{y}_4 = 0$$
$$\bar{y}_2 = \bar{y}_3 = A/2.$$



Example of even symmetric function over observation M

Fig. 3

According to AVAR, the sample statistic of eq (6), $\sigma^2(2\Delta k) = 0$. Obviously, the variance is not really 0 yet our estimate is 0. In this regard, determining and removing a polynomial (particularly a drift coefficient) over the interval has meaning only if the procedure is physically correct. Drift removal is model-dependent [Barnes, 1983]. A popular method involves quantifying drift as an overall second-difference of $\{x_{k'}\}$ which, when removed, results in the usual sample Allan variance being precisely 0 at the longest τ [Weiss, Allan, Howe, 1992] [Weiss and Hackman, 1992].

We replace the Allan two-sample function (which I will denote as "real" or "in-phase") shown in figure 1 and reproduced in figure 4(a) by a phase-shifted version of it to pick up variations in a phase-quadrature component of the data over the coarsest interval or longest scale of τ. This turns out to be a three-sample sequence over each interval which is shown in figures 4(b) and 4(c) and denoted as "imaginary" or "phase-quadrature" functions. The Allan two-sample in-phase function has odd functional symmetry in the interval 2τ as shown in figure 4(a) and hence is suitably an "odd sampling window function." The three-sample sequence has even symmetry in the same interval 2τ as shown in figures 4(b) and 4(c) given by sequence

$$\pm\,[\bar{y}_{k+3} - (\bar{y}_{k+2} + \bar{y}_{k+1}) + \bar{y}_k],$$

where the average frequency is taken over $\tau/2$ rather than τ.

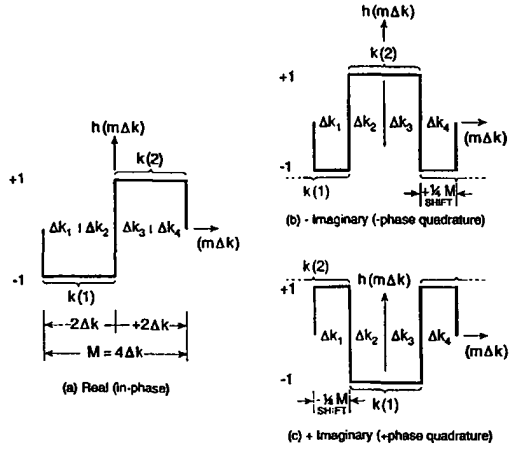


Fig. 4

Without loss eq (7) can be rewritten as

$$\sigma^2_{\bar{y}}(\tau) = \frac{1}{2}<\left[\frac{(\bar{y}_{k+3}+\bar{y}_{k+2})}{2} - \frac{(\bar{y}_{k+1}+\bar{y}_k)}{2}\right]^2>, \qquad (8)$$

$$\bar{y}_k = \bar{y}_k(t_k - \tau/2,\ t_k),\ \ k=1,3,5...$$

in which the variance is still for scale τ but where $\{\bar{y}_k\}$ are average frequencies over $\tau/2$ rather than τ.

Modified in the form of finite impulse responses shown in figure 4 and rewritten as a mean-square combination of an "in-phase" variance and "phase-quadrature" variance gives

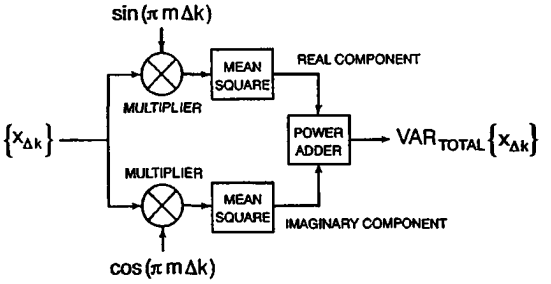
$$\sigma^2_{\bar{y}*}(\tau) = \frac{1}{4}\left[<\left[\frac{(\bar{y}_{k+3}+\bar{y}_{k+2})}{2} - \frac{(\bar{y}_{k+1}+\bar{y}_k)}{2}\right]^2> + <\left[\bar{y}_{k+3} - \frac{(\bar{y}_{k+2}+\bar{y}_{k+1})}{2} + \bar{y}_k\right]^2>\right], \qquad (9)$$

where the average frequency is taken over $\tau/2$.

Another variance can be defined from two variances of orthogonal in-phase and phase-quadrature elements. This formalism bears a similarity to the method of complex demodulation used in signal processing [Bingham, Godfrey, and Tukey, 1967]. In this method, time-series data are assumed to contain periodicities and viewed as consisting of separate real and imaginary parts; the variance is calculated for each part and the total is taken as

$$\sigma^2_{\Delta\,total} = Re|\sigma^2| \oplus Im|\sigma^2| = [Re^2\sigma^2 + Im^2\sigma^2]^{1/2}, \qquad (10)$$

where "$\oplus$" means power-added together. Functionally, the method of complex demodulation is shown in figure 5. The real part is comprised of



Method of complex demodulation

Fig. 5

deviates of the actual data (the in-phase observation) and the $\pm$ imaginary part are the deviates created by shifting the observation window by $\pm$ ¼T (the phase-

quadrature observation). In the case of a first-order difference filter, we obtain impulse responses precisely as shown in figure 4(a) for the real part and figures 4(b) and 4(c) for the $\pm$ imaginary part. The sign of the imaginary part becomes unimportant as the deviates are squared. The imaginary part that is outside the actual data are wrapped; that is, h(M+k) = h(k). This is justified because one assumes the measurement is for periodic (sine and cosine) functions which are in process, that is, unwindowed and periodically extended.

The estimate to eq (9) is an equivalent description to an "in-phase" and "phase-quadrature" discrete functional component separation. Using this method, the variance of the even-symmetric function is computed by moving the observation window by m/2, for the "phase-quadrature" variance and adding this to the "in-phase" variance. This has a sample statistic given by

$$\hat{\sigma}^2_{\bar{y}}(m\tau_o) = \frac{1}{4(M-1)}\left[\sum_{k=1}^{M-1}(\nabla\bar{y}_{k+1,m})^2 + \sum_{k=1}^{M-1}(\nabla\bar{y}^*_{k+1,m})^2\right], \quad (11)$$

$$where\ \bar{y}_{k,m} = \frac{x_{k'-m}(t_{k'}-m\tau_o) - x_{k'}(t_{k'})}{m\tau_o},$$

$$and\ \bar{y}^*_{k,m} = \frac{x_{k'-m/2}(t_{k'}-\frac{m\tau_o}{2}) - x_{k'+m/2}(t_{k'}+\frac{m\tau_o}{2})}{m\tau_o},$$

$x_1, x_2, \ldots x_N$ with wrap such that $x_\varsigma = x_{\varsigma \bmod N}$ for $\varsigma < 1$ and $\varsigma > N$, i.e., $x_1 = x_{N+1}$ which reindexes to x_N as noted earlier. Shifting the data and using a wrap simplifies the form of the sample variance of eq (11) corresponding to eq (9) [for a discussion of wrapped data, see Howe and Percival, 1995; also briefly in Bloomfield, 1976]. Equation (11) is an average of variances whereas eq (10) implies an RSS for orthogonal (sin vs. cos) basis functions or an RMS for random functions. Using simulation of common power laws, the difference between an average and RMS of variances turns out to be negligible in practice. Returning to the previously mentioned hypothetical example and using eq (11), the result is not 0 but rather

$$\hat{\sigma}^2_{\bar{y}^*}(m=2) = \frac{1}{12}\left(\frac{A}{2}\right)^2.$$

HIGHER ORDER DIFFERENCES

Given continuous function f(k) with nth order derivative ($\frac{d^n f(k)}{dk}$ where "k" is the independent variable), there are n derivatives for n possible orders. Continuous function f(k) is an idealization; all physical measurements will sample f(k) in discrete, usually equispaced, increments Δk_k despite the fact that Δk_k may be infinitesimally small. In all cases, therefore, we are representing f(k) as discrete continuous f(mΔk) which describes a space- or time-ordered sampled function with nth order possible differences. The variance is like an average of these squared differences. In this case, there are not n, but always 2^n independent differences in the average for n orders. To preserve independence, the proper variance is therefore the square root of the sum of squares of 2^n variances. To obtain stationary deviates, the order of differencing is usually quite small, frequently 1 or 2. This is because the corresponding (high-pass) impulse response makes new deviates which have a PSD which goes as f^{+2n} for each order n. Hence an f^{-4} PSD would need to be differenced only twice (order 2) to realize white Gaussian noise. Similarly, many functions can be well approximated on an interval of finite length by a polynomial of low degree. For the simple high pass first difference (order 1), there are 2^1 or two independent sets of differences and the proper variance is just the RMS of the two variances from each set as discussed in this paper. For the second difference (order 2), there are 2^2 or four independent sets of differences, and so on.

AN IMPROVED STATISTIC AT LONG TERM

I construct a sample variance (eq (11)) that is an average of variances of separate in-phase and phase-quadrature deviates (time-shifted by m/2) and given by a procedure which wraps $\{x_{k'}\}$ or equivalently $\{\bar{y}_{k'}\}$. This is sufficient to avoid a potentially serious estimation error compared to traditional AVAR at long averaging times. Historically the Allan variance is the composite variance of interest. It formally derives from the standard variance (see eq (2)). We compute various things from finite data which are calculations of statistics (one being AVAR) as estimators of the Allan variance. For a set of noise processes, a good estimator (1) has the same mean as the variance and (2) itself has a low variance. A question is whether the three-sample sequence introduced in this paper when combined with the

original AVAR implies a new definition having better confidence or is a better estimate of the existing definition. Better statistical certainty usually derives from more degrees of freedom. An in-phase convolving function (first term of eq (9)) is being combined with another convolving function at phase quadrature (second term of eq (9)). This is either an extension of the historically used two-sample variance or is a different concept.

We can establish an even better long-term estimator which describes and hence predicts longer-term frequency stability from available measurements. We can average all possible N-1 sample variances, each time shifted by τ_0. For $\{\overline{y}_{k'}\} = \overline{y}_1,\ldots,\overline{y}_{N-1}$, this statistic is given by

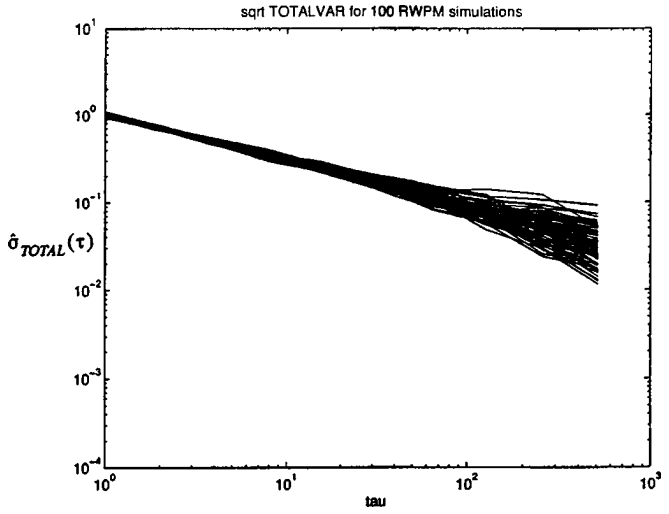
$$\hat{\sigma}^2_{total}(\tau) = \frac{1}{N-1}\sum_{j}^{N-1}\left[\frac{1}{2(M-1)}\sum_{k=1}^{M-1}(\nabla\overline{y}_{k+1,j})^2\right], \qquad (12)$$

where $\{\overline{y}_{k',j}\} = \overline{y}_{j+1}, \overline{y}_{j+2},\ldots, \overline{y}_{N-1}, \overline{y}_1, \overline{y}_2,\ldots,\overline{y}_j$ are spaced by τ_0 and $\{\overline{y}_{k'}\}$ is therefore wrapped and reindexed by j. $\{\overline{y}_{k,j}\}$ are averages implied over $\tau = m\tau_0$. Equation (3) can be combined into eq (12) for $\hat{\sigma}^2_{total}$ in terms of $\{x_{k'}\}$.

This statistic and its associated impulse responses average the variances of all possible fractional frequency deviates. The new statistic properly picks up and normalizes all functional variations at all τ scales, and at the longest ones, namely as $\tau \rightarrow T/2$. Its use is recommended for at least the longest four scales.

SIMULATION WITH RANDOM WALK OF $\{x_{k'}\}$

Figure 6(a) shows 100 plots of calculations of the square root of $\hat{\sigma}^2_{total}(\tau)$ for 100 simulations of random walk of phase modulation (RWPM). Equation (12) is the expression used for these calculations and N=1024 for each simulation. Each of the simulation averages of two-sample variances is equal to one. The bottom plot is the 100-simulation total mean and shows excellent agreement with theory. Figure 6(b) is the same set of calculations using traditional square root of maximally-overlapped AVAR. We see that the spread in the estimates is greater using AVAR instead of the new statistic in figure 6(a) especially at $\tau = T/2$. The new statistic is preferred at long averaging times since it yields a distribution which is less skewed and which is less susceptible to optimistic (sometimes very optimistic) estimation errors.



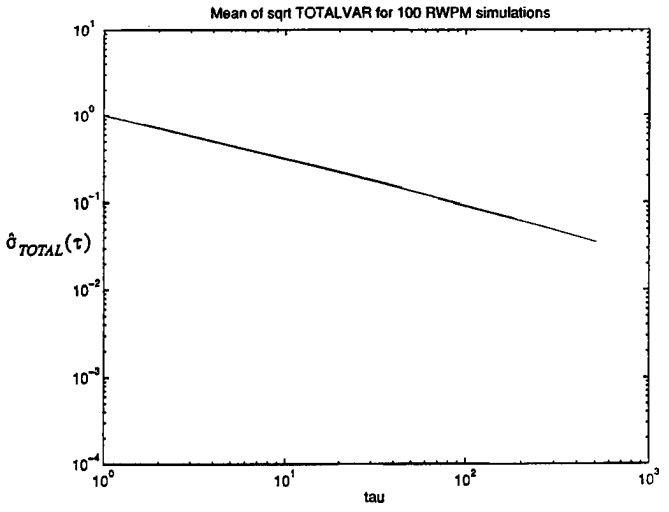
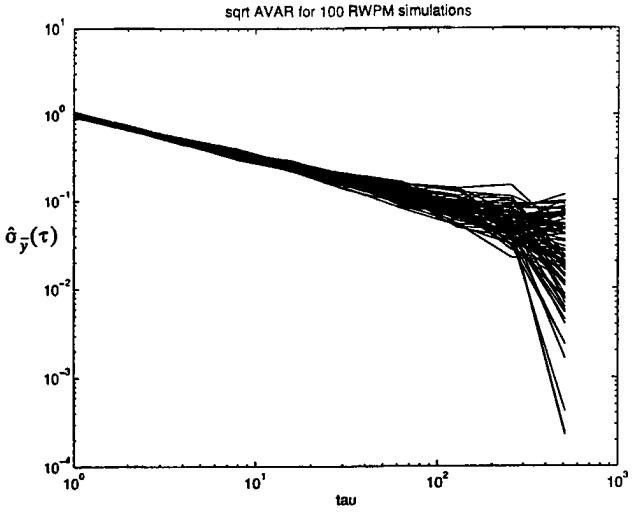


Fig. 6(a)

Top: New statistic (indicated as square root of TOTALVAR) calculated for 100 RWPM simulations with unit (two-sample) mean. Note the reduced skew and spread in the response for long term averaging times compared to traditional square root of maximally-overlapped AVAR shown at right at the top of figure 6(b).

Bottom: Mean of the 100 simulations using square root of TOTALVAR agrees satisfactorily with theory.



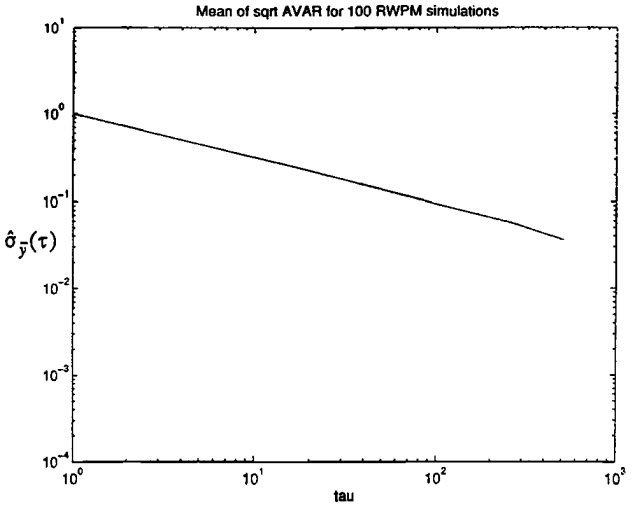


Fig. 6(b)

Top: For comparison, traditional square root of maximally-overlapped AVAR is calculated for the same 100 RWPM simulations as used at left for square root of TOTALVAR in figure 6(a). Note the wider skew and spread for long term averaging times and the tendency toward an optimistic (sometimes very optimistic) response.

Bottom: Mean of the 100 simulations using square root of maximally-overlapped AVAR is shown.

REFERENCES

Allan, D. W., Statistics of atomic frequency standards, Proc. IEEE 54, 221-230 (1966).

Allan, D.W. and Barnes, J.A., A modified "Allan Variance" with increased oscillator characterization ability, Proc. 35th Freq. Control Symp., 470-476 (1981).

Allan, D.W., Weiss, M.A., and Jespersen, J.L., A frequency-domain view of time-domain characterization of clocks and time and frequency distribution systems, Proc. 45th Frequency Control Symposium, 667-678 (1991).

Barnes, J.A., Atomic timekeeping and the statistics of precision signal generators, Proc. IEEE 54, 207-220 (1966).

Barnes, J. A., Chi, A. R., Cutler, L. S., Healy, D. J., Leeson, D. B., McGunigal, T. E., Mullen, J. A., Jr., Smith, W. L., Sydnor, R. L., Vessot, R. C. F., and Winker, G. M. R., Characterization of frequency stability, IEEE Trans. Instrum. Meas., IM-20, 105-120 (1971).

Barnes, J.A., The measurement of linear frequency drift in oscillators, Proc. 15th Ann. PTTI Meeting, 1983. Article is also contained in Natl. Inst. Stand. Technol. TN-1337, 1990.

Beran, J., Statistical methods for data with long-range dependence, Statistical Science, 7, 404-427 (1992).

Bingham, C., Godfrey, M. D., and Tukey, J. W., Modern techniques of power spectrum estimation, IEEE Trans. Audio Electroacoustics, AU-15, 56-66 (1967).

Bloomfield, P., Fourier analysis of time series: an introduction, Wiley, New York, 1976. Pages 55-60 are a useful discussion of the "circular definition" of a noise (or roughness) coefficient, namely the variance, using "wrapped" data.

Box, G. E. P., Hunter, W. G., and Hunter, J. S., Statistics for experimenters, Wiley, New York, 1978.

Box, G. E. P. and Jenkins, G. M., Time series analysis: forecasting and control, 1970, Holden-Day, San Francisco.

Brockwell, P. J. and Davis, R. A., Time Series: Theory and Methods, 1987, Springer-Verlag, New York.

Cutler, L.S. and Searle, C.L., Some aspects of the theory and measurement of frequency fluctuations in frequency standards, Proc. IEEE 54, 136-154 (1966).

Howe, D.A., Allan, D.W., and Barnes, J.A., Properties of signal sources and measurement mehtods, Proc. 35th Ann. Symp. Frequency Cont., Philadelphia, PA, May 27-29, 1981, A25-A35, 1981. Article is also contained in Natl. Inst. Stand. Technol. TN-1337, 1990.

Howe, D.A. and Percival, D.B., The wavelet variance, Allan variance, and leakage, Proc. IEEE Trans. on I & M, IM-44, no. 2, 94-97 (1995).

Howe, D.A. and Percival, D.B., The role of circularity in analyzing the variance of a time series, to be published. Contact D.A. Howe, NIST, 325 Broadway, Boulder, CO 80303.

Howe, D. A., Circular representation of infinitely extended sequences, to be published, Proc. 1995 IEEE Freq. Control Symp. (this issue).

Lindsey, W. C. and Simon, M. K., Telecommunication system engineering, 1973, Prentice-Hall, New Jersey.

Rabiner, L. R. and Gold, B., Theory and application of digital signal processing, 1975, Prentice-Hall, Inc., New Jersey.

Weiss, M.A., Allan, D.W., and Howe, D.A., Confidence on the second difference estimation of frequency drift, Proc. IEEE 46th Freq. Control Symp., 300-305, 1992.

Weiss, M.A. and Hackman, C., Confidence on the three-point estimator of frequency drift, Proc. 24th Ann. PTTI Meeting, 451-460, 1992.

Yoshimura, K., Characterization of frequency stability: Uncertainty due to the autocorrelation of the frequency fluctuations, IEEE Trans. Instrum. Meas., IM-27, 1-7 (1978). The number of degrees of freedom is about half the total number of deviates, so the scaling factor ends up being $\simeq 1/2(M-1)$ as shown in eq (6). It is exactly ½ in the infinite limit. This paper is concerned with the typical finite case and the special case of just a few deviates.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

CUT-OFF FREQUENCIES AND NOISE POWER LAW MODEL OF SPECTRAL DENSITY : ADAPTATION OF THE MULTI-VARIANCE METHOD FOR IRREGULARLY SPACED TIMING DATA USING THE LOWEST MODE ESTIMATOR APPROACH

F. Vernotte*, G. Zalamansky*, M. McHugh*, E. Lantz**
*** Observatoire de Besançon**
41 bis, avenue de l'Observatoire - B.P. 1615 - 25010 BESANÇON Cedex - FRANCE
☎ : (33) 81.66.69.22 - Fax : (33) 81.66.69.44 - E-mail : FRANCOIS@OBS-BESANCON.FR
**** Laboratoire d'Optique P. M. Duffieux, U. A. CNRS 214, Besançon, FRANCE**

Abstract - The concept of structure functions [1], which is an extension of the variance approach, is useful to determine the variance (the structure function) which is optimized for a type of noise and for an order of drift. The multi-variance method was developed to use different variances over the same signal. It is then possible to select a set of variances in which each variance is optimized to the determination of one parameter (of one noise level, drift, or cut-off frequency). Recently, we adapted this method to irregularly spaced timing data. In this connection, we replaced the structure functions by another method of spectral density estimation : the lowest mode estimator, introduced by J.E. Deeter and P. E. Boynton for the analysis of pulsar timing data [2, 3]. Different lowest mode estimators can be constructed according to two priorities : the order of drifts that must be removed and the type of noise for which the sensitivity must be maximum. Thus, a multi-variance system is developed using a set of different estimators. The details of this method are described and the results for different signals are discussed in this paper.

1. INTRODUCTION

In some applications of time metrology, the measurement cannot be regularly spaced in time. This problem occurs for instance in the case of time transfer using GPS satellites in common view, since it is necessary to wait to see the same satellite from the two clocks being compared. This is also what happens in the study of the pulsars, since pulsars cannot be observed at regularly spaced dates. However, it is of importance to measure the time stability in these cases.

Following the works of Groth [4], Deeter and Boynton [2, 3] developed a new method for the estimation of the power spectrum of the pulsar timing data : the lowest mode estimator. This approach may be considered as an extension of the structure function concept, introduced by Lindsey and Chie [1], to the case of irregularly spaced timing data, since the lowest mode estimator may be optimized to one type of noise.

We have shown in a previous paper [5] the importance of the choice of the spectral density estimator (variance or structure function), particularly in the case of the multi-variance method [6]. For instance, a very low frequency noise (f^{-3} or f^{-4} frequency noise which may occur in pulsar timings) could be confused with a linear frequency drift. In this case, a set of variances containing variances insensitive to linear frequency drift and converging for very low frequency noises associated with variances sensitive to linear drift must be chosen in order to determine both the low frequency levels and the linear drift coefficient.

In this connection, different lowest mode estimators may be defined according to an optimization to different types of noises and a sensitivity (or insensitivity) to different orders of drift. With such a set of lowest mode estimators, it is then possible to build a multi-estimator system which is the extension of the multi-variance analysis to the case of irregularly spaced timing data.

2. PRINCIPLE OF THE LOWEST MODE ESTIMATOR

2.1. Spectral density estimation

The Power Spectral Density (PSD) of time error of an oscillator may be modeled as a sum of power laws :

$$S_x(f) = k_0.f^0 + k_{-1}.f^{-1} + k_{-2}.f^{-2} + k_{-3}.f^{-3} + k_{-4}.f^{-4}$$

Moreover, for the millisecond pulsar study, a f^{-5} phase noise should be added.

On the other hand, since the instantaneous normalized frequency deviation y(t) is the derivative of the time error x(t), its PSD is given by :

$$S_y(f) = 4\pi^2 f^2 S_x(f)$$

Then :

$$S_y(f) = \sum_{\alpha=-3}^{+2} h_\alpha f^\alpha$$

0-7803-2500-1/95 $4.00 © 1995 IEEE

Estimating the PSD of y(t) consists then in measuring the noise levels h_α.

The most obvious way to find the PSD is by taking the square of the Fourier transform (or the numerical Fast Fourier Transform) of the time series. Although straightforward, this is not always the best approach nor is it always a possibility. Variances are calculated using r^{th} order difference operators, one example being the two sample variance or Allan variance:

$$\sigma_y^2(\tau) = \frac{1}{2} < [y(t+\tau) - y(t)]^2 >$$

Variances in general have a smoother response than the Fourier transform. However, a single variance is not well suited to measure all possible power laws. The multi-variance method was developed to better deal with a wide range of power law noise spectra [6]. This method uses different variances and gives the most probable set of power law noise levels (most probable in the sense of the least-squares). One severe drawback for the study of pulsar timing data is the fact that none of these methods can be used on non-equispaced data.

In order to see a solution to this limitation let us formulate a general method of analysis. Let x(t) represent the time series or noise process and $\{h_1(t), ..., h_j(t), ..., h_n(t)\}$ be a set of n sampling functions. We can now define :

$$P(f_j) = \left[\int_{-\infty}^{+\infty} h_j(t)\, x(t)\, dt \right]^2$$

where $P(f_j)$ is an estimator of $S_x(f)$, the PSD of x(t). This continuous function formula can easily be adapted to the case of discrete data. We have m discrete data obtained at times $\{t_1, ..., t_k, ..., t_m\}$

$$h_j(t) = \sum_{k=1}^{m} h_j(t_k)\, d(t-t_k)$$

denoting $h_{jk} = h_j(t_k)$, the estimation is given by :

$$P(f_j) = \left[\sum_{k=1}^{m} h_{jk}\, x_k \right]^2$$

The question is what is the relation between $P(f_j)$ and $S_x(f_j)$ for a given frequency f_j ? It can be shown that the expectation of $P(f_j)$ for a large number of processes may be written as :

$$<P(f_j)> = \int_{-\infty}^{+\infty} |H_j(f)|^2\, S_x(f)\, df$$

where $H_j(f)$ is the Fourier transform of the sampling function $h_j(t)$. If for example we choose $e^{i2\pi f_j t}$ as $h_j(t)$, we see that $H_j(f)$ is a delta function and $P(f_j)$ is identified with $S_x(f)$ at each $f = f_j$ just as we expect. For a variance calculation $h_j(t)$ is an r^{th} order difference function, and $S_x(f_j)$ is related to $P(f_j)$ via the transfer function $H_j(f)$. We may choose anything we like for $h_j(t)$, and as we shall see, a polynomial can give useful results. Regardless of our choice for $h_j(t)$, P(f) is a filtered form of $S_x(f)$, related by the transfer function $|H_j(f)|^2$.

We can now state that the sampling function $h_j(t)$ gives an unbiased estimate of $S_x(f)$ if $H_j(f)$ is centered on some nominal frequency f' and has a narrow bandwidth. For a finite length data set, $H_j(f)$ will not be a single peak but will have sidebands. A sinusoidal $h_j(t)$ will no longer give a delta function and instead yields a sinc function. This does not pose too great a problem if $S_x(f)$ is flat, i. e. a white noise spectrum. But if $S_x(f)$ rises too steeply for low frequencies, i. e. a red noise spectrum, the sidebands in fact dominate the product of $|H_j(f)|^2.S_x(f)$ and the integral diverges.

2.2. The moment condition

To ensure the convergence of the estimator the moment condition is imposed. We want $H_j(f)$ to satisfy

$$\int_{-\infty}^{+\infty} |H_j(f)|^2\, f^{-\alpha}\, df \qquad \text{finite}$$

where $f^{-\alpha}$ is the red noise part of the power spectrum. If $h_j(t)$ satisfies the following moment condition :

$$\int_{-\infty}^{+\infty} h_j(t)\, t^q\, dt = 0 \qquad \text{for } q < n$$

it can be shown that the estimator will converge for $\alpha < 2n+1$. Thus for f^{-5} noise (f^{-3} frequency noise) the moment condition must be verified for n = 2. For discrete data the moment condition is

$$\sum_{k=1}^{N} h_{jk}\, t_k^{\,q} = 0 \quad \text{for} \quad q \le 2$$

Moreover, this condition implies $h_j(t)$ is non-sensitive to quadratic phase drift, i. e. linear frequency drift. In fact we shall see that it is equivalent to having removed a linear and quadratic term from the time series.

2.3. The orthogonality conditions

The choice of an estimator verifying the moment condition ensures that it will be unbiased on a red noise, since it will reject the (unavoidable) drifts. If we want it to be insensitive to the measurement white noise, and be mostly sensitive to the suspected red one, we are led to the orthogonality condition and to the lowest mode estimator [2, 3].

A set $\{h_1(t), ..., h_N(t)\}$ of estimators applied over a measurement noise (assumed as white phase noise) has a covariance matrix, of which the elements may be defined as :

$$\langle h_i|h_j\rangle = \sum_{k=1}^{m} h_{ik}\, h_{jk}\, \sigma_k^2$$

where σ_k^2 is the variance of the observational noise at the time t_k and m the number of time measurements. This is a kind of scalar product. Uncorrelated estimators must have a diagonal covariance matrix and are consequently "orthogonal".

On a red noise of known spectrum $S_x(f) = k_{-2r}\, f^{-2r}$, the estimation is given by :

$$P(f_j) = \int_{-\infty}^{+\infty} |H_j(f)|^2\, k_{-2r}\, f^{-2r}\, df \qquad (1)$$

Moreover, denoting $h_j^{(-r)}(t)$, the r^{th} integral of $h_j(t)$, its Fourier transform is given by :

$$h_j(t) = \frac{d^r}{dt^r} h_j^{(-r)}(t) \Leftrightarrow H_j^{(-r)}(f) = \frac{H_j(f)}{f^r} \qquad (2)$$

Then, from (1) and (2), it becomes :

$$P(f_j) = k_{-2r} \int_{-\infty}^{+\infty} |H_j^{(-r)}(f)|^2\, df$$

Thus, the covariances of the estimations for a f^{-2r} noise defines a new r-scalar product :

$\langle h_i^{(-r)}|h_j^{(-r)}\rangle$

The independance of the estimations for a f^{-2r} noise leads us to a new diagonal r-covariance matrix.

The choice of sets of estimators, orthogonal on both measurement noise and red noise has the advantage that we can estimate the variance of any sum of them, as the sum of individual variances (assuming that uncorrelation implies independence, even if opposite examples are well known for non-gaussian distributions).

Deeter and Boynton suggest a basis of m orthogonal estimators for measurement noise $h'_i(t)$, using m orthonormal polynomials $p_i(t)$ of ascending degrees i, defined iteratively by :

$$\sum_{k=1}^{m} p_i(t_k)\, p_j(t_k)\, \sigma_k^2 = \delta_{ik}$$

and

$$h'_i(t) = \sum_{k=1}^{m} p_i(t)\, \sigma_k^2\, \delta(t - t_k)$$

$\{h'_1(t), ..., h'_m(t)\}$ are orthogonal on measurement noise even for irregularly sampled data. It may be demonstrated that each polynomial $p_i(t)$ verifies the moment condition up to an order equal to i, the degree of this polynomial. Thus, removing the first n+1 polynomials ensures that they respect the moment condition up to order n.

From $\{h'_{n+1}(t), ..., h'_m(t)\}$ we derive an r-orthogonal set of estimators by rotating the r-covariance matrix to its principal axes, obtaining $\{h_{n+1}(t), ..., h_m(t)\}$:

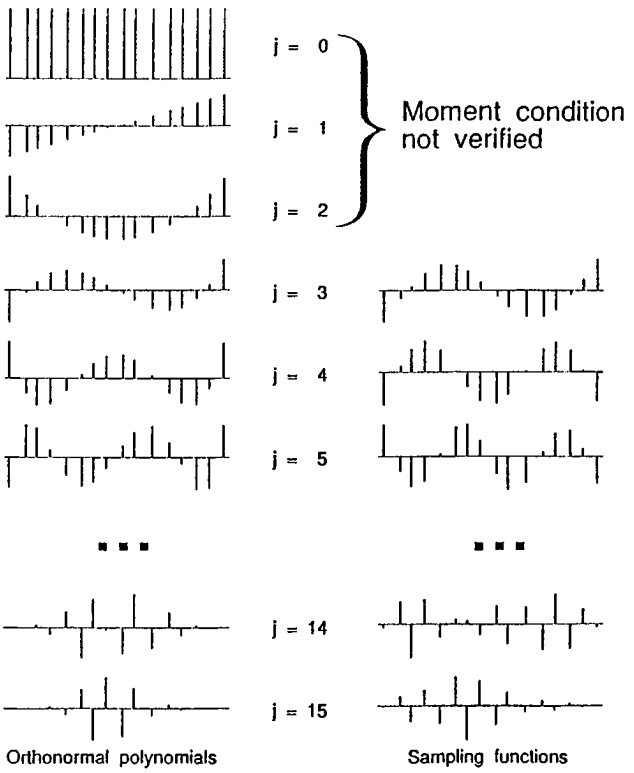


Figure 1 : Orthonormal polynomials $p_j(t)$ and the corresponding sampling functions $h_j(t)$ for a set of 16 irregularly spaced data.

$$h_j(t) = \lambda_j^{-1/2} \sum_i u_{ji}\, h'_i(t)$$

u_{ji} the eigen-vectors components

with λ_j the eigen values of the r-covariance matrix

Thus :

$\langle h_i^{(-r)}|h_j^{(-r)}\rangle = \delta_{ij}$

Figure 1 shows an example of estimators calculated for m = 16, a moment condition verified up to order n = 2, and an optimization for r = 2 (f^{-4} phase noise).

The "signal to noise ratio" sn of one of those doubly orthogonal estimators may be estimated as the average over an infinite set of experiments :

$$sn = \frac{\langle h_i^{(-r)}|h_i^{(-r)}\rangle}{\langle h_i|h_i\rangle}$$

The lowest mode estimator is the one which minimizes it.

2.4. Practical calculation of the lowest mode estimator

We want to study f^{-3} and f^{-4} (if it exists) frequency noise in pulsar timing, and we assume that no "redder" noise is present. We have 165 residual times of the monitoring of PSR 1937+21, obtained at Arecibo Observatory (courtesy Joe Taylor). Our estimators must then verify the moment

condition up to order 2 in order to converge up to f^{-4} frequency noise (f^{-6} phase noise) :

$$\sum_{k=1}^{165} h_{ik}\, t_k^q = 0 \quad \text{for} \quad 0 \le q \le 2$$

On the other hand, the a priori knowledge of the spectrum of the residuals leads us to optimize the estimator for f^{-2} frequency noise by diagonalizing the 2nd order covariance matrix.

The computation of all 165 $p_i(t_k)$ requires a large number of significant digits, but once it is done we easily get the 162 doubly orthogonal estimators $\{h_0, ..., h_{161}\}$, and we normalize them so that $< h_i^{(-r)} | h_j^{(-r)} > = \delta_{ij}$.

We want to outline as a conclusion of this paragraph that the synthesis of estimators optimized for a given noise requires some previous knowledge of the spectrum. The moment condition will not be fulfilled if the signal contains a noise redder than suspected. And the higher we choose r_{max}, the less sensitive we will be to low frequencies. Covariance matrices of given sets of estimators also depend on the expected noise we took into account. Independence on a f^{-2r} noise does not imply independence on a $f^{-2r'}$ noise for $r' \neq r$.

2.5. Estimator bandwidth

Table I gives the relationship between the PSD of x(t) and the PSD of z(t). Thus, the estimation process may be rewritten as :

$$P(f_j) = \int_{-\infty}^{+\infty} |H_j(f)|^2 \, S_x(f)\, df = \int_{-\infty}^{+\infty} |H_j(f)|^2 \, \frac{S_z(f)}{16\pi^4 f^4}\, df$$

Moreover, since

$$h_j(t) = \frac{d^2}{dt^2} h_j^{(-2)}(t)$$

where $h_j^{(-2)}(t)$ is the second integral of $h_j(t)$, the Fourier transform of these functions, $H_j(f)$ and $H_j^{(-2)}(f)$, are linked by :

$$\frac{H_j(f)}{4\pi^2 f^2} = H_j^{(-2)}(f)$$

Hence :

$$P(f_j) = \int_{-\infty}^{+\infty} |H_j^{(-2)}(f)|^2 \, S_z(f)\, df \approx S_z(f_j)$$

Thus, an estimation of the PSD of z(t) is performed for the frequency f_j. This is obvious if $S_z(f)$ is constant (i.e. f^{-2} frequency noise, f^{-4} phase noise) because the second integral of the sampling functions are normalized.

This is right also for the other types of noises if the transfer functions of the sampling functions have equal relative bandwidths. Unfortunately, this is not true.

Time domain	x(t)	$y(t) = \frac{dx}{dt}$	$z(t) = \frac{dy}{dt} = \frac{d^2x}{dt^2}$
Spectral density	$S_x(f)$	$S_y(f) = 4\pi^2 f\, S_x(f)$	$S_z(f) = 4\pi^2 f^2\, S_y(f)$ $S_z(f) = 16\pi^4 f^4\, S_x(f)$

Table I : Relations between x(t), y(t), z(t) and their PSD

However, in order to obtain equal relative bandwidths, the estimations given by different sampling functions may be averaged. Each average is weighted according to the frequency f' at which we want to measure the PSD : the closer to f' the maximum f_j of the transfer function $H_j(f)$ is, the higher the weight of this estimation is. This means that it is possible to perform an estimation of the PSD of z(t) for any frequency.

3. RESULTS

3.1. Simulation of pulsar timing sequence

In order to check the efficiency of this method, we generated sequences of simulated pure noises, from white phase noise to f^{-5} phase noise (f^{+2} to f^{-3} frequency noises). The knowledge of the noise level, entered for the simulation, allows us to compare it with the level given by the estimation.

The first step is the computation of a sequence of 8192 data of the pure noise selected with the entered noise level. These data are considered as x(t) data, regularly sampled, with a sampling step of 1 Arbitrary Time Unit (ATU).

The second step consists in removing data of this sequence, in order to get the same time distribution as the timing sequence of the pulsar PSR 1937+21, recorded in Arecibo from 1984 to 1992. Only 165 data are kept, from t_0 = 0 ATU to t_{164} = 8191 ATU. The smallest time interval between two consecutive data is about 2 ATU, whereas the largest gap between two consecutive data is about 160 ATU.

This method was used to obtain 25 independent noise processes for each type of noise.

3.2. Sampling functions and transfer functions of the lowest mode estimator

Performing the method described in section 2 with a moment condition set to 3 and estimators optimized for f^{-4} phase noise, yields 162 sampling functions (from $h_0(t)$ to $h_{161}(t)$).

Figure 2 shows the first and the 101st sampling functions. Of course, these functions are only defined for the 165 dates $\{t_0, ..., t_{164}\}$ of the noise sequences. The first sampling function is useful for the lowest frequency

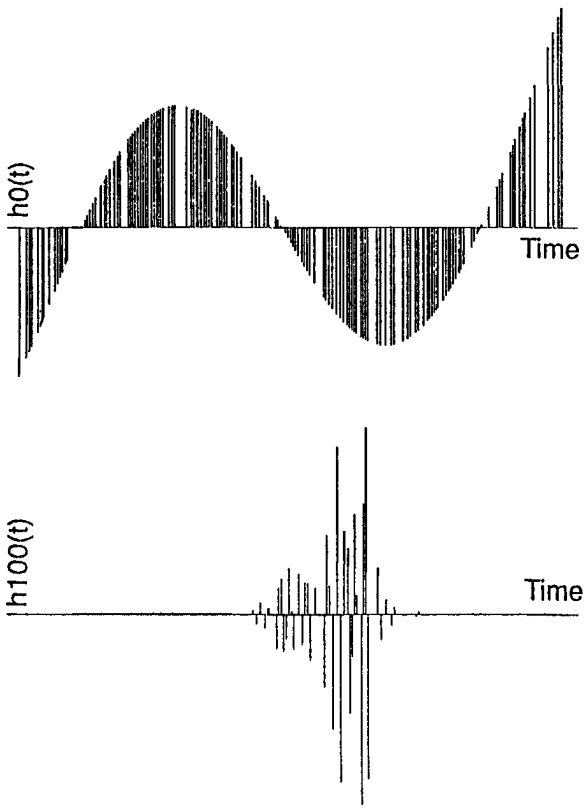


Figure 2 : Sampling functions $h_0(t)$ and $h_{100}(t)$.

analysis (long period undulations) and the other one for higher frequency analysis (short period undulations).
This is confirmed by their transfer functions shown on figure 3 : the maximum of $H_0(f)$ is centered at about 2.10^{-4} Arbitrary Frequency Unit ($\frac{1}{5000\ ATU}$) whereas the other one is centered at about 5.10^{-3} AFU ($\frac{1}{150\ ATU}$).

3.3. Responses for the different types of noises

The PSD of a noise process remains a random process : only the expectation of this PSD may be modeled as a deterministic function (i. e. power laws). Thus, in order to obtain the response of the lowest mode estimator for one type of noise, it is necessary to study several independent noise processes generated with the same noise levels.
For each type of noise, we used 25 independent noise processes. We will call "*response of the lowest mode estimator for a f^{α} noise*", the average of the estimations performed over 25 independent processes generated with a noise level h_{α} equal to 1 and the other noise levels set to 0.
Figure 4 shows the response of the lowest mode estimator for f^{+2} to f^{-3} frequency noises. The frequency range of this estimation stretch from 2.10^{-4} AFU to 5.10^{-3} AFU with 5 measurements per octave. The measurements

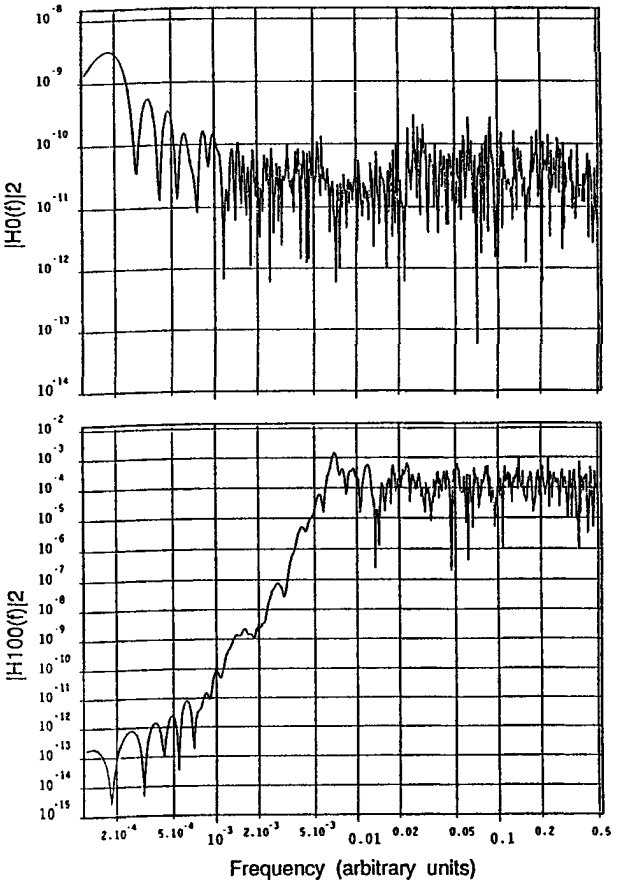


Figure 3 : Square of the modulus of the transfer functions associated with $h_0(t)$ and $h_{100}(t)$.

obtained for frequency higher than 5.10^{-3} AFU seems to be meaningless.
As we expected, this method estimates the PSD of z(t) since the slopes of the responses vary from f^{+4} to f^{-1} while the types of noise vary from f^{+2} to f^{-3} frequency noise. Moreover, this method is much more sensitive to very low frequency noises than high frequency noises.

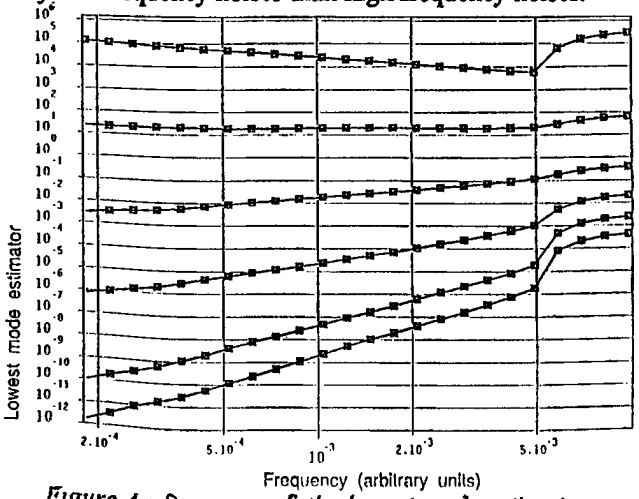


Figure 4 : Response of the lowest mode estimator for f^{+2} frequency noise (below, f^{+4} slope) to f^{-3} frequency noise (above, f^{-1} slope).

3.4. Results for pure noises

Figure 5 presents results obtained for sequences composed of pure noise. Each curve exhibits the estimation of the PSD of z(t) for only one process. Therefore, these curves are not so straight as the curves of responses (figure 4).
However, from these measurements, we estimated the noise level with an accuracy of about 10 % compared with the noise level entered for the simulation.

3.5. Results for mixed noises

The millisecond pulsar should be affected by both white phase noise and a very low frequency noise (e. g. f^{-3} frequency noise). Thus, it is of importance to estimate the accuracy of this method for mixed noises.
Figure 6 shows the results obtained for two sequences composed of two noises : the first one is composed of f^{-3} frequency noise and white frequency noise, and the second one of f^{-3} frequency noise and white phase noise, as expected in pulsar timings.
In these two cases, the measurement of the 2 noise levels was performed with an accuracy of 25 %.

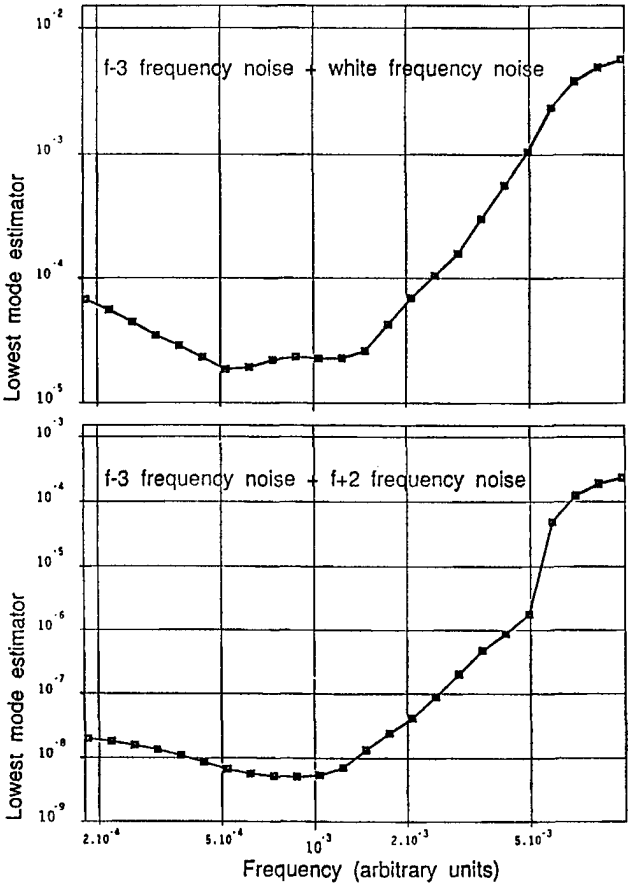


Figure 6 : Results for processes composed of 2 mixed noises.

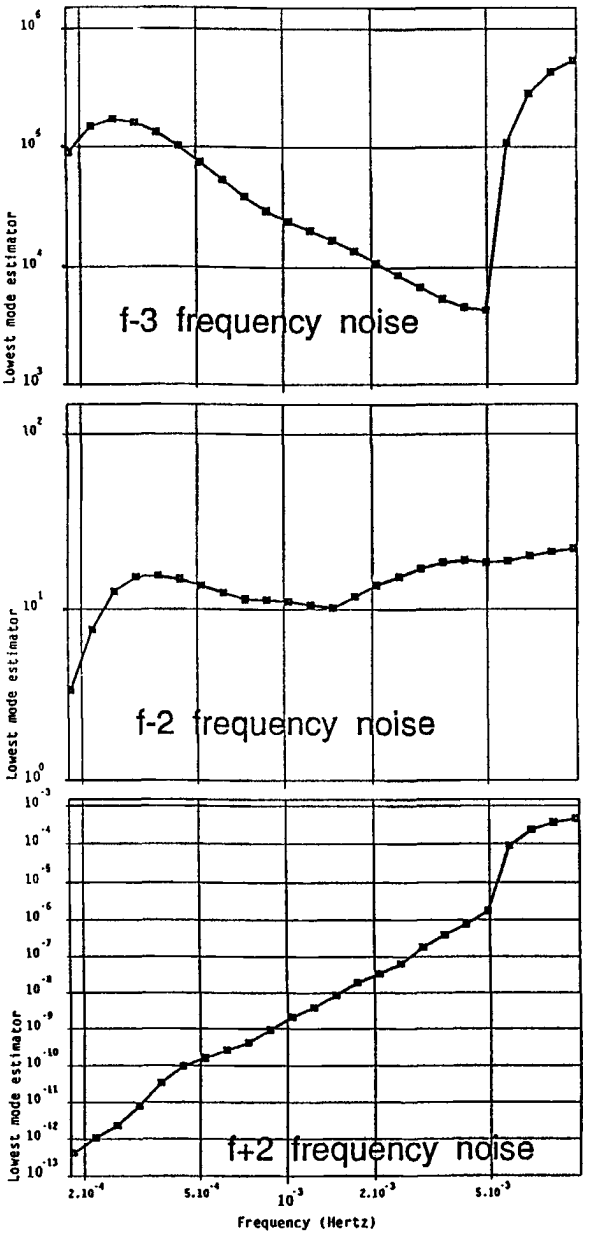


Figure 5 : Results for processes composed of 1 pure noise.

4. CONCLUSION

With this set of 165 data, the lowest mode estimator yields a frequency range of about 5 octaves, with a frequency resolution of about ½ octave (the graphs were made with finer resolution redundant points for easier plotting).
These values may be compared to the ideal case : equispaced data with a sampling period equal to the smallest time interval between two consecutive data (i. e. 4096 data with a sampling period of 2 ATU). In this case, the frequency range should be of 11 octaves.
The leakage induced by the gaps between data concerns only the highest frequencies. However, this problem is

not a severe drawback for millisecond pulsar study since we are mostly interested in the long term behavior.
Concerning the prospects of this work, we will improve this method by using it in the multi-variance framework. Thus, we will define several lowest mode estimators and use them over the same signal.
First, these estimators may be optimized for different types of noises (the exponent must be even) :

- for f^{+2} frequency noise
- for white frequency noise
- for f^{-2} frequency noise (as we did)
- for f^{-4} frequency noise.

Secondly, these estimators may be defined with different moment conditions :

- to first order, this yield an estimator sensitive to linear frequency drift but which does not converge for very low frequency noises (as the modified Allan variance)
- to second order (as Picinbono variance).

Thus, it is possible to build up to 8 different estimators. Using all these estimators in a multi-estimator system yields the advantage of the multi-variance method [6]:

- a wider frequency range
- a better accuracy for the noise level measurement.
- an estimation of the uncertainties about each noise level measurement.

5. REFERENCES

[1] W. C. Lindsey, C. M. Chie, "Theory of oscillator instability based upon structure functions", *Proceedings of the IEEE*, vol. 64, December 1976, pp. 1652-1666.

[2] J. E. Deeter, P. E. Boynton, "Techniques for the estimation of red power spectra. I. Context and methodology", *The Astrophysical Journal*, vol. 261, October 1982, pp. 337-350.

[3] J. E. Deeter, "Techniques for the estimation of red power spectra. II. Evaluation of alternative methods", *The Astrophysical Journal*, vol. 281, June 1984, pp. 482-491.

[4] E. J. Groth, *Astrophysical Journal Supplement*, vol. 29, 1975, pp. 443.

[5] F. Vernotte, E. Lantz, M. McHugh, G. Zalamansky, "Cut-off frequencies and noise power law model of spectral density : adaptation of the multi-variance method using the structure function approach", to be published in *Proc. of 9th EFTF*, Besançon, France, 1995.

[6] F. Vernotte, E. Lantz, J. Groslambert, J. J. Gagnepain, "Oscillator noise analysis : multi-variance measurement", *IEEE Trans. on Instrum. and Meas.*, vol. IM-42, no. 2, April 1993, pp. 342-350.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

CIRCULAR REPRESENTATION OF INFINITELY EXTENDED SEQUENCES

D. A. Howe
Time & Frequency Division National Institute of Standards & Technology
Boulder, CO 80303

ABSTRACT All data in a space-ordered or time-ordered series are always observed for a finite distance or time. A sample variance of the data represents a formalized means of capturing the extent of observed variations over a finite interval. Different types of variances abound and for the existence of a "true" particular variance, assumptions are made regarding ergodicity, stationarity and statistical independence of the random variables. This writing is not about variance *per se* but rather about the fact that proper estimations of variance come from understanding the implications of these basic assumptions. In particular, there is usually an underlying assumption of ergodicity. Ergodicity means that we treat one statistical average as an ensemble of smaller statistical averages. If we use the ergodic assumption within a data set, then we acknowledge its generalization as an included assumption for data outside the set, namely before and after, in the case of a time series. Ergodicity implies that any series will have a likelihood of recurrence which inversely depends on the number of independent observations. If only one series is ever observable, absolute recurrence of that series is the consequence under the implied assumption of ergodicity and an assumption that a "true" variance indeed exists. This yields the model that time-ordered data may be treated as circular or wrapped.

Mathematical formalisms such as a sample variance (or any estimator) are based on assumptions involving what might happen if given infinite time. These formalisms can easily be misinterpreted and even misunderstood. Therefore, I clarify the implications of specific infinite-time variance-related concepts in what I hope is a compact, speculative manner and style.

INTRODUCTION

The circular representation of two-dimensional measurements means that one of the coordinates is wrapped so that the endpoints of the coordinate match. This paper focusses on the notable example of measurements with respect to time. If the measurements derive from physical systems described as having periodicity, circularity, or any of the set of wave-like observations, then we assume that periodicity underlies a time-series measurement. Fourier analysis proceeds from this assumption. Building a model of the observation is then driven by a search for the nature of this periodicity which was assumed. Furthermore we might go further and infinitely extend and circularly represent the finite time series so that the measurement is recurrent for all time again because periodicity is an assumed property of the data set. This is precisely what is done for most frequency-domain representations of a time series. There is a vast literature on Fourier analysis techniques. See for example [Percival and Walden, 1992] for a comprehensive review.

The assumption of periodicity as underlying a time series is often physically correct. For example using Fourier analysis to model the motion of an orbiting satellite, the motion of underwater ocean currents, or the current in a resonant electronic circuit seems plausible. But here is a major problem. What if we are not given a basis for assuming that periodicity is intrinsic to the data? What if we are not at liberty to assume any known deterministic cause whatsoever for that matter? Then we resort to a model that can be used to calculate a probability of a future value lying between two calculable limits which is based on past values. It is formally called a stochastic model [Box and Jenkins, 1970]. In other words, in analyzing a time series as non-deterministic functions, we regard it as a realization of a stochastic process, one supposedly dependent on chance within formalized limits given by past information. I claim however that the interpretation of a stochastic process inadvertently includes a fundamental assumption, namely ergodicity. This paper discusses the

U.S. Government work not protected by U.S. copyright

appropriateness of an assumption of ergodicity and shows that we can infinitely extend and circularly represent finite periodic processes and, in addition, finite random or stochastic processes under such an assumption.

BASIC CONCEPTS

Science is the trained observation and interpretation of a part of something based on the fundamental principle that everything is contained in or generalized from any part of itself. See for example [Poincaré, 1912]. This assumes that order and consistency are a part of all that we observe. Basically, scientists try to predict events based on a model derived from limited observation. The purpose of models is prediction. For example, every measurement of the phase difference x(t) between two clocks, or its derived frequency difference y(t) has a beginning at, say, t_0 and an ending at, say, t_M. Furthermore, the values of x(t) or y(t) are always sampled, preferably equispaced, in intervals of, say, Δt. A useful figure of merit is frequency stability, which is an estimate of the variance of y(t) from this limited view of the data. I say "estimate" in the sense that the true variance requires that we somehow obtain all the data for all time, a physical impossibility requiring infinite time and infinitesimal Δt spacing. If we remove systematic trends so that to the best of our available knowledge any value of x(t) or y(t) is unrelated to any other value, then the residuals of x(t) or y(t) are said to be processes that are random, and their variance remains more or less constant and independent of when and for how long we have made observation. There is, therefore, some expectation that the degree of deviation within x(t) or y(t) has fixed limits. The methodology is depicted in figure 1. I will not pursue the issue that the assumption of continuous randomness is, however, idealistic and that "expecting randomness" is a contradiction of itself.

Practically speaking, all that we can do is obtain a data set from t_0 to t_M sampled at Δt intervals. We say that we ignore everything else, but in fact we must be definite about what we mean by and the expectations of the "ignored" data. Granted, the goal of statistical analysis in many cases is to model at least two elements, the deterministic and stochastic processes, so that we can somehow know within finite bounds the "unmeasurable" part of the data (outside our measurement time) without actually measuring it. This is what we do with an accurate model. But besides ignoring the turn-on/turn-off effects at t_0, t_M, and Δt, ignoring the unmeasured or unmeasurable data strictly means in the usual sense that we are assuming that it is noiseless with zero mean as shown in figure 2. With surprising regularity, traditional statistical treatments do not address properly the assumptions of (1) noiseless zero mean up to t_0, (2) the effect of turn-on at t_0, (3) sampling at Δt, (4) the effect of turn-off at t_M, and (5) the effect of noiseless zero-mean after t_M.

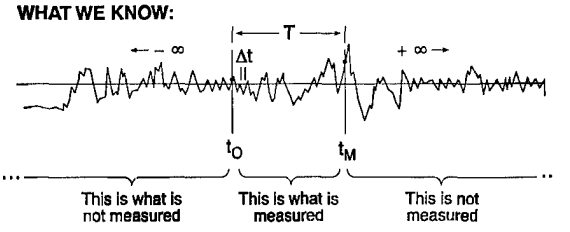


Figure 1

In every sense, "estimating" the variance is akin to estimating what the rest of the unmeasurable data are. So why assume from the outset that the rest of the data are noiseless with zero mean? Obviously this is not a good model. As I will point out, we never rightfully use this model, but worse yet, we erroneously might think we are using it.

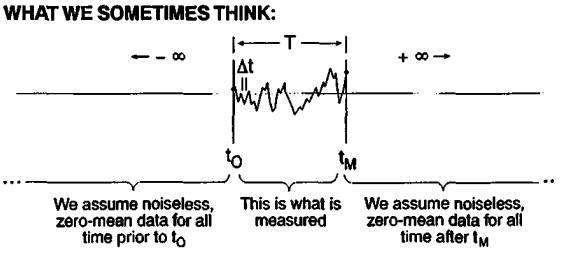


Figure 2

ERGODICITY

We assume that the outcome of flipping a coin ten times is equivalent to flipping ten coins all at once. The property is called "ergodicity." When dealing with a population, this is a reasonable starting point. Applied to a "time series," a stationary random process is defined as "ergodic" if all types of ensemble averages are interchangeable with the corresponding time averages. Thus in "ergodic" processes, the averages across an ensemble are equal to the averages over time of a single function of

infinite extent.[1] For example, given the single function y(t), with zero mean, we have

$$\sigma^2 = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} y^2(t)\, dt, \qquad (1)$$

where σ^2 is a type of average called the standard variance and the form of eq. (1) yields the variance spectrum. By "spectrum" we mean that y(t) is ordered or arrayed by an independent varying component, in this case "T" which is the time interval from some origin. Its reciprocal (frequency) can be used as the basis for assigning coefficients of a Fourier series to determine the spectral components in y(t) [Bingham, Godfrey, Tukey, 1967].

Let $y(t \mid n)$ represent the value of the nth sample function drawn from the population y at time t. If y is ergodic, then we have (by definition):

$$\sigma^2 = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} [y(t \mid n)]^2\, dt \qquad (2)$$

for any specific time series n in the population y. This must be equal to the variance at one specific instant t across the ensemble of sample functions (that is, all values of y at one instant t). This variance can be written as

$$\sigma^2 = \lim_{M\to\infty} \frac{1}{M} \sum_{n=1}^{M} [y(t \mid n)]^2 = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} y^2(t)\, dt. \qquad (3)$$

eq (1) and eq (2) say that the true value of the standard variance of the entire data is equivalent to the average of variances of finite length = T as shown in figure 3. One activity observed forever has an equivalence to a finite observation as shown in figure 4.

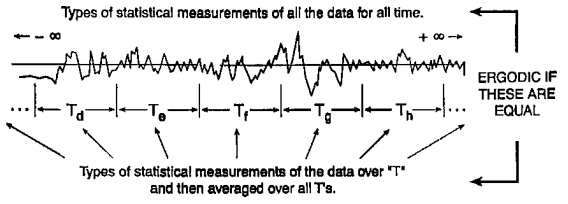


Figure 3

The equality of eq (1) and eq (2) is the ergodic hypothesis when expressed as eq (3). By "equality," I mean that within a statistical uncertainty given by say, a standard deviation over many samples which test the hypothesis, I cannot tell that the expressions are not equal in the limit given.

We say the finite observations are statistically independent if the measurement times are different, that is, if no two measurements over T overlap. Then the time average over one infinite record is equivalent to an ensemble average of finite records of length T. Of course, the concept of a true variance is an idealization since it implies that we have collected data infinitely, a physical impossibility. Interestingly, ergodic theory originated from the opposite view that an ensemble of infinite members can be viewed as one member for all time as shown in figure 4.

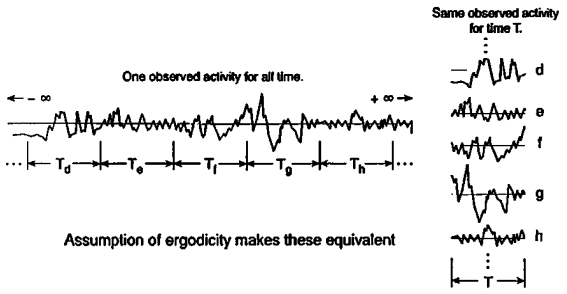


Figure 4

Ergodic theory originated in the mid-1920's from experiments in classical statistical mechanics. At that time, surmising the statistical properties of a single particle was virtually impossible, and experiments involved a large number of particles or members of an ensemble. It was verifiable by experiment that in most cases and under proper conditions, ensemble statistics were approximately equal to time-averaged statistics for a single particle or member. Mathematicians attempted to prove this for arbitrary data rather than for observations in statistical mechanics. Ergodic theory was the subject of widespread mathematical interest for a fairly brief period of time from the 1930's to 1950's. Ergodicity was never proved to be true in all situations but was assumed in many situations. Physicists accepted it based on seemingly sufficient mathematical tractability, and mathematicians accepted it because of experimental results. Judging from events, ergodicity (eq (3)) was eagerly accepted as a way to replace phase statistics by time statistics. By the 1960's, ergodic theory was an accepted model and was not pursued seriously thereafter, and the

remaining unproven aspects were left more or less hanging [Halmos, 1956].

The conditions for ergodicity are related to the uniformity in the data and measurement procedures. The only test for ergodicity is that any given statistical average is uniformly the same over any interval throughout a given interval. There are many decidedly nonergodic situations. For example, we cannot combine the time readings of an ensemble of different clocks and view them as one clock unless the readings have identical statistical averages. In any application nonergodicity of one time-series just means that the random numbers concerned are, in fact, an artificial union of a number of distinct and otherwise (ergodic) stationary sequences [Yaglom, 1987]. An important consideration in frequency metrology is whether the power-law of the fractional RMS frequency stability changes with scale (or averaging time τ). If so, the series is clearly non-ergodic.

In practice, we cannot realize the exact limit of eq (3) and must resort to a finite version. This is a source of uncertainty, and, by all rights, eq (3) should be an approximation given as

$$\frac{1}{M}\sum_{n=1}^{M}[y(t|n)]^2 \simeq \frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} y^2(t)\,dt \ . \qquad (4)$$

But then questions are raised regarding the degree and conditions of the approximation. Experimentalists working under deadlines eventually say that close enough is good enough. But the assumption of equality in eq (4) has dramatic implications when applied to any data. And in problems involving time-series data, ergodicity has been regularly assumed, perhaps properly or perhaps not. For one important statistical measure, the variance, the assumption of ergodicity applied to any finite observation of ordered deviates is central to many other properties based on the variance.

Our observed time series shown in figure 2 is an ensemble consisting of one measurable member, namely from t_0 to t_M. We might think all other members are zero-valued. Any infinite time average of variance of such an ensemble (represented as either eq (1) or eq (2) will naturally approach 0 because we are dividing a finite sample by ∞. But in the case here, I am assuming an infinite sample, which in the limit is supposed to converge to a nonzero value, its so-called true value. Hence our assumption that all other members are zero must not be correct.

What assumption regarding the unobservable events is or should be made when computing any type of nonzero average for a finite record over T (since the infinite average cannot be zero in every case)? The possibilities are that the unobservable events (those outside our observation window) are either different (possibly zero) or the same. Zero as shown is not possible based on our assumption that a nonzero variance must still exist when we extend our observation period to infinite time. If the unobservable events are arbitrarily different, then I argue that there is less and less we can say about the events as we imagine the observation period to be lengthened. And over infinite time, we know nothing. Saying with great mathematical abstraction the degree to which these outside events "behave like" or "are similar to" the observed events is a trap of sorts in which we are denying the obvious; that is, we unknowingly project some expectation of "sameness" onto the unobserved events. At a fundamental level, we are predisposed to thinking that a future possibility (our imagination) is similar to or consistent with the record of the past (our memory). But imagination which is dependent on memory in any way is misleading. That which is imagined becomes nothing more than a goal or expectation.

In short, if the unobservable (future) events are different, there is nothing we can say about them if the events derive from complicated systems. This is the basis of randomness. The assumption then is that we commonly (and inadvertently) assume that the measurements are identical to the one we observe no matter when we observe (a condition of stationarity). This is shown in figure 5 as the same observed measurements from t_0 to t_M and is repeated for all time with length T prior to t_0 and after t_M. This assumption satisfies the conclusion that any finite, nonzero average is exactly its equivalent infinite average. Of course, any number of different processes can yield an identical, specific average. But it is not our privilege to simply create ad hoc any random process yielding same specific averages. Hence, we conclude that the observation is everywhere recurrent under an ergodic hypothesis given by eq (3) and shown in figure 6.

Without provocation, we accept that conclusions are derived and built from assumptions. I must point out, though, that this is not what I am doing here.

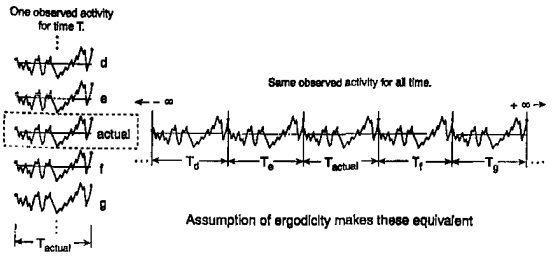


Figure 5

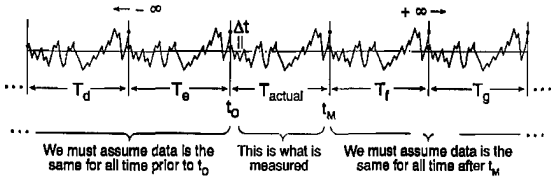


Figure 6

Here I am simply deriving a different view of an assumption from an existing assumption. That the data in an observation window repeat themselves outside the observation window is an assumption derived from the common assumption of ergodicity. Furthermore, one assumption is not so much derived from the other as both are the same assumption. Implicit throughout subjects on data analysis is the pervasive notion of extended periodicity, yet admitting this as applied directly to the observed data seems somehow untenable. Frankly, the notion is very tenable and leads to the practical consequence that we can represent the data as circular as shown in figure 7.

A finite observation period T is always a high-pass filter that is insensitive to any change slower than T. A normal starting point is that the ergodic hypothesis requires stationarity of data. Determining convergence, divergence, or stationarity of data requires that we are at liberty to change T and see the resultant change, if any, in a statistical average. As profound as it seems, however, stationarity, is a condition of our chosen models and not actual data [Barnes, et al., 1971]. The assumption of stationarity is satisfied here by the fixed finite period of observation. The circular representation derives wholly from the ergodic assumption, and stationarity is maintained. Therefore, in the end, it does not matter if we decide that the actual data are stationary or not. The circular model is stationary.

As is always true, it is up to the reader to assign meaning to a representation or model. At least three

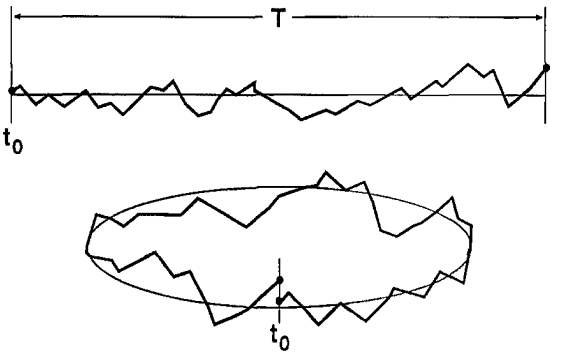


Figure 7

aspects are up for consideration regarding the circular model: (1) Is the model rich enough to include other accepted similar models? (2) Is the model plausible and satisfying in some sense? (3) Is it practical, and how can it work for us? I would like to add a fourth consideration, simplicity. It behooves us to pick simple models.

OTHER MODELS

That we can rightfully assume that a finite time series of random variables recur to an arbitrarily long extent are corroborated by several related, notable examples:

1. Mathematical reasoning proceeds from the particular to the general as inductive reasoning. In the asymptotic limit of inductive reasoning, we proceed from the finite to the infinite and must revert to "reasoning by recurrence"; that is, any finite succession can only be deemed as recurring to an infinite extent [Poincaré, 1912].

2. Any spatial or temporal measures in which the endpoints can be matched without discontinuity are said to be "circular." All analysis can proceed from an assumption of circularity. Endpoint mismatches affect statistical averages only by a scaling factor. See [Bloomfield, 1976] and the discussion on smooth functions.

3. A finite, bounded standing wave can be viewed as counter-propagating travelling waves of arbitrarily long extent. The extended distance to any "virtual" boundary is always a rational multiple of the actual

boundary. See, for example, [Churchill, 1963].

4. For a finite observation period, the method of complex demodulation states that data can be viewed as having in-phase and phase-quadrature components. We must extend the time series by at least half its length to completely account for the phase-quadrature component [Bloomfield, 1976].

5. In computing the power spectrum, commonly used digital processing instruments treat a finite time series as a "block" which is periodic and infinitely extended because all variances still properly converge if given arbitrarily more independent blocks [Welch, 1967].

6. Formulations involving the Fourier transform assume that both discrete continuous functions $\{y_k\}$ and their discrete Fourier transforms $\{Y_k\}$ are circular (periodic) in the sense that evaluations of the functions outside the range $k=1, 2, 3,\ldots,M$ will yield results modulo M [Otnes and Enochson, 1978].

PLAUSIBILITY

At first glance, it appears that we cannot use this model of recurrence because our experience is inconsistent with a hypothesis that activity recurs; that is, if we wait, the same thing happens. So an experience-based model suggests that something different can (in fact, usually does) happen. But it is known in physics and mathematics that acting exclusively and exactly on the appearance of our experience leads to inconsistency in the extreme. For example, pure deductive logic can lead to contradictory results from one set of assumptions and one set of rules [Gödel, 1931]. In fact, it is the principle of causality which is under consideration. Any measurement is not the effect of what precedes it but rather is the effect of a set of conditions. It has been argued that what we see as activity actually remains fixed, and it is only a set of conditions, principally the observer's viewpoint, which changes randomly. Thus he incorrectly perceives that the observation itself is changing. It follows that an inconsistency within a model is not an inconsistency at all but rather perhaps only a different view of the same model. Recurrent or nonrecurrent behavior is within the observer's subjective perception [Hofstadter, 1979]. All the same, specifically, no one can honestly say whether a true variance for a time series exists, yet we take on faith that we are "estimating" it in the following way. Virtually every thesis on the estimation of variance starts with three statements taken as fact: (1) a "true" variance exists, when (2) time is infinitely extended, and (3) we do not know (1) because we cannot realize (2). We argue that we could know (1) if our observation were infinite in time. But what constitutes an infinite observation time? Is it from the big bang (call it t_0) to now (call it t_M)? No, that itself is a finite comprehension. Is every event prior to the moment of the big bang classed as uninteresting or 0 as in figure 2? Then all variances (for that matter, all statistical averages) approach 0, and the concept of a "true" variance is a hoax. Hence, whatever we determine as a variance for any observation period is not an estimate at all but is in fact the true (and only) variance for that observation time, however short or long the time may be.

The only correct answer to the question of the existence of a true variance in the statistically pure sense is to say, "I don't know." Furthermore, if in any way it is implicit that the estimate of a variance for an observation time T is taken to be the true variance (which is a reasonable assertion), then the variables measured cannot be treated as independent of those not measured at any other imagined observation time. In summary, all statistical tools applied to a time series based on the assumption of ergodicity will, taken as a whole, require that the actual finite measurement repeat itself for all times t.

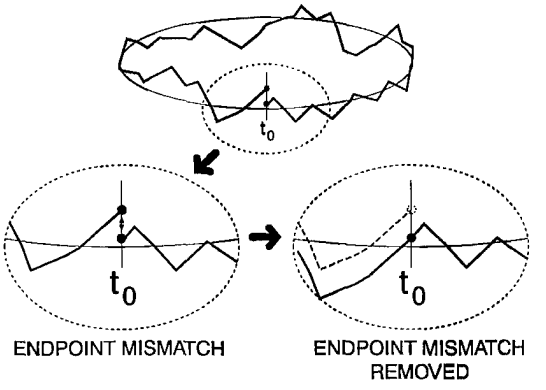


Figure 8

PRACTICALITY

With the circular representation presented here, residual data (that is, data with trends removed) which are wrapped often will have the endpoints miss each other, that is, not match. This is shown in figure 8. The discontinuity of the endpoint mismatch may produce components outside of the physical, system-related, or measurement bandwidth and will create an artifact and a commensurate error in a statistical average such as variance. Furthermore, endpoint mismatch affects overall scaling. More accurate results are obtained if the ends do not miss each other when data is wrapped. For computing statistical averages, we can arrange coordinates so that the ends will match [Howe, to be published].[2]

For any observation period starting at time t_o, continuing with length T, and ending at t_M, the next point after t_M (if our period is extended by an interval Δt) is t_o. In other words, the epoch which marks the end of our observation allows us to predict the next point, the variable at t_o since t_o starts the next repetition. Since this next variable is predictable, it is not a random variable, and the occurrence of the end of the actual period forces a correlation of t_M with t_o The next point is the future, and is outside our window of observation, but at any instant, the next point is surmised to be the one at the beginning time t_o. Furthermore any piece of the total coordinate of time is self-contained and can be treated as isolated from any events outside of that piece. Any observation is treated as circular as shown in figure 7. Now in the extreme case, suppose t_o marks the moment of creation or the big bang and duration T extends to now. Therefore, t_M marks the present (now). The end then becomes the beginning. This is not to say that all events repeat themselves nor that periodicity exists everywhere. Quite the contrary. Random events are not redundant and repeated; each event is unique and need not have any determinable cause (satisfying the observer, at any rate). That periodicity exists everywhere (including outside our observation window) would be a man-made assumption.[3] Let me emphatically state that the conclusion that the end and beginning are one derives from a reasonable assumption that all of the time we can possibly comprehend is still a finite observation time. This model of time correlates "now" with our chosen time origin, or if you wish, the so-called "beginning of time." This concept suitably derives from the question, "Do unobservable events exist?" For the purpose here, I assume they do not, since there can be no cognizance of such events. This is opposed to many physicists' views on realism and takes the stand that a thing does not exist for the observer before his measurement of it. I believe that the observer, however, does not give it its existence, so the future is not given any attribute prior to its observation. Needless to say, there is a great deal left for further philosophical discussion.[4] However, if we accept the hypothesis of ergodicity and accept that a true variance indeed exists, then we are at liberty to treat finite time-series data $\{x_k\}$ as circular. We can wrap the data so that $x_{M+1} = x_1$, and this is a practical tool. Any resultant mismatch of x_1 and x_M produces undesirable effects and ought to be treated as a separate element in our overall model of the observed process being measured.

Is there a coordinate system or reference frame in which time-ordered events can be viewed as nonrepeating or noncircular? This can be answered with an analogous geometrical question regarding space-ordering: Is there a reference frame in which a line (of infinite length and consisting of an infinite number of points) can be viewed on its end, so to speak, as a single point? No. We assume that a line cannot be viewed on end. If it is nowhere viewable as anything except a line, Einstein suggested that it comes back on itself. Every end of the line is also its beginning. We define absolute straightness as a minimum distance between two points in space. We are predisposed to thinking that the line cannot curve around; it is straight, and infinitely long. If space is not curved, how can this be? If space is curved, then how and to what extent is it curved so that everything comes back on itself (since the line becomes a geodesic of unknown dimension)? Perhaps it is not describable by finite human comprehension except to say that infinite space is somehow circular in that it comes back on itself, but any piece is somehow straight by every conceivable test. And all time vs. finite periods may be similarly viewed.[5]

CONCLUSION

Predicting time-dependent measures means predicting a future which is always outside qof our observation window. Observables have varying degrees of reproducibility, hence predictable reoccurrence. There are accurate models for fairly simple physical realizations. Randomness, however, is associated with nondeterministic, complex systems in which simple trends have been modeled and removed. This paper has described the appropriateness of wrapping

a finite time-dependent measurement as a tool for the determination of statistical averages such as the variance of non-deterministic processes. It behooves us to pick uncomplicated models. Conditions which satisfy an assumption of ergodicity yield the very simple model that time-ordered data may be treated as circular or wrapped. An endpoint mismatch must be removed because it affects statistical averages by a transient and by a scaling factor which can be a dominant source of errors. Future analysis will include the circular representation (wrapping the data) if appropriate. A remaining topic is whether the ergodic assumption is reliable for arbitrary constant power-law processes such as the kinds used to characterize frequency stability.

ACKNOWLEDGEMENTS

I gratefully acknowledge review and insightful discussion of this manuscript by J. A. Barnes, J. L. Jespersen, J. Levine, D. B. Percival, and G. M. R. Winkler.

FOOTNOTES

[1]The general case of nth-order function yields [Panter, 1965]:

$$\tilde{x}^n = \int_{-\infty}^{\infty} x^n p_n(x,t)dt$$

which is equal to the time average of x^n ($p_n(x,t)$ is the probability density of x^n and is assumed to be uniform-random and normalized to T):

$$\bar{x}^n = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [x(t)]^n dt$$

so that $\tilde{x}^n = \bar{x}^n$.

[2]It is thoroughly recognized that the start and stop of the observation window can adversely affect the outcome of particular analyses. For example, an assumption that periodicity is within some set of data might be physically correct. Therefore, Fourier analysis is suitable. Smoothly tapering the ends of a series to zero is often done to reduce errors at the expense of analysis resolution in the periodic extension of the observation, a condition brought about in the implementation of the Fourier transform integral.

[3]Joseph Fourier in 1822 rigorously treated the suggestion of Pythagoras, Kepler, Galileo, and Newton that wave or cyclical motion was everywhere in nature. So far it holds that all observable activity can be broken down into wave-like building blocks. Many types of analysis proceed from this assumption.

[4]Certainly we must reconcile whether the existence of anything depends exclusively on its finite observation.

[5]A long-standing view is held that space and time are sensible geometries for all human experience, but that they are different forms, one for the outer sense (space) and one for the inner (time) [Kant, 1781].

REFERENCES

Barnes, J., et al., "Characterization of Frequency Stability," IEEE Trans. Instrum. Meas., vol. IM-20, May 1971, p. 107.

Bingham, C., Godfrey, M. D., and Tukey, J. W., "Modern techniques of power spectrum estimation," IEEE Trans. Audio Electroacoustics, AU-15,56-66, 1967.

Bloomfield, P., Fourier analysis of time series: An introduction, Wiley, New York, 1976. Pages 55-60 are a useful discussion of the "circular definition" of a roughness coefficient, the variance, using "wrapped" data. Bloomfield does not, however, give reasons for permitting data to be wrapped.

Box, G. E. P. and Jenkins, G. M., Time series analysis forecasting and control, Holden-Day, San Francisco, 1970.

Churchill, R.V., Fourier series and boundary value problems, McGraw-Hill, New York, 1963, p. 103-109.

Gödel, K., On formally undecidable propositions (translation of 1931), Basic Books, New York, 1962.

Halmos, P. R., Lectures on ergodic theory, The Mathematical Society of Japan, Kenkyushu Printing Co., Ltd., Tokyo, 1956.

Hofstadter, D. R., Gödel, Escher, Bach: an eternal golden braid, Random House, Vintage Books, New York, 1980.

Howe, D. A., "Simple elimination of the window transient for accurate measures of variance," to be published. The subject of the window start and stop transients is addressed in this article.

Kant, I., The Critique of Pure Reason as explained by Bertrand Russell, A History of Western Philosophy (orig. 1945), Simon and Schuster, New York, 1972, pg. 712.

Otnes, R. K. and Enochson, L., Applied Time Series Analysis, Vol. 1, Basic Techniques, John Wiley and Sons, New York, 1978, pg.31.

Panter, P.F., Modulation, noise, and spectral analysis, McGraw-Hill, New York, 1965, pg. 134.

Percival, D. and Walden, A., Spectral analysis for physical applications, Cambridge University Press, New York, 1992.

Poincaré, H., Science and hypothesis, (translation of 1912), Dover Publication, New York, 1952.

Welch, P. D., "The use of fast Fourier transform for the estimation of power spectra: a method based on time-averaging over short, modified periodograms," IEEE Trans. Audio Electroacoust., Vol. AU-15, pp. 70-73, June, 1967.

Yaglom, A. M., Correlation theory of stationary and related random functions, Vol. I: Basic Results, Springer-Verlag, 1987, p. 216.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ESTIMATING THE MODIFIED ALLAN VARIANCE

Charles A. Greenhall

Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Dr, MS 298-100
Pasadena, CA 91109 USA

Abstract

The third-difference approach to modified Allan variance (MVAR) leads to a tractable formula for a measure of MVAR estimator confidence, the equivalent degrees of freedom (edf), in the presence of power-law phase noise. The effect of estimation stride on edf is tabulated. A simple approximation for edf is given, and its errors are tabulated. A theorem allowing conservative estimates of edf in the presence of compound noise processes is given.

Introduction

The ingredients for this work were presented three years ago at this Symposium. The first ingredient, a paper by the present author [6], shows how the labor of computing modified Allan variance (MVAR) estimates can be reduced by expressing MVAR in terms of third differences of the cumulative sum of time residuals. This approach shows that an MVAR estimate is hardly more difficult to compute than a conventional Allan variance (AVAR) estimate. A review of the method is given below. The second ingredient is a paper by Kasdin and Walter [10] on simulating a class of discrete-time power-law noises. In a subsequent paper [12], Walter exploits these noise models to derive a formula for the variance of a fully overlapped MVAR estimator. Combined with a formula for the estimator mean (MVAR itself), this formula can be used for computing an estimator confidence measure, the equivalent degrees of freedom (edf, defined below). In turn, edf can be used for assigning confidence intervals.

Walter's expression is difficult to evaluate. Happily, the combination of Walter's models with the third-difference approach has led to another formula for edf, mathematically equivalent to Walter's formula, but easier to evaluate because it has fewer summation terms. This formula is given below, together with additional results as follows.

- An assessment of the dependence of the edf of an MVAR estimator on its estimation period τ_1, defined as the time interval by which the summands of the estimator are shifted. It turns out that a wide range of choices of τ_1 gives essentially the same edf. The user can choose τ_1 from considerations of convenience and computational effort.
- A simple approximation formula for edf, with coefficients drawn from a brief lookup table. Most users will not need the exact edf formula.
- A theorem that allows one to calculate conservative values of estimator edf in the presence of a polynomial phase noise spectrum, i.e., linear combinations of power laws with unknown coefficients. This theorem is also valid for AVAR estimators, but is more useful for MVAR estimators because their edf varies less with power-law noise exponent.

The most critical assumption underlying these results is a negligible rate of linear frequency drift, or a drift rate that is known a priori; in this case, it can be removed from the data.

This paper mainly gives results; a longer paper with more derivations [7] has been submitted elsewhere.

MVAR and Its Estimators

Third-Difference Formulation

Let $x_1, x_2, \ldots$, with sample period τ_0, be a sequence of time residuals obtained from a comparison of clocks or from a phase comparison of two frequency sources. The conventional Allan variance for an averaging time $\tau = m\tau_0$ is defined by

$$\sigma_y^2(\tau) = \frac{1}{2\tau^2} E[\Delta_m^2 x_n]^2, \qquad (1)$$

where E denotes mathematical expectation (ensemble average), and Δ_m is the backward

0-7803-2500-1/95 $4.00 © 1995 IEEE

difference operator with step m, that is,

$$\Delta_m f_n = f_n - f_{n-m},$$

$$\Delta_m^2 f_n = f_n - 2f_{n-m} + f_{n-2m},$$

for any sequence f_n.

For the modified Allan variance, define the moving averages of x_n by

$$\bar{x}_n(m) = \frac{1}{m} \sum_{j=0}^{m-1} x_{n-j}.$$

The conventional definition of MVAR is

$$\text{mod } \sigma_y^2(\tau) = \frac{1}{2\tau^2} E\left[\Delta_m^2 \bar{x}_n(m)\right]^2. \qquad (2)$$

The third-difference formulation of MVAR uses an auxiliary sequence w_n of cumulative sums of x_n, defined by

$$w_0 = 0, \quad w_n = \sum_{j=1}^{n} x_j. \qquad (3)$$

This sequence can be generated from the recurrence $w_n = w_{n-1} + x_n$. Observe that

$$\bar{x}_n(m) = \frac{1}{m} \Delta_m w_n, \quad n \geq m.$$

When this is substituted into (2), the difference operators multiply to give

$$\text{mod } \sigma_y^2(\tau) = \frac{1}{2\tau^2 m^2} E\left[\Delta_m^3 w_n\right]^2 \qquad (4)$$

$$= \frac{1}{2\tau^2 m^2} E\left[w_n - 3w_{n-m} + 3w_{n-2m} - w_{n-3m}\right]^2.$$

This is the third-difference form of MVAR. The advantage of (4) over (2) is that it expresses MVAR in terms of four values of w_n instead of 3m values of x_n.

MVAR Estimator with Stride

To estimate MVAR with limited data, we replace the E operator in (4) by a finite average over n. For such a time average, we have to decide how much to increase n from one term to the next. This increase, denoted here by m_1, is called the estimation stride. The corresponding time shift $\tau_1 = m_1\tau_0$ is called the estimation period. When computing AVAR from (1), it is customary to use $m_1 = 1$, called "full overlap", or $m_1 = m$, called "τ overlap". The effect of these choices on AVAR estimator confidence has previously been computed ([5] and references therein). In the context of MVAR, the overlap formalism becomes awkward, and is replaced here by the stride formalism. The existing literature on MVAR ([1], for example) customarily assumes a stride of 1 (with good reason, as we shall see later). Here, we shall allow m_1 to vary between 1 and m, and investigate the effect on the confidence of the resulting estimator.

Suppose that N time residuals $x_1, x_2, \ldots, x_N$ are available. From these come N+1 cumulative sums $w_0, w_1, \ldots, w_N$, and N-3m+1 samples of $\Delta_m^3 w_n$, $3m \leq n \leq N$. Let M be the number of samples of $\Delta_m^3 w_n$ that are separated by the stride m_1. Then

$$M = \text{int}\left[\frac{N - 3m + m_1}{m_1}\right], \qquad (5)$$

where int(x) is the greatest integer that is $\leq x$. The MVAR estimator to be studied is given by

$$V = \frac{1}{2\tau^2 m^2 M} \sum_{k=0}^{M-1} \left[\Delta_m^3 w_{3m+km_1}\right]^2. \qquad (6)$$

Equivalent Degrees of Freedom

One measure of the statistical confidence of an estimator X is its equivalent degrees of freedom (edf), defined by

$$\text{edf } X = \frac{2(EX)^2}{\text{var } X}. \qquad (7)$$

Higher values of edf mean that the distribution of X is more concentrated about its mean. If X is distributed as a constant multiple of a chi-squared random variable with ν degrees of freedom, then edf $X = \nu$. Even if X does not have such a distribution, edf X can still serve as a convenient dimensionless measure of the confidence of X as an estimator of its mean. In this case, edf X need not be an integer. I take this point of view with regard to V, not having studied the nature of its distribution under the noise models discussed below. In frequency-stability analysis, it is customary to assume that estimators of AVAR or MVAR obey an approximate chi-squared law, and, on this basis, to construct confidence intervals for AVAR or MVAR [9][15] from levels of

the appropriate chi-squared distribution function.

Noise Models

The statistical properties of V, its edf in particular, depend on the random process chosen to model the time residuals x_n. The classical continuous-time spectral model for phase or time deviations is a linear combination of power laws:

$$S_x^+(f) = \sum_{\beta=-4}^{0} g_\beta f^\beta, \qquad (8)$$

whose components, for $\beta = 0,-1,-2,-3,-4$, are called white phase, flicker phase, white frequency, flicker frequency, and random-walk frequency. (The plus sign indicates one-sided spectral density.) It is understood that there is some high-frequency cutoff, the "hardware bandwidth", and that the power-law components of (8) might only behave asymptotically like f^β as $f \to 0$. Bernier [2] studied the behavior of MVAR for each of these spectral components, tackling the complex interaction among the hardware bandwidth B, the sample period τ_0, and the averaging time τ. Here, we follow Walter [12] in using explicit discrete-time power-law models for the samples x_n of the time residual process. These are the so-called fractional-difference processes [3][8], which have one-sided spectral densities proportional to

$$S_x^+(f) = 2[2 \sin(\pi f \tau_0)]^\beta, \quad f < \frac{1}{2\tau_0}. \qquad (9)$$

Nonintegral values of β are allowed.

There are two reasons for using these models here. First, the abovementioned complications of sampling the continous-time models are avoided. Second, the models fit perfectly into the MVAR third-difference framework. In particular, the sequence w_n defined by (3) is also a fractional-difference process with exponent β-2, that is,

$$S_w^+(f) = 2[2 \sin(\pi f \tau_0)]^{\beta-2}. \qquad (10)$$

Now, since MVAR has been given in terms of w_n, there is no need to use x_n in the theory.

Generalized Autocovariance

The frequency-domain description (10) of the model for w_n has an equivalent time-domain description, called the generalized autocovariance (GACV) and denoted by $R_w(n)$, where n runs through all integers, positive and negative. The concept of autocovariance (ACV) as a function of one time variable applies to stationary processes only. With some care, though, the ACV admits an extension to certain nonstationary processes in such a way that their covariance properties can be described in terms of a function, the GACV, that still depends on one time variable. Although the GACV cannot be regarded as a covariance function in the usual sense, it can be used like one under certain restrictions.

Because the GACV $R_w(n)$ plays a central role in the formula for edf V given below, we give this function here for all the required values of β, namely $-4 \le \beta \le 0$. Bear in mind that the noise-type label applies to x_n, a power-law process with exponent β, while $R_w(n)$ applies to a power-law process w_n with exponent β-2. For the flicker noises we need an auxiliary sequence L_n, a discrete version of the logarithm, defined by

$$L_0 = 0, \quad L_n = \sum_{j=1}^{n} \frac{1}{j - 1/2}.$$

Following are the required GACV formulas.

$\beta = 0$; white phase

$$R_w(n) = \frac{-|n|}{2\tau_0}$$

$\beta = -1$; flicker phase

$$R_w(n) = \frac{-1}{2\pi\tau_0}\left[\frac{1}{4} - n^2\right]L_{|n|}$$

$\beta = -2$; white frequency

$$R_w(n) = \frac{-|n|(1 - n^2)}{12\tau_0}$$

$\beta = -3$; flicker frequency

$$R_w(n) = \frac{-1}{24\pi\tau_0}\left[\frac{1}{4} - n^2\right]\left[\frac{9}{4} - n^2\right]L_{|n|}$$

$\beta = -4$; random-walk frequency

$$R_w(n) = \frac{-|n|(1 - n^2)(4 - n^2)}{240\tau_0}$$

β nonintegral

$$R_w(n) = \frac{-\Gamma(1-\beta/2+n)}{2\tau_0 \cos(\pi\beta/2)\ \Gamma(2-\beta)\ \Gamma(\beta/2+n)}$$

The formula for nonintegral β is equivalent to the form used by Kasdin and Walter [10] and by Walter [12], but for the GACV of x_n, not of w_n.

Additional Mathematical Assumptions

For technical correctness, it is assumed that the the time residuals x_n have stationary, Gaussian, mean-zero second increments. Assuming that these increments have zero mean is the same as assuming that the frequency drift rate is zero.

Results

MVAR Estimator edf: Exact Formula

In the estimator defined by (5) and (6), recall that the averaging time is $m\tau_0$, and that the estimation period is $m_1\tau_0$. For nonnegative integers n, let

$$R_n = -R_w(n-3m) + 6R_w(n-2m) - 15R_w(n-m) + 20R_w(n) - 15R_w(n+m) + 6R_w(n+2m) - R_w(n+3m). \quad (11)$$

In other words, $R_n = -\delta_m^6 R_w(n)$, where δ_m^6 is the sixth central difference operator with step m. Actually, R_n is just the ordinary ACV of the stationary process $\Delta_m^3 w_n$. Let

$$\rho_n = \frac{R_n}{R_0},$$

be the corresponding autocorrelation sequence. The formula for edf V is given by

$$\frac{1}{\text{edf V}} = \frac{1}{M}\left[1 + 2\sum_{k=1}^{M-1}\left[1 - \frac{k}{M}\right]\rho_{km_1}^2\right]. \quad (12)$$

This formula is mathematically equivalent to Walter's formula for var V ([12], eq (32)), but requires less computation. Evaluation of (12) requires 7M evaluations of $R_w(n)$. Walter's formula, which is given only for $m_1 = 1$, is a double sum requiring $5(2m-1)(2M-1)$ evaluations of the GACV of x_n. This shows the advantage of the third-difference approach, which derives MVAR estimator summands from four values of w_n instead of from 3m values of x_n.

In connection with a recent conference paper [15], tables of edf V for $m_1 = 1$ and integral β were generated by the method given here, by Walter's method, and by Monte Carlo simulation. The two theoretical methods agreed within 0.1 percent; the simulations agreed with the theoretical results within a few percent.

A note on computation: The ACV R_n tends to zero as $n \to \infty$, yet is obtained from differences of $R_w(n)$, which tends to ∞ with n. Clearly, one should use double precision for evaluating (11). Even so, the computed values of R_n can deteriorate for large n, especially for nonintegral β, where $R_w(n)$ involves Γ functions. I was able to cure this problem by replacing the upper limit M-1 of the summation in (12) by K-1, where $K = \min(M, 10m/m_1)$. (In all actual computations, m/m_1 is assumed to be an integer.)

Effect of Estimation Period

From here on, we assume that the estimation period divides evenly into the averaging time, that is, we have

$$\frac{\tau}{\tau_1} = \frac{m}{m_1} = r,$$

where r is an integer. Under this assumption, (12) was used to generate tables of edf V for combinations of N, m, and m_1. For each combination, the number M of estimator summands is calculated from (5), and the parameter p is defined by

$$p = \frac{M}{r} = \frac{M\tau_1}{\tau}. \quad (13)$$

A selection of edf values is shown in Table 1 for integral values of β. Values for half-integral β were also computed, but are not shown; as expected, they interpolate the given values. For now, ignore the "%" rows, and observe how edf depends on r (or m_1) for N = 1024, m fixed. For each β, and for $m \geq 4$, it is clear that any value of r between 4 and m gives a value of edf that is nearly maximal for that m and β. If $m < 4$, then we should take $\tau_1 = \tau_0$ ($m_1 = 1$, $r = m$). For $\beta \geq -2$, an estimation period of τ ($m_1 = m$, $r = 1$) gives inferior results. Here is an empirical result that summarizes the observations.

Assume an averaging time τ at most 1/4th the duration of the time-deviation record. For each power law between white phase and random-walk frequency, any estimation period τ_1 between τ_0 and $\max(\tau_0, \tau/4)$ that divides evenly into τ gives an MVAR estimator V whose edf is within 8 percent of the maximal value for τ.

Table 1 shows that the variation of edf V with r is greatest for white phase

Table 1. Values and Approximation Errors for MVAR Estimator edf

N = 1024					β					
m	r	m_1	M	p	0.0	-1.0	-2.0	-3.0	-4.0	
1	1	1	1022	1022.	525.9	589.3	681.6	828.6	1022.	
					+0.0	+0.0	+0.0	+0.0	+0.0	%
2	1	2	510	510.0	262.6	310.1	380.8	459.1	432.3	
	2	1	1019	509.5	477.0	496.5	515.2	523.6	441.4	
					-0.1	-0.1	-0.1	-0.1	-0.1	%
3	1	3	339	339.0	174.6	210.3	260.1	304.4	271.0	
	3	1	1016	338.7	373.9	349.9	341.5	334.6	274.0	
					+11.1	-2.8	-3.9	-4.1	-5.0	%
16	1	16	62	62.00	32.15	39.57	48.69	55.29	47.55	
	2	8	123	61.50	58.06	59.26	59.68	58.73	47.60	
	4	4	245	61.25	72.74	61.99	59.93	58.57	47.43	
					+4.1	+0.1	-0.2	-0.3	-0.2	%
	8	2	489	61.13	77.60	62.26	59.84	58.46	47.33	
					-2.6	-0.6	-0.2	-0.3	-0.2	%
	16	1	977	61.06	78.88	62.26	59.78	58.40	47.29	
					-4.3	-0.7	-0.2	-0.3	-0.2	%
128	1	128	6	6.000	3.375	4.061	4.909	5.552	4.766	
	2	64	11	5.500	5.754	5.841	5.857	5.716	4.535	
	4	32	21	5.250	7.005	5.922	5.706	5.525	4.367	
					+3.4	+0.4	-0.1	-2.3	+0.2	%
	8	16	41	5.125	7.354	5.840	5.599	5.417	4.277	
					-3.6	-0.3	-0.3	-2.5	+0.0	%
	16	8	81	5.063	7.410	5.784	5.542	5.361	4.231	
					-5.3	-0.4	-0.4	-2.6	+0.0	%
	32	4	161	5.031	7.405	5.755	5.513	5.332	4.207	
					-5.8	-0.4	-0.4	-2.6	+0.0	%
	64	2	321	5.016	7.394	5.739	5.498	5.318	4.196	
					-5.9	-0.4	-0.4	-2.6	+0.0	%
	128	1	641	5.008	7.386	5.732	5.491	5.311	4.190	
					-5.9	-0.4	-0.4	-2.6	+0.0	%
N = 16										
1	1	1	14	14.00	7.475	8.327	9.561	11.51	14.00	
					-3.7	-3.1	-2.4	-1.4	+0.0	%
2	2	1	11	5.500	5.754	5.946	6.117	6.146	5.061	
					-10.6	-10.0	-9.2	-8.1	-5.9	%
3	3	1	8	2.667	3.815	3.526	3.386	3.224	2.508	
					+9.9	-2.0	-3.0	-7.2	-3.5	%
			noise type:		wh ph	fl ph	wh fr	fl fr	rw fr	

(β = 0). Also, we see that p itself is a rough estimate of edf V, especially for m_1 in the recommended range $1 \leq m_1 \leq \max(1,m/4)$.

The choice of estimation period τ_1 might depend on a tradeoff between convenience and computational effort. For small data sets that are held entirely in memory, the minimal choice $m_1 = 1$ is convenient, and the computational cost is probably negligible. For larger data sets that are read sequentially from a file, the maximal choice $m_1 = m/4$ allows sequential accumulation of MVAR sums from the stream of w_n with moderate use of memory. As an example, take m = 32, m_1 = 8. To update the sum of squares of $\Delta_{32}^3 w_n$ at every eighth sample of w_n, the program has to remember the previous 12 values of w_{8j}. Alternatively, if there are many thousand data points, one can simply use $m_1 = m$ to accumulate sums of squares of third differences for smaller values of m, while collecting a global buffer of w_n subsampled by some factor m_2. After all the data are read, the buffer is used for calculating MVAR estimates with $m_1 = m_2$, $m = rm_1$ for various r.

MVAR Estimator edf: Approximate Formula

Because the power-law models are only an approximate fit to actual phase noise, the precision of the theoretical values of edf V in Table 1, four significant figures, is meaningless for a user who needs to construct error bars for MVAR measurements. Therefore, the following simple approximation is offered as an empirical result.

Assume power-law phase noise with exponent β between -4 (random-walk frequency) and 0 (white phase), at least 16 time-residual points, an averaging time τ at most 1/5th the duration of the measurement, and an estimation period τ_1 between τ_0 and $\max(\tau_0, \tau/4)$ that divides evenly into τ. In our notation, $N \geq 16$, $m \leq N/5$, and $m = rm_1$, where r is an integer between $\min(m,4)$ and m. For the estimator V defined by (6), we have

$$\text{edf V} \approx \frac{a_0 p}{1 - \frac{a_1}{p}}, \qquad (14)$$

where $p = M/r$, M is given by (5), and the coefficients a_0, a_1, as functions of m and β, are drawn from Table 2.

The relative error of this approximation is observed to be at most ±11.1 percent.

Each "%" row in Table 1 shows the percentage errors of (14) for the row above. The table entries were chosen to represent the full range of observed errors. This approximation holds only under the above restrictions on data set size and averaging time. For example, if $m = N/4$ then the exact edf formula (12) must be used.

This approximation was derived from two rigorous lower bound formulas, one for edf V, the other for the edf of a continuous-time analog of V. The choice between these two bounds as approximations was made partly by insight, partly by trial and error.

Compound Noise Spectra

The foregoing results assume a power-law phase noise spectrum proportional to (9) for some fixed exponent β. If that were indeed the case, our statistical efforts ought to be directed toward estimating the two-parameter set consisting of β and the constant of proportionality. Instead, as usual, we find ourselves using parametric tools to evaluate the confidence of a nonparametric statistic. The value of edf V depends on β. What can we do in the presence of a compound phase noise model

$$S_x^+(f) = \sum_\beta g_\beta \sin^\beta(\pi f \tau_0), \qquad (15)$$

a finite sum of fractional-difference spectra? Some help is given by the following theorem, which, although weak and perhaps obvious, is better than nothing.

Theorem. Let the phase noise be a finite sum of independent component noises with stationary Gaussian mean-zero second increments. Form an MVAR estimator V from the given phase noise, and corresponding estimators V_k from the components. Then

$$\text{edf V} \geq \min_k \text{edf } V_k.$$

In other words, we never do worse than the worst component.

To apply this theorem to the situation (15), assume that the component β values are all in some subinterval of [-4,0] (the whole range, perhaps). Use (14) and Table 2 to compute edf V_β for each tabulated β in the subinterval, and take the smallest value as a conservative estimate of edf V. For example, if one believes that the noise has components between white phase and flicker phase,

Table 2. Coefficients for Approximating MVAR Estimator edf

noise type	β	m = 1 a_0	m = 1 a_1	m = 2 a_0	m = 2 a_1	m > 2 a_0	m > 2 a_1
wh ph	0.0	.51429	0	.93506	0	1.2245	.58929
	-0.5	.54277		.95407		1.0739	.59605
fl ph	-1.0	.57640		.97339		1.0030	.60163
	-1.5	.61688		.99246		.97732	.59769
wh fr	-2.0	.66667		1.0101		.96774	.57124
	-2.5	.72948		1.0237		.96102	.50974
fl fr	-3.0	.81057		1.0266		.94663	.41643
	-3.5	.91389		.99981		.90604	.34276
rw fr	-4.0	1.0000		.86580		.76791	.41115

perhaps from prior knowledge, perhaps from a log-log σ-τ plot with slopes between -3/2 and -1, then one can minimize (14) over the first three rows of Table 2.

This theorem can be generalized to AVAR estimators and other situations involving averages of the square of a stationary Gaussian mean-zero process. Its usefulness for MVAR, as opposed to AVAR, is enhanced by the relatively weak dependence of MVAR estimator edf on β, as can be seen in Table 1. Similar tables for fully overlapped AVAR estimators [5][11] show a sharper dependence on β, especially for large τ/τ_0. Thus, minimizing estimator edf over β causes a smaller loss of accuracy for MVAR than for AVAR.

Concluding Remarks; Future Work

The previous paper on the third-difference approach [6] showed that MVAR estimates are almost as easy to calculate as AVAR estimates. The results given here extend this conclusion to the exact formulas for the confidence of the estimators. In addition, the approximation formulas for MVAR confidence are simpler and more uniform than existing approximation formulas for AVAR confidence [5][9], and the confidence values are more robust to spectral uncertainties. Having overcome the apparent increase in complexity of the extra moving-average filter in MVAR, we are free to enjoy all its advantages.

The problem of frequency drift removal now needs to be addressed. For AVAR, it is known that estimation of drift rate from the data themselves, and removal therefrom, causes negative estimator bias that worsens as averaging time τ increases. The use of three-point [13][14] or four-point [4] drift estimators, which extract a quadratic component of the time-residual sequence x_n, simplifies the calculation of the mean and variance of AVAR estimators with drift removed. I have no doubt that similar calculations for MVAR estimators can be made by using four-point drift estimators that extract a cubic component of the cumulative-sum sequence w_n.

The work described in this paper was performed by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] D. W. Allan, "Time and frequency (time-domain) characterization, estimation, and prediction of precision clocks and oscillators", IEEE Trans Ultrasonics Ferroelectrics Freq Control, vol UFFC-34, pp 647-654, 1987

[2] L. G. Bernier, "Theoretical analysis of the modified Allan variance", Proc 41st Ann Freq Control Symp, pp 116-121, 1987

[3] C. W. J. Granger, R. Joyeux, "An introduction to long-memory time series models and fractional differencing", J Time Series Analysis, vol 1, pp 15-29, 1980

[4] C. A. Greenhall, "The fundamental structure function of oscillator noise models", Proc 14th Precise Time and Time Interval Applications and Planning Meeting, pp 281-294, 1982

[5] C. A. Greenhall, "Recipes for degrees of freedom of frequency stability estimators", IEEE Trans Instrum Meas, vol 40, pp 994-999, 1991

[6] C. A. Greenhall, "A shortcut for estimating the modified Allan variance", Proc 1992 IEEE Freq Control Symp, pp 262-264, 1992

[7] C. A. Greenhall, "The third-difference approach to modified Allan variance", submitted to IEEE Trans Instrum Meas, 1995

[8] J. R. M. Hosking, "Fractional differencing", Biometrika, vol 68, pp 165-176, 1981

[9] D. A. Howe, D. W. Allan, J. A. Barnes, "Properties of signal sources and measurement methods", Proc 35th Ann Freq Control Symp, pp 1-47, 1981

[10] N. J. Kasdin, T. Walter, "Discrete simulation of power law noise", Proc 1992 IEEE Freq Control Symp, pp 274-283, 1992

[11] S. R. Stein, "Frequency and time--their measurement and characterization", Precision Frequency Control, vol 2, E. A. Gerber and A. Ballato, eds, pp 191-232, 399-416, New York, 1985

[12] T. Walter, "Characterizing frequency stability: a continuous power-law model with discrete sampling", IEEE Trans Instrum Meas, vol 43, pp 69-79, 1994

[13] M. A. Weiss, D. W. Allan, D. A. Howe, "Confidence on the second difference estimation of frequency drift", Proc 1992 IEEE Freq Control Symp, pp 300-305, 1992

[14] M. A. Weiss, C. Hackman, "Confidence on the three-point estimator of frequency drift", Proc 24th Precise Time and Time Interval Applications and Planning Meeting, pp 451-460, 1992

[15] M. A. Weiss, F. L. Walls, C. A. Greenhall, T. Walter, "Confidence on the modified Allan variance and the time variance", 9th European Frequency and Time Forum, 1995

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SOME NEW DEVELOPMENTS OF PRECISION FREQUENCY MEASUREMENT TECHNIQUE

Wei Zhou, Zongqiang Xuan, Jianguo Yu

Dept. of Measurement and Instrumentation, Xidian University
Xi'an, 710071, P. R. China

Abstract

In recent years, a distinguishing feature of developments of precision frequency measurement techniques is highly precision measurement of wider frequency range. It is very useful for the development of communication, post and metrology. In this paper conventional frequency standard comparison techniques and corresponding wide frequency measurement techniques developed in recent years are compared. Based on the study of the characteristics of the greatest common factor frequency [1] and "phase coincidence" detection technique [2][3], the advantages of frequency standard comparison technique and comparison of two signals with the same nominal frequency can also be obtained in the frequency measurement of a very wide frequency range.

In the application of the new techniques, some special design methods are used. They are: 1. The method combined "phase coincidence" detection method with no dead time measurement method. 2. In the measurement of some special frequency signals, the uses of regular distribution of "phase coincidence" between two frequency signals. 3. The method with which the noise of measuring device and the noise of measured oscillator are divided in frequency characteristics. 4. For higher precision some analog circuits can be used instead of digital circuits.

This paper also analyses the possibility that the new technique and instruments are used instead of conventional digital frequency meters.

1. Introduction

The conventional frequency standard comparison techniques are highly precision, but their measuring frequency range is very narrow . The frequency measurement techniques based on "phase coincidence" detection is not only highly precision, but also of wide frequeny range.

If the two techniques are compared, a lot of similar characteristics can be found. The frequency measuring technique by broad- band phase detection (BPD) [2] is similar to beat frequency method, and the frequency measuring technique by dual broad-band phase detection (DBPD) is similar to the dual mixer time difference method. Based on the research and uses of the characteristics of the greatest common factor frequency (GCFF), the high precision can be obtained in a very wide frequency range. Because corresponding instruments are highly accurate, simple in structure, of low cost, they have been widely used in China.

In the uses of the new techniques, some special design methods are very useful for enhancing accuracy, simplifying structure, which is to be described later in this paper.

2. Comparisons of some frequency measurement techniques

The conception of the greatest common factor frequency and "phase coincidence" detection method can be used to overcome ± 1 count error[2] [3]which exists in most conventional frequency and time measuring instruments. The frequency measuring techniques by broad- band phase detection and dual broad-band phase detection are based on the new conception and method.

A lot of characteristics of the new frequency measuring methods are almost the same as some conventional frequency standard comparison methods [4], if one uses the greatest common factor frequency instead of

0-7803-2500-1/95 $4.00 © 1995 IEEE

frequency. the least common multiple period (LCMP) instead of period. In frequency measurement of wide frequency range, distribution of "phase coincidence points" can influence measurement accuracy and measuring response time. Through corresponding comparison, many frequency standard comparison techniques can be widely used in wide range frequency measurement.

Take a simplest example, we compare the Beat frequency method with The frequency measuring technique by broad-band phase detection. They have many same characteristics They are: simple structure, excellent precision, but with measuring dead time. They can be combined with a frequency synthesizer to improve the measuring response time. Beat frequency method is narrow in measuring frequency range, and BPD is wide. There is a frequency transformation in beat frequency method, but BPD is based on time process of signals. Compared with beat frequency method, BPD can be used more flexibly. The comparison of beat frequency method with BPD is shown in table 1.

Table 1 Comparison of beat frequency method with BPD

Measurement method	Beat frequency	BPD
measuring precision	high $10^{-13}/\tau$	high $10^{-10}\sim10^{-12}/\tau$
Frequency range	narrow	wide
Device	simple	simple
Dead time	exist (can be eliminated)	exist (can be eliminated)
Aided device	frequency synthesizer	frequency synthesizer
Key circuit	mixer	"phase coincidence" detection circuit
Processing method	frequency transformation	time process
Minimum measurement time τ	determined by period of beat frequency	influenced by LCMP or period of combination beat frequency

The comparison of Dual mixer time difference method (DMTD) with DBPD is similar to the comparison mentioned above. DBPD has the same advantages as DMTD. The noise of common oscillators can also be offsetted. Wide frequency measurement range and flexible uses of DBPD are its better characteristics than those of DMTD.

Compared with conventional frequency measurement techniques, BPD and DBPD are more accurate. Combined with aided device and technique, a normal measuring time or gate time can also be obtained.

3. Some special design methods

In the uses of the wide range frequency measurement techniques, some special circuits and design methods can be used to enhance accuracy and simplify structure. They are very useful in designing many general and special frequency measuring instruments.

(1) *The method of "phase coincidence" detection combined with no dead time measurement.* It is very useful for highly accurate frequency and frequency stability measurement of arbitrary signal with no dead time. By using this method, the reference time interval signal is combined with the "phase coincidence points" between two compared signals to form the continuous sampling time interval signal. At the moments of beginning and end of every sampling time interval, the contents of the counters are fed into the registers. The microcomputer is used to process the counts. The measuring equipment is shown in Figure 1.

The equation calculating measured signal frequency is:

$$f_x = f_r \frac{N_{x2} - N_{x1}}{N_{r2} - N_{r1}} \qquad (1)$$

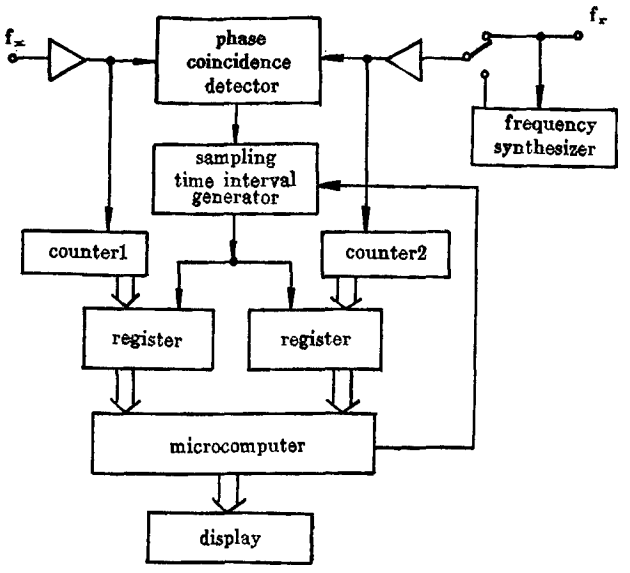


Figure 1. A no dead time precision frequency measurement equipment

where f_r is the frequency of reference signal; N_{x2}, N_{x1} are the cycle number counts of f_x at the end and beginning of sampling time interval, respectively; and N_{r2}, N_{r1} are the cycle number counts of f_r at the end and beginning of sampling time interval, respectively.

(2) *For some special frequency signals, utilizing the special "phase coincidence" distribution between two frequency signals, higher accuracy can be obtained and the device can be simpler.* When some wide frequency measuring instruments are used in frequency standard comparison, the phase difference variation between frequency signals in a least common multiple period or a preiod of difference frequency can be from large to small or from small to large regularly. In this case, one can use the stability of "phase coincidence" detection circuit to obtain higher detection accuracy. Generally, the stability of a phase coincidence detection circuit is one or more orders better than its accuracy. Based on this property, the measurement accuracy of an instrument can be enhanced by several orders. When the frequencies of two signals are approximately the same, or the frequency of a signal is at several times frequency of another signal, the distribution of "phase coincidence" is favourable for highly accurate measurement.

(3) *According to the characteristics of the greatest common factor frequency, many linear phase comparators with excellent linearity at high frequency can be designed [1] [5] [6].* Using these instruments, not only long term performance of a frequency signal, but also its short term performance can be measured with high precision. With these instruments, one can divide the noise of measuring device and the noise of measured oscillator in frequency characteristics. Some measuring instruments with great noise can give higher precision.

The linear phase comparator designed with the characteristics of the greatest common factor frequency is shown in Figure 2.

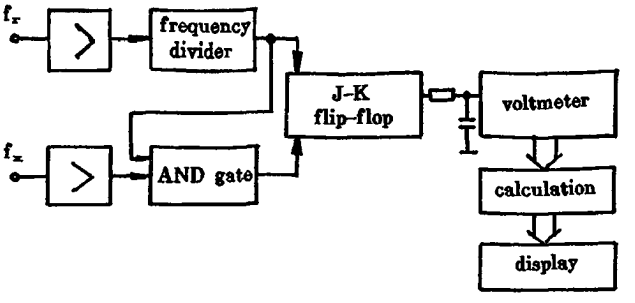


Figure 2. The linear phase comparator designed with the characteristics of the greatest common factor frequency

When this phase comparator is used, the nominal frequency of the two phase comparison signals should be the same, the phase discrimination period being equal to the period of the compared signals.

With this method, in any phase difference case at 5MHz frequency, the linearity of phase comparison is better than 0.5% (<1ns), and the "dead range" is less than 0.1% of phase comparison period; and about 0.1% linearity

can be obtained in most phase difference case. The variation of the linearity is stable. The structure of the device is simple, but its accuracy is better than many other instruments . Using the comparator, the "zero" and "full scale" of phase comparison must be regulated in operating conditions.

In some precision measurement of frequency signals, the noise of measuring device is larger than that of measured signal. For example, the triggering error is always unavoidable. Because the noise frequency of the triggering error is very high, one can divide it from the noise of measured signals. When the device is used to measure the short term stability of a measured oscillator, (τ =0.1s, 1s, 10s etc.) a digital voltage meter whose integration time is changeable is used as a terminal device. When the integration time of the digital voltage meter is selected, the noise of the measuring device can be restrained obviously, but the noise of the measured oscillator is not improved. For most osillators, the integration time of the voltage meter can be from 1ms to 20ms. The main measuring device noise (for example, the triggering error) has been averaged, and it can not influence the display of the measurement. The precision of measuring device is improved obviously. Changing the integration time of the digital voltage meter, the measuring results are shown in table 2.

From table 2, one can understand the importance of the integration time. For most frequency sources, the suitable integration time to measure phase comparison voltage is about 1ms. Long integration time may improve the noise of measured oscillator and display a wrong result.

By using this method, the equation for calculating the stability of measured oscillator is:

$$\sigma_y(\tau) = \frac{1}{\tau}\sqrt{\sum_{i=1}^{m}\frac{[(T_{i+1}-T_i)-(T_i-T_{i-1})]^2}{2m}} \quad (2)$$

Table 2. The measurement results of several frequency signals with different integration time

integration time \ frequency signals & their stability	crystal oscillator 1	crystal oscillator 2	Rb frequency standard
	$2.6\sim4.1\times10^{-10}$/s	$2.8\sim3.3\times10^{-12}$/s	$7.5\sim9.7\times10^{-13}$/s
1s	5.5×10^{-12}/s	$1\sim1.4\times10^{-12}$/s	
100ms	6.2×10^{-12}/s	$1.2\sim1.7\times10^{-12}$/s	$5\sim8.5\times10^{-12}$/s
20ms	$7\sim8\times10^{-12}$/s	$1.3\sim1.8\times10^{-12}$/s	$6.1\sim9\times10^{-12}$/s
2.5ms	2×10^{-10}/s		
1ms	$2.7\sim3.9\times10^{-10}$/s	$2.8\sim3.1\times10^{-12}$/s	$7.1\sim9.9\times10^{-12}$/s
0.3ms	$3.4\sim4.2\times10^{-10}$/s		

where T_{i+1}, T_i and T_{i-1} are the phase differences between two signals, respectively, at times t_{i+1}, t_i and t_{i-1}. They are represented with voltages.

The self - checking precision of the device is about $1\times10^{-13}/\tau$. When the same input circuit is used in other frequency standard comparator, the precision is one more orders lower than that of the device. Therefore, the method can improve the niose of measuring instruments obviously.

(4) *Many analog circuits can be used in the new frequency measuring instruments as digital functions.* In new wide range frequency measuring instruments, almost all circuits are digital. If one wants to obtain higher precision, some low noise analog circuits can take the place of the digital circuits. Because digital circuits operate with large signals, their speed limitation and noise of phase coincidence detection circuit will certainly have some influence on the detection precision. Therefore, accomplishing the phase coincidence detection with small

signal and low noise can enhance the precision further. A successful approach is to reshape the two comparison signals into small-amplitude narrow pulse, and detect the phase coincidence points by using a balanced mixer. The wide measuring range can be obtained , without any filter in the circuit. For the measurement of some special frequency signals, the precision can be two orders greater than that of devices with digital phase coincidence detection circuit.

4. Improvement of conventional digital frequency meters

Conventional digital frequency meters can be replaced by these new techniques and instruments. Because of their wide frequency measurement range, simple structure, higher accuracy, a lot of new time - frequency measuring instruments can be designed and used widely. For more convenient uses and regular gate time, some assistant techniques ought to be used. Using processing and control of soft ware a standard signal and a frequency synthesizer will work in co- operation with each other to accomplish accurate measurement. The diagram of the frequency meter is shown in Figure 3.

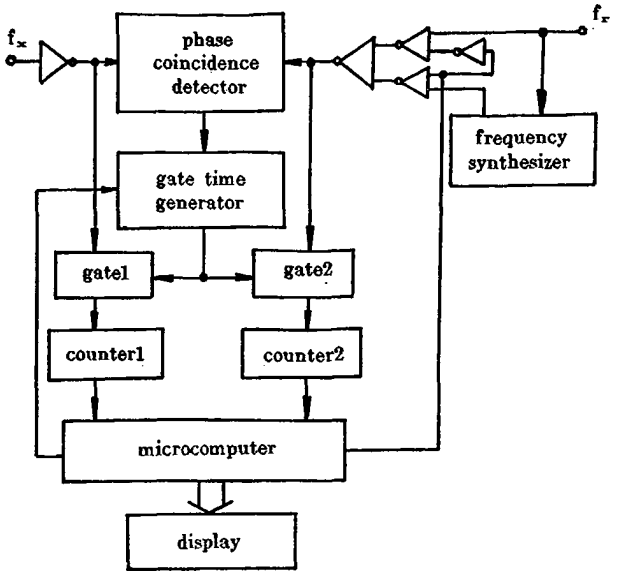


Figure 3. a new digital frequency meter

In general case, the measuring gate time signal depends on the reference gate time which is obtained by computer soft ware and the "phase coincidence" between standard and measured signal[1]. It is close to the reference gate time in length, but its starting and stopping moments correspond strictly to two "phase coincidence" between the two frequency signals. When only a standard frequency signal is used as reference signal, in a very wide frequency range the least common multiple period, period of combination difference frequency and distribution of "phase coincidence" vary in a very wide range. Therefore, more than one reference signal must be used, and the measurement of wide frequency range can be achieved. A frequency synthesizer can be used to cooperate with the standard signal as the reference signals to accomplish "phase coincidence" detection with the measured signal. The simple synthesizer can be locked by the standard frequency signal with a suitable frequency deviation. The reference [3] has illustrated that the greatest common factor frequency between two signals varies uncontinuously according to certain periodic property. As a frequency synthesizer with a single frequency cooperates with a standard frequency, the wide range measurement can be made satisfactorily. The frequency deviation between standard signal and the synthersizer can be decided by:

$$\Delta f = F_{maxc} \qquad (3)$$

where F_{maxc} is the maximum f_{maxc} in most cases for the synthesizer and standard signal cooperating with each other. Because the measuring precision and regular gate time are also required, Δf must not be too small. The computer soft ware decides which signal is

used as the reference signal of measurement. In most cases the reference signal is the standard frequency signal. After the reference gate time starts or stops, the measuring gate time can not come at a fixed time, then the synthesizer is chosen as the reference signal. The computer sends out a level signal to select the reference signal. The frequencies of standard signal and synthesizer are stored in the memory of the computer. The gate time of the frequency meter designed by this method can be 10ms, 100ms, 1s and 10s. Its precision is 2 orders higher than that of ordinary frequency meters. For more normal gate time, Δf is chosen as 1kHz.

5. conclusion

A development direction of high precision frequency measurement technique is to accomplish the high precision measurement in a very wide frequency range. The method described in this paper is a simple approach. In the measurement process, there is not any frequency transformation. Therefore, it is of wide frequency range. This method is based on the processing of the time relationship between the two signals, and can be used to eliminate ± 1 count error which exists in conventional frequency measuring instruments. Besides frequency and time measurement instruments, the new method has been used in some simple frequency synthesizers and network analyzers.

References

[1] Wei Zhou "The Greatest Common Factor Frequency and Its Application in the Accurate Measurement of Periodic Signals", Proceedings of the 1992 IEEE Frequency Control Symposium, pp. 270 - 273, 1992

[2] Wei Zhou "The Frequency Measurement Technique by Broad-band Phase Detection", Chinese Journal of Scientific Instrument, Vol. 14, No. 4, (1993), pp. 358 - 362

[3] Wei Zhou, Ping Liang "Phase Difference Variation Characteristics between Frequency Signals and Its Uses in Measurement" Proceedings of 1994 IEEE Instrumentation and Measurement Technology Conference, pp. 810 - 813, 1994

[4] D. A. Howe, D. W. Allan, J. A. Barnes "Properties of Signal Sources and Measurement Methods" Proceedings of the 35th Annual Symposium on Frequency Control, 1981

[5] Wei Zhou "Linear Phase Comparison Methods for Measuring the Time Delay of Devices and Equipment", Journal of Xidian University, Vol. 21 No. 1, pp. 122 - 125, 1994

[6] Wei Zhou, Zongqiang Xuan "A New Method of Phase Fluctuation Measurement of Signals with any Frequency" Chinese Journal of Scientific Instrument, Vol. 15, N0. 4, pp. 425 - 428, 1994

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A TEST SUITE FOR THE CALCULATION OF TIME DOMAIN FREQUENCY STABILITY

W.J. Riley, EG&G, Inc., Salem, MA 01970

ABSTRACT

This paper describes a test suite to check the accuracy of frequency stability analysis software. It contains the values of several common statistics and frequency stability measures for two data sets, a small one suitable for manual entry, and a larger one produced by a portable pseudo-random number generator. The paper also discusses related issues such as data gaps and outliers, phase-frequency conversions, drift removal, numerical precision, plotting, and simulation.

INTRODUCTION

Specialized calculations are necessary to express the results of time domain frequency stability measurements [1,2,3]. A common example is the Allan variance for a set of fractional frequency data. Such calculations are generally performed by a computer, for which a custom program may need to be written and debugged. Each generation of computer hardware and operating system usually requires an update of the software, which must then be validated before use. A suite of test data, for which correct values of common frequency stability measures are known, can be a valuable tool for checking the accuracy of frequency stability analysis software.

DATA TYPES AND TERMINOLOGY

The time domain stability of a frequency source can be measured by either phase or frequency data. The former is normally expressed as $x(t) = \phi(t)/2\pi\nu_0$, where ν_0 is the nominal frequency. This quantity has units of time, but is generally called "phase" to avoid confusion with the independent time variable, t. Frequency is normally expressed as fractional frequency, $y(t) = [\nu(t) - \nu_0]/\nu_0 = x'(t)$, which is dimensionless. When making stability measurements, it is preferable to take phase data, since it is the more fundamental quantity. The terms "frequency standard" and "clock" are often used interchangeably. A frequency source may be called a clock even though it does not contain any actual clock hardware, especially if it is used for timing purposes. The term "oscillator" is best used for an active source like a crystal oscillator rather than a passive device such as a cesium beam tube or rubidium gas cell frequency reference.

DATA PREPROCESSING

Preprocessing of the measurement data is often necessary before performing the actual analysis, which may require data averaging, or removal of outliers, frequency offset, and drift [4].

Phase data can be modified for a longer averaging time by simply removing the intermediate points. This is accomplished for frequency data by calculating the average of each group of points.

Frequency offset may be removed from phase data by subtracting a line determined by the average of the first differences, or by a least-squares linear fit. An offset may be removed from frequency data by normalizing it to have an average value of zero.

Frequency drift may be removed from phase data by a least-squares or 3-point quadratic fit [5], or by subtracting the average of the second differences. Frequency drift may be removed from frequency data by subtracting a least-squares linear fit, by subtracting a line determined by the mean of the first-differences of the frequency data, or by calculating the drift from the difference between the average of the two halves of the data. The latter, called here the bisection drift, is equivalent to the 3-point fit to the phase data.

Other, more specialized, methods of drift removal may also be used. For example, the frequency data may be fit to a particular model such as the log function of MIL-O-55310B [6]. The latter is particularly useful to describe the stabilization of a frequency source. In general, the objective is to remove as much of the deterministic behavior as possible, obtaining random residuals for subsequent noise analysis.

Phase data may be converted to frequency data by calculating the first differences of the phase data and dividing by the measurement interval or averaging time. Frequency data may be converted to phase data by piecewise integration, using the averaging time as the integration interval.

FREQUENCY STABILITY MEASURES

The most common time-domain measures of frequency stability are as follows:

0-7803-2500-1/95 $4.00 © 1995 IEEE

1. The normal Allan variance, $\sigma_y^2(\tau)$ [7].
2. The normal Allan variance calculated from fully overlapping samples, $\sigma_y^2(\tau)$ [8].
3. The modified Allan variance, Mod $\sigma_y^2(\tau)$ [9].
4. The time variance, $\sigma_x^2(\tau)$ [10,11].

where the variances are functions of the averaging time, τ. They are often denoted as AVAR, MVAR and TVAR, respectively. These quantities are not affected (except possibly for reasons of numerical precision) by the nominal frequency offset. They are usually calculated after removal of frequency drift, and are expressed as their square roots, ADEV, MDEV and TDEV. Each can be calculated from either phase or frequency data, which give the same result.

These calculations are reasonably fast on a modern computer with a math coprocessor, except, perhaps, for the modified Allan variance (and the related time variance). For a data set of several thousand points, calculation of the normal Allan variance is practically instantaneous, and the overlapping Allan variance calculation takes only a few seconds. Calculation of the modified Allan variance from phase data also takes only seconds, while its calculation from frequency data requires a triply-nested loop that can take several minutes. Algorithms faster than the obvious ones are available [12,13], particularly if the data have no gaps. A complete stability run is commonly done over a range of averaging times by doubling the τ at each successive analysis point for which there is sufficient data.

TEST DATA

A "classic" suite of frequency stability test data is the set of nine 3-digit numbers from Annex 8.E of NBS Monograph 140 [14] shown in Table I. Those numbers were used as an early example of an Allan variance calculation. This frequency data is also normalized to zero mean by subtracting the average value, and then integrated to obtain phase values. A listing of the properties of this data set is shown in Table II. While nine data points are insufficient to calculate large frequency averages. they are, nevertheless, a very useful starting point to verify frequency stability calculations, since a small data set can easily be entered and analyzed manually. A small data set is also an advantage for detecting "off-by-one" errors.

The larger frequency data test suite used here consists of 1000 pseudo-random frequency data points. It is produced by the following prime modulus linear congruential random number generator [15]:

$$n_{i+1} = (16807 \cdot n_i) \; Mod \; 2147483647$$

This expression produces a series of pseudo-random integers ranging in value from 1 to 2147483646 (the prime modulus, $2^{31} - 1$, avoids a collapse to zero). When started with the seed $n_0 = 1234567890$, it produces the sequence $n_1 = 395529916$, $n_2 = 1209410747$, $n_3 = 633705974$, etc. These numbers may be divided by 2147483647 to obtain a set of normalized floating-point test data ranging from 0 to 1. Thus the normalized value of n_0 is 0.5748904732. A spreadsheet program is a convenient and reasonably universal way to generate this data. The frequency data set may be converted to phase data by assuming an averaging time of 1, yielding a set of 1001 phase data points. Similarly, frequency offset and/or drift terms may be added to the data. These conversions can also be done by a spreadsheet program.

The values of this data set will be uniformly distributed between 0 and 1. While a data set with a normal (Gaussian) distribution would be more realistic, and could be produced by summing a number of independent uniformly distributed data sets, or by the Box-Muller method [16], this simple data set is adequate for software validation.

RESULTS

Statistics for the 1000-point test suite are shown in Table III. These values, reported to 7 significant figures, were obtained using IEEE 754 64-bit double-precision (15-digit) floating point arithmetic. While this reported precision is unwarranted for actual stability measures, it is useful for software validation. The theoretical expected value for the mean of a random variable uniformly distributed over the interval (0,1) is 0.5, and is independent of the averaging factor. The linear slope (per data interval) and the intercept are calculated as a least-squares linear regression fit. The standard deviation is that for the sample (not the population). The theoretical expected value for the standard deviation is $1/\sqrt{12} = 0.2886751$. The normal Allan deviation, $\sigma_y(\tau)$, is calculated for the full data set without averaging ($\tau = \tau_0$), using adjacent differences. For white frequency noise, it is equal to the standard deviation. The modified Allan deviation, Mod $\sigma_y(\tau)$, is, by definition, equal to the normal Allan deviation for an averaging time equal to that of the basic data interval, τ_0. The overlapping Allan deviation, calculated using fully overlapping samples (every available difference at a certain averaging time) is also equal to the normal Allan deviation for $\tau = \tau_0$. The time deviation, $\sigma_x(\tau)$, is simply the modified Allan deviation multiplied by $\tau/\sqrt{3}$.

GAPS, JUMPS, and OUTLIERS

It is not uncommon to have gaps and outliers in a set of raw frequency stability data. Missing or erroneous data may occur due to power outages, equipment malfunctions, and interference. For long-term tests, it may be impossible or impractical to repeat the run, or to otherwise avoid such bad data points. Usually the rea-

son for the gap or outlier is known. It is particularly important to explain all phase discontinuities. Plotting the data will often show the bad points, which may have to be removed before further analysis to obtain meaningful results.

A simple, yet effective, technique for finding outliers is to compare each frequency data point, y_i, with the median value, m, of the data set plus or minus some multiple of the median absolute deviation (MAD):

$$MAD = Median\{|y_i - m| / 0.6745\}$$

where $m = Median\{y_i\}$, and the factor 0.6745 makes the MAD equal to the standard deviation for normally distributed data [17]. These median statistics are more robust because they are insensitive to the size of the outliers. Outlier detection is normally applied only to frequency data.

More elaborate techniques exist for the recognition of outliers in marginal cases [17,18]. A particularly effective means is statistical comparison of each data point against an optimum predictor based on an appropriate noise model.

Often a bad data point is replaced with a gap. The gaps should be kept in the record because they serve as "place-keepers" in time, and because "truth-in-packaging" may require them to be identified. A value of zero is often used as an obvious and unique way to indicate a gap, especially in fractional frequency data, where zero almost never appears. It is also an easy value to test for. However, a value of zero does occur at the beginning and end of a set of normalized phase data, so, while a zero is suitable to indicate an embedded gap in phase data, it is not unique as the first or last point.

Stability analysis algorithms can be modified to handle gaps. Two sample variances can be formed for the available pairs, taking into account the actual number of analysis points. Averages can be formed where there is at least one point within the group. Otherwise a gap is inserted into the averaged data. Phase-frequency conversions can also be written to handle gaps.

Optionally, a gap may be replaced by an interpolated value. While this may be desirable in some cases, it masks the existence of the missing data, and creates a fictitious value. Filling in gaps has the advantage, however, that the plotting and stability analysis methods do not need provisions for handling gaps.

Frequency jumps can also be a problem for stability analysis. Intuitively, a frequency jump is an indication that the statistics of the device being measured are not stationary. It may be necessary to divide the record into two portions before and after the jump and analyze them separately. A jump may be defined and recognized by moving a sliding window through the frequency data, looking for a change in the average value between the two halves of the window. The magnitude of the jump can be judged in relation to the scatter of the data.

DATA ARRAYS

A one-column vector is all that is required for a phase or frequency data array. Because the data points are equally spaced, no time tags are necessary. While time tagging may be needed for archival storage of clock measurements, a vector of extracted gap-filled data is sufficient for analysis.

NUMERICAL PRECISION

There are relatively few numerical precision issues relating to the analysis of frequency stability data. One obvious case, however, is phase data for a highly stable frequency source having a relatively large frequency offset. The raw phase data will be essentially a straight line (representing the frequency offset), and the instability information is contained in the small deviations from the line. A large number of digits is necessary unless the frequency offset is removed by subtracting a linear term from the raw phase data. Similar considerations may also apply to the quadratic phase term (linear frequency drift).

Many frequency stability measures involve averages of first or second differences. Thus, while their numerical precision depends upon the variable digits of the data set, there is little error propagation in forming the summary statistics.

DATA PLOTTING

Data plotting is often the most important step in the analysis of frequency stability. Visual inspection can provide vital insight into the results, and is an important "preprocessor" before numerical analysis. A plot also shows much about the validity of a curve fit.

Phase data is generally plotted with line segments connecting the data points. This presentation properly conveys the integral nature of the phase data. Frequency data is often plotted the same way, simply because that is the way plotting is usually done. But a better presentation is a flat, horizontal line between the frequency data points. This shows the averaging time associated with the frequency measurement, and mimics the analog record from a frequency counter. As the density of the data points increases, there is essentially no difference between the two plotting methods. In a plot, missing data should be shown as a gap without a line connecting the adjacent points.

Stability plots generally take the form of graphs of log σ versus log τ, often with error bars to show the

precision of the results. The slope of the $\sigma_y(\tau)$ characteristic depends on the type of noise. It is customary to show points at octave increments of τ. These are equally spaced on the log scale, and are the result of successive averaging by a factor of two. Such a run usually ends when there are too few analysis points (say<7) for reasonable confidence. A run for all possible τ values, while slow, can provide valuable information since it is, in effect, a form of spectral analysis that can show power-law spectra and periodic instabilities such as environmental effects.

ERROR BARS

Several approaches exist for the setting of error bars on a stability plot to indicate the precision of the results. A rough approximation to a standard 1-sigma confidence interval can be obtained by simply dividing the point value by the square root of the number of analysis points. This approximation degenerates for a small number of degrees of freedom. More exact single- and double-sided confidence intervals can be determined by the noise and variance type [2,8,19,20], as shown in Table IV.

NOISE RECOGNITION AND MODELING

It is often desirable to have a means for identifying the type of noise that is being analyzed, and to be able to fit it to a power-law noise model. Noise recognition can generally be accomplished by using the B1 bias function (the ratio of the standard variance to the Allan variance) and the R(n) ratio function (the ratio of the modified Allan variance to the normal Allan variance). These functions can be calculated by the methods described in Reference 1. Power-law noise can be modeled over a range of averaging times by fitting a line to the results of an Allan variance stability analysis on a log σ versus log τ plot.

SIMULATED CLOCK NOISE

It is valuable to have a means of generating simulated power-law clock noise having the desired noise type (white phase, flicker phase, white frequency, flicker frequency, and random walk frequency), Allan deviation, frequency offset and frequency drift. This can serve as an additional means to validate stability analysis software, particularly for checking numerical precision and noise recognition and modeling. An excellent method for power-law noise generation is described in Reference 21. This reference also provides very useful analysis insights.

VALIDATION METHODS

Several methods are available to validate frequency stability analysis software:

1. Manual Analysis: The results obtained by manual analysis of small data sets (such as in NBS Monograph 140 Annex 8.E) can be compared with the new program output. This is always good to do to get a "feel" for the process.
2. Published Results: The results of a published analysis can be compared with the new program output. The purpose of this paper is to serve as such a reference.
3. Other Programs: The results obtained from other specialized stability analysis programs (such as that from a previous generation computer or operating system) can be compared with the new program output.
4. General Purpose Programs: The results obtained from industry standard, general purpose mathematical and spreadsheet programs (such as MathCAD and Lotus 1-2-3) can be compared with the new program output.
5. Consistency Checks: The new program should be verified for internal consistency, such as producing the same stability results from phase and frequency data. The standard and normal Allan variances should be approximately equal for white FM noise. The normal and modified Allan variances should be identical for an averaging factor of 1. For other averaging factors, the modified Allan variance should be approximately one-half the normal Allan variance for white FM noise and $\tau >> \tau_0$. The normal and overlapping Allan variances should be approximately equal. The overlapping method provides better confidence of the stability estimates. The various methods of drift removal should yield similar results.
6. Simulated Data: Simulated clock data can also serve as a useful cross check. Known values of frequency offset and drift can be inserted, analyzed, and removed. Known values of power-law noise can be generated, analyzed, plotted, and modeled.

CONCLUSIONS

There is a continuing need to validate the custom software used to analyze time domain frequency stability. The suite of test data presented here, along with the other suggestions made, can help ensure that correct results are obtained.

ACKNOWLEDGMENTS

The author wishes to thank Messrs. David W. Allan, Leonard S. Cutler, Samuel R. Stein, and Fred L. Walls for their help in checking the results presented here. Any remaining errors are my responsibility alone.

REFERENCES

1. D.B. Sullivan, D.W. Allan, D.A. Howe, and F.L. Walls (Editors), Characterization of Clocks and Oscillators, *NIST Technical Note 1337*, U.S. Department of Commerce, National Institute of Standards and Technology, March 1990.
2. S.R. Stein, "Frequency and Time - Their Measurement and Characterization," Chapter 12, pp.191-416, *Precision Frequency Control*, Vol. 2, Edited by E.A. Gerber and A. Ballato, Academic Press, New York, 1985.
3. "IEEE Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology," *IEEE Std. 1139-1988.*
4. J.A. Barnes, "The Measurement of Linear Frequency Drift in Oscillators," *Proc. 15th Annu. PTTI Meeting*, pp. 551-582, Dec. 1983.
5. M.A. Weiss and C. Hackman, "Confidence on the Three-Point Estimator of Frequency Drift," *Proc. 24th Annu. PTTI Meeting*, pp. 451-460, Dec. 1992.
6. MIL-O-55310B, Oscillators, Crystal, General Specification For.
7. D.W. Allan, "The Statistics of Atomic Frequency Standards," *Proc. IEEE*, Vol. 54, No. 2, pp. 221-230, Feb. 1966.
8. D.A. Howe, D.W. Allan, and J.A. Barnes, "Properties of Signal Sources and Measurement Methods," *Proc. 35th Annu. Symp. on Freq. Contrl.*, pp. 1-47, May 1981.
9. D.W. Allan and J.A. Barnes, "A Modified Allan Variance with Increased Oscillator Characterization Ability," *Proc. 35th Annu. Symp. on Freq. Contrl.*, pp. 470-474, May 1981.
10. D.W. Allan, D.D. Davis, J. Levine, M.A. Weiss, N. Hironaka, and D. Okayama, "New Inexpensive Frequency Calibration Service From NIST," *Proc. 44th Annu. Symp. on Freq. Contrl.*, pp. 107-116, June 1990.
11. D.W. Allan, M.A. Weiss, and J.L. Jespersen, "A Frequency-Domain View of Time-Domain Characterization of Clocks and Time and Frequency Distribution Systems," *Proc. 45th Annu. Symp. on Freq. Contrl.*, pp. 667-678, May 1991.
12. D.W. Allan, "Time and Frequency Metrology: Current Status and Future Considerations," *Proc. 5th European Freq. and Time Forum*, pp. 1-9, March 1991.
13. C.A. Greenhall, "A Shortcut for Computing the Modified Allan Variance," *Proc. 1992 IEEE Freq. Contrl. Symp.*, pp. 262-264, May 1992.
14. B.E. Blair (Editor), Time and Frequency: Theory and Fundamentals, *NBS Monograph 140*, Annex 8.E, p. 181, May 1974.
15. S.K. Park and K.W. Miller, "Random Number Generators: Good Ones are Hard to Find," *Comm. ACM*, Vol. 31, No. 10, pp. 1192-1201, Oct. 1988.
16. W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes in C*, Cambridge Univ. Press, Cambridge, U.K., 1988, pp. 216-217.
17. D.B. Percival, "Use of Robust Statistical Techniques in Time Scale Formation," *Preliminary Report*, U.S. Naval Observatory Contract No. N70092-82-M-0579, 1982.
18. Gernot M.R. Winkler, "Introduction to Robust Statistics and Data Filtering," *Tutorial at 1993 IEEE Freq. Contrl. Symp.*, Sessions 3D and 4D, June 1, 1993.
19. C.A. Greenhall, "Recipes for Degrees of Freedom of Frequency Stability Estimators," *IEEE Trans. Instrum. Meas.*, Vol. 40, No. 6, pp. 994-999, Dec. 1991.
20. C.A. Greenhall, "Estimating the Modified Allan Variance," *Proc. IEEE 1995 Freq. Contrl. Symp.*, (to be published).
21. N.J. Kasdin and T. Walter, "Discrete Simulation of Power Law Noise," *Proc. 1992 IEEE Freq. Contrl. Symp.*, pp. 274-283, May 1992.

Table I. NBS Monograph 140, Annex 8.E Test Data			
Data Point #	Frequency Data	Normalized Frequency	Phase Data (τ=1)
1	892	103.11111	0.00000
2	809	20.11111	103.11111
3	823	34.11111	123.22222
4	798	9.11111	157.33333
5	671	-117.88889	166.44444
6	644	-144.88889	48.55555
7	883	94.11111	-96.33333
8	903	114.11111	-2.22222
9	677	-111.88889	111.88889
10			0.00000

Table II. NBS Monograph 140, Annex 8.E Test Data Statistics		
Averaging Factor	1	2
# Data Points	9	4
Maximum	903	893.0
Minimum	644	657.5
Average	788.8889	802.875
Median	809	830.5
Linear Slope	- 10.20000	-2.55
Intercept	839.8889	809.25
Standard Deviation [1]	100.9770	102.6039
Normal Allan Deviation	91.22945	115.8082
Overlapping Allan Deviation	91.22945	85.95287
Modified Allan Deviation	91.22945	74.78849
Time Deviation	52.67135	86.35831

Table II Notes:

[1] Sample (not population) standard deviation.

Table III. 1000-Point Frequency Data Set				
Averaging Factor		1	10	100
# Data Points		1000	100	10
Maximum		9.957453e-01	7.003371e-01	5.489368e-01
Minimum		1.371760e-03	2.545924e-01	4.533354e-01
Average [1]		4.897745e-01	4.897745e-01	4.897745e-01
Median		4.798849e-01	5.047888e-01	4.807261e-01
Linear Slope [2,3]		6.490910e-06	5.979804e-05	1.056376e-03
Intercept [3]		4.865258e-01	4.867547e-01	4.839644e-01
Bisection Slope [2]		-6.104214e-06	-6.104214e-05	-6.104214e-04
1st Difference Slope [2]		1.517561e-04	9.648320e-04	1.011791e-03
Log Fit [2,4] $y(t) = a\ln(bt+1)+c$ $y'(t) = ab/(bt+1)$	a	5.577220e-03	5.248477e-03	7.138988e-03
	b	9.737500e-01	4.594973e+00	1.420429e+02
	c	4.570469e-01	4.631172e-01	4.442759e-01
	Slope at end	5.571498e-06	5.237080e-05	7.133666e-04
Standard Deviation [5]		2.884664e-01	9.296352e-02	3.206656e-02
Normal Allan Deviation [6]		2.922319e-01	9.965736e-02	3.897804e-02
Overlapping Allan Deviation [8]		2.922319e-01	9.159953e-02	3.241343e-02
Modified Allan Deviation [7,8]		2.922319e-01	6.172376e-02	3.897804e-02
Time Deviation [8]		1.687202e-01	3.563623e-01	1.253382e-02

Table III Notes:

[1] Expected value = 0.5.

[2] All slopes are per interval.

[3] Least-squares linear fit.

[4] Exact results will depend on iterative algorithm used. Data not suited to log fit.

[5] Sample (not population) standard deviation. Expected value = $1/\sqrt{12}$ = 0.2886751.

[6] Expected value equal to standard deviation for white FM noise.

[7] Equal to normal Allan deviation for averaging factor = 1.

[8] Calculated with listed averaging factors from the basic $\tau = 1$ data set.

Table IV. Error Bars for n=1000 Point =1 Data Set with Averaging Factor=10

Allan Deviation			Noise		Confidence Interval		
Type	#	Value	Ratio	Type	Value	χ^2	Remarks
Normal	99	9.965736e-02	B1=0.870	W FM [1]	CI= 8.713870e-03 [2,3]	-	$\pm 1\sigma$ CI. [6]
Overlapping	981	9.159953e-02	-	W FM # χ^2df = 146.177	Max $\sigma_y(\tau)$= 1.014923e-01	119.07	95% CI. [7]
					Max $\sigma_y(\tau)$= 1.035201e-01	114.45	95% CI. [8]
					Min $\sigma_y(\tau)$= 8.223942e-02	181.34	
Modified [4]	972	6.172376e-02	R(n)=0.384	W FM [5] # χ^2df = 94.620	Max $\sigma_y(\tau)$=7.044412e-02	72.64	95% CI. [7]
					Max $\sigma_y(\tau)$=7.224944e-02	69.06	95% CI. [8]
					Min $\sigma_y(\tau)$ = 5.419961e-02	122.71	

Table IV Notes:

[1] Theoretical B1=1.000 for W FM noise and 0.667 for F and W PM noise.

[2] Simple, noise-independent CI estimate = $\sigma_y(\tau)/\sqrt{N}$ = 1.001594e-02.

[3] This CI includes κ_α factor that depends on noise type:

Noise	α	κ_α
W PM	2	0.99
F PM	1	0.99
W FM	0	0.87
F FM	-1	0.77
RW FM	-2	0.75

[4] BW factor = $2\pi f_h \tau_0 = 10$. Applies only to F PM noise.

[5] Theoretical R(n) for W FM noise = 0.500 and 0.262 for F PM noise.

[6] Double-sided 68.3% confidence interval: p = 0.158 and 0.842.

[7] Single-sided 95% confidence interval: p = 0.950.

[8] Double-sided 95% confidence interval: p = 0.025 and 0.975.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

MEASUREMENT OF VOLTAGE NOISE IN CHEMICAL BATTERIES

Chadwick K. Boggs, Alan D. Doak, F. L. Walls
National Institute of Standards and Technology
325 Broadway Boulder, CO 80303

ABSTRACT

Ultra low noise voltage sources are often required in measurement systems and other applications. Common voltage regulators have performed inadequately in some applications. As an alternative, battery cells have been used. Of the various types, Hg cells have been credited with the best performance. However, actual values for the voltage noise in batteries have not, to our knowledge, been reported. In this paper a measurement system capable of measuring voltage noise below -200 dBV/Hz is discussed and its ability to characterize experimentally high performance voltage references is explored. The results of such measurements on common batteries are presented, and potential applications are considered.

I. Introduction

Chemical batteries have often been used when power supply noise is a significant problem. It is widely known that chemical batteries provide noise performance that is superior to common regulators and regulated power supplies. It is not, however, widely known exactly what performance can be realized with batteries, and what performance differences exist between battery types. Such knowledge is significant in some applications and yields insight into the fundamental noise processes present in batteries. In this paper, chemical battery noise performance is explored and measurements are presented. The voltage noise and current noise measurement techniques developed in these experiments have very high resolution and may be applied in other similar measurement systems.

II. Fundamental Noise

Johnson Noise

The fundamental source of voltage noise across any resistor is Johnson noise, given by [1,2]*

$$V_{noise} = \sqrt{4hfR\Delta\nu/[\exp(hf/kT)-1]}, \quad (1)$$

where k = Boltzmann's constant, T = temperature in kelvin, R = resistance in ohms, h is Planck's constant, f is frequency in Hertz, and $\Delta\nu$ is the measurement bandwidth.

At low frequencies Johnson noise can be approximated by

$$V_{noise} \cong \sqrt{4kTR\Delta\nu} \quad (2)$$

This fundamental noise process is dependent only upon temperature and resistance.

Shot Noise

Broadband current noise is found when two conditions are satisfied: (1) the charge carriers are quantized; and (2) the arrival times are random. These conditions cause current fluctuations of magnitude

$$i_{shot} = \sqrt{2qI\Delta\nu}, \quad (3)$$

where q is the charge quantization (in this case the charge of an electron) in coulombs, I is the bias current in amperes, and $\Delta\nu$ is the measurement bandwidth in Hertz. Current noise of this type is commonly called shot noise.

III. Voltage Noise

Voltage Noise Measurement System

To see voltage fluctuations in batteries, a measurement system with a resolution of approximately -200 dBV/Hz is required. This was accomplished by the use of two low noise

U.S. Government work not protected by U.S. copyright

amplifiers and a two channel cross-correlation technique[3], shown in Fig. 1. The two independent and similar amplifiers provided ~50 dB of gain with a noise floor of -180 dBV/Hz. The amplifiers boosted the noise from the batteries to approximately -150 dBV/Hz, which was just above the noise floor of the Fast Fourier Transform (FFT) spectrum analyzer used. The output noise from the amplifier was, however, -130 dBV/Hz which obscured the noise signal from the batteries.

To recover the battery signal, cross-correlation was performed between the two channels. The cross-correlation operation caused the noise from the independent amplifiers to average to zero as $1/\sqrt{n}$ (n = number of averages) revealing the underlying correlated voltage noise from the batteries. Two batteries were used to double the battery noise power and permit the use of high gain DC coupled amplifiers. The limiting factor in this technique was the time taken to average the data. A practical resolution of -206 dBV/Hz at 1 kHz was reached after 100,000 averages. This noise floor was sufficient to measure the voltage noise of batteries. The effect of cross-correlation in this system is described by

$$\overline{(V_1)_N \cdot (V_2)_N} = (V_{B1})_N \cdot (V_{B2})_N \cdot \frac{Sys1_N + Sys2}{\sqrt{N}} \quad (4)$$

The terms are as labeled in Fig. 1.

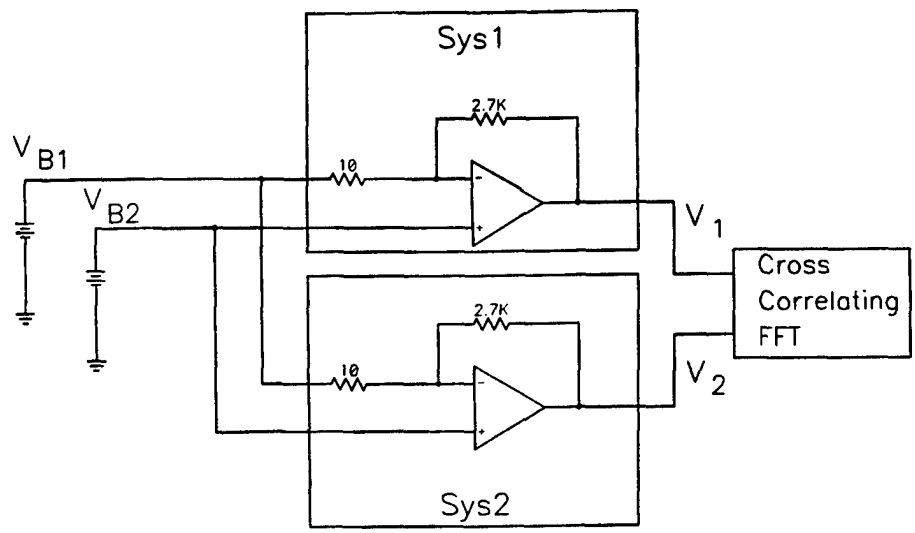


Figure 1: Voltage Noise Measurement System

Measurement of Chemical Battery Noise Voltage

Figs. 2 and 3 depict the results of measurement on various types of batteries. The top line in Fig. 3 shows the noise floor of the measurement system without cross-correlation, and the bottom line shows the ultimate noise floor of the system obtained through cross correlation.

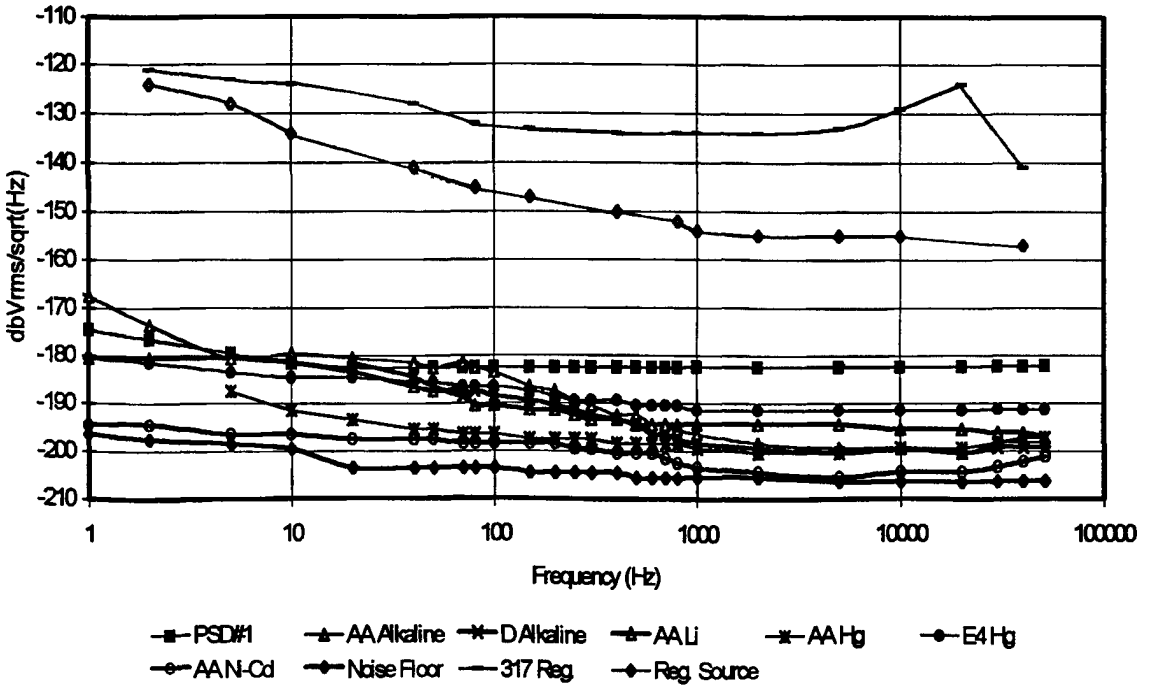


Figure 2: Voltage Noise Measurements

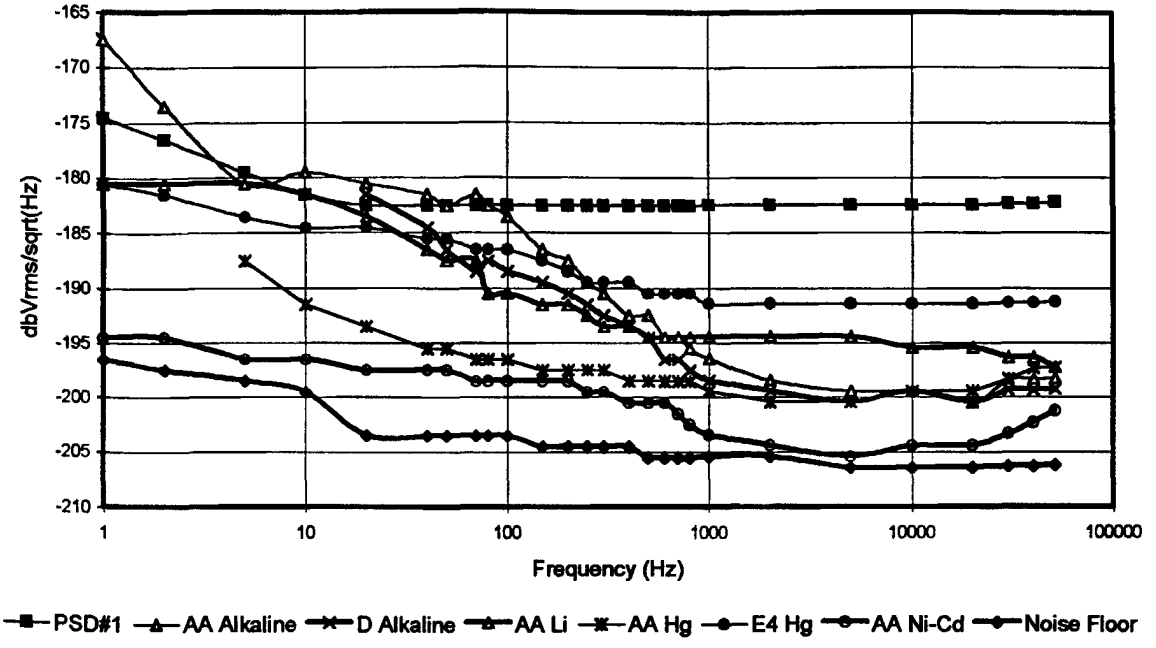


Figure 3: Battery Voltage Noise Measurements

Our results show that Ni-Cd batteries have the best noise performance of any of the chemical batteries tested. A value of -205 dBV/Hz at 10 kHz was recorded. In fact, measurements this close to the system noise floor were possibly inflated due to the addition of noise from the measurement system.

Different types and sizes of batteries produced different voltage noise. The battery's capacity influenced voltage noise. As can be seen from the graph, there is a slight difference between the noise in the AA and D size batteries. It was also found that the voltage noise of batteries was constant over the range of current from trickle charge to ~1mA discharge.

Noise Processes In Chemical Batteries

Batteries have very low internal resistance, generally less than 1 Ω. A value of 0.20 Ω was measured for Ni-Cd batteries, corresponding to a Johnson noise of -205 dBV/Hz at room temperature. As can be seen from Figs. 2 and 3, this agrees closely with the measured broad band voltage noise of the Ni-Cd batteries. This agreement suggests that the dominant broadband noise process in batteries is the Johnson noise from the internal resistance of the batteries.

This conclusion suggests that larger capacity batteries, which generally have smaller internal resistances, would produce a lower voltage noise. Voltage noise was found to be independent of bias current.

IV. Current Noise

The current noise of a battery is directly related to its voltage noise by Ohm's Law (see Fig. 4). In our measurements, battery voltage noise was found to be independent of bias current for typical current levels. Thus, the noise in a bias current from a battery is inversely proportional to the load impedance across the battery, $i = \frac{v}{Z}$. The current noise, then, should be known once the voltage noise is determined.

A battery can be modeled by the circuit in Fig. 4. The battery is represented by a dc voltage, V, an ac noise voltage, v, and an internal resistance, r. In normal applications, the battery's load, R, is much greater than the battery's internal resistance. The current noise in this model is $i = \frac{v}{r+R} \approx \frac{v}{R}$, R >> r. As discussed previously, it was found that v was approximately described by Eq. 2. This produces a current noise, i, that is small compared to shot noise.

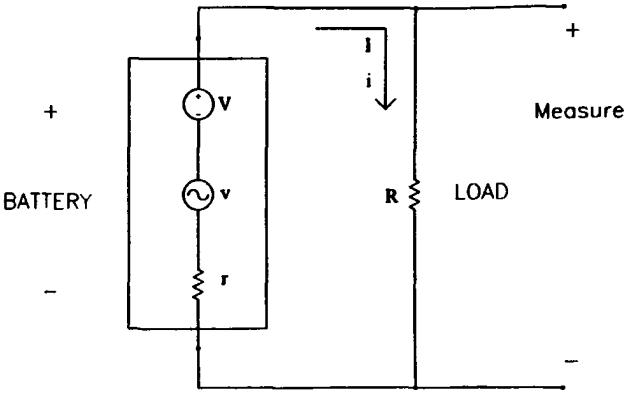


Figure 4: Low Frequency Chemical Batter Circuit Model

Batteries suppress shot noise by correlating the arrival times of charge carriers at the battery terminals. This can be shown by looking at an example of typical values in the model in Fig.4: V = 1.2 V, r = 0.2 Ω, R = 1000 Ω. If shot noise, i_{shot}, was dominant in the current noise, the result would be

$$v = i_{shot}R = R\sqrt{2q\frac{V}{R}\Delta\nu} = \sqrt{2qVR\Delta\nu} \quad (5)$$

For the example values, v = -154 dBV/Hz assuming shot noise. However, we measured $v \cong \sqrt{4kTr\Delta\nu} = -204 dBV/Hz$ which is nearly 50 dB below the shot noise in 1mA.

It can be shown that for any power supply consistent with the model in Fig. 4, and R >> r, the shot noise will induce a voltage

noise greater than the Johnson noise of r. Assuming shot noise is present, v is given by Eq. 5. If the voltage noise is dominated by the Johnson noise in r (note that the noise of r||R is roughly that of r), v is given by Eq. 2 as $v \cong \sqrt{4kTr\Delta v}$. Shot noise dominates whenever Eq. 5 is greater than Eq. 2, or

$$\sqrt{2qVR\Delta v} > \sqrt{4kTr\Delta v},$$
$$VR > \frac{2kTr}{q},$$
$$V > \frac{2kT}{q}\frac{r}{R},$$
$$V > 0,\ R >> r. \qquad (6)$$

In the above model shot noise would be expected to dominate over Johnson noise for source voltages greater than 52 mV at T = 300 K and r=R.

Current Noise Measurement System

To verify the model upon which the previous analyses were made, direct measurements of the current noise in batteries were made with the measurement system shown in Fig. 5. The transconductance of this circuit is

$$T = \frac{v_{out}}{i} = Z(K), \qquad (7)$$

where Z is the impedance of the circuit inductor, and K is the gain of the non-inverting amplifier.

An inductor was used to obtain high impedance with low Johnson noise. At high frequencies the input noise of this system was dominated by the Johnson noise of the 825 Ω resistor, yet the load impedance was on the order of 1 MΩ. A load impedance on the order of 1 MΩ produced a transconductance of 140 dBΩ. This means that the input current fluctuations were amplified by 140 dB and converted to voltage for measurement.

Input Noise of Measurement System

Measurements confirmed that the broadband input voltage noise of the system in Fig. 5 was white from 20 Hz to at least 100 kHz at -136 dBV/Hz. The broadband input current noise was

$$i = \frac{v_{out}}{T} = (-136 - 140)\ \text{dBA/Hz} \qquad (8)$$
$$= -276\ \text{dBA/Hz}$$
$$= 16\ \text{fA/Hz}$$

This agrees with the calculated Johnson noise from the 825 Ω resistor. At low frequencies, the input Johnson noise drops to that of the 203 Ω resistance in the inductor. However, at frequencies below 20 Hz the input noise becomes dominated by power law processes that raise it above that due to the Johnson noise of the input resistors.

To predict accurately the performance of the physical circuit we need to consider the inductive impedance more carefully. A real inductor of 1 H has large parasitic capacitance shown by the transfer function in Fig. 6. The impedance of our inductor was ideal to approximately 5 kHz. The parallel parasitic capacitance resonated at 13 kHz, and beyond 20 kHz the impedance dropped significantly from ideal.

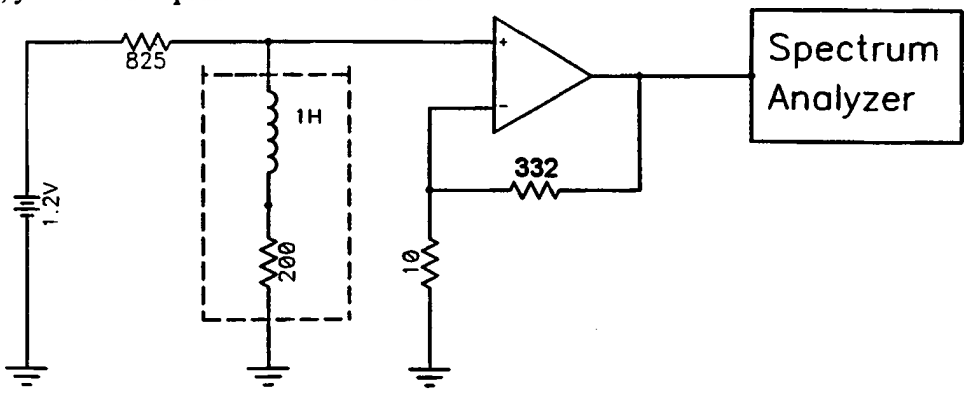


Figure 5: Current Noise Measurement System

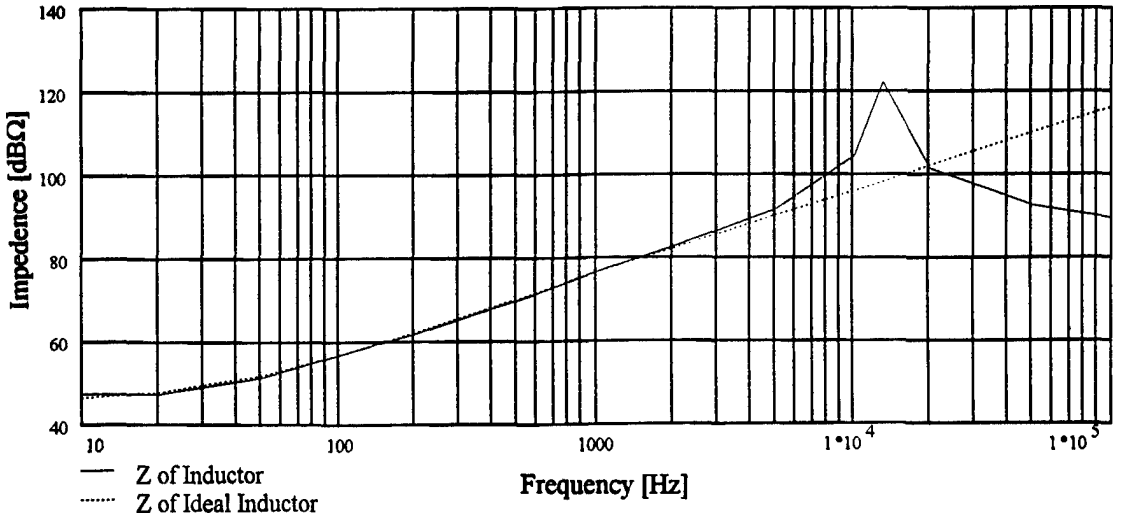


Figure 6: Effect of Parasitic Capacitance

Replacement of the battery in Fig. 5 with a high impedance current source made possible direct measurement of the system's broad band transconductance, from which the actual inductor impedance was calculated. The input current noise floor of the system is shown in Fig. 7. As a reference, shot noise for a 1 mA bias current is also plotted.

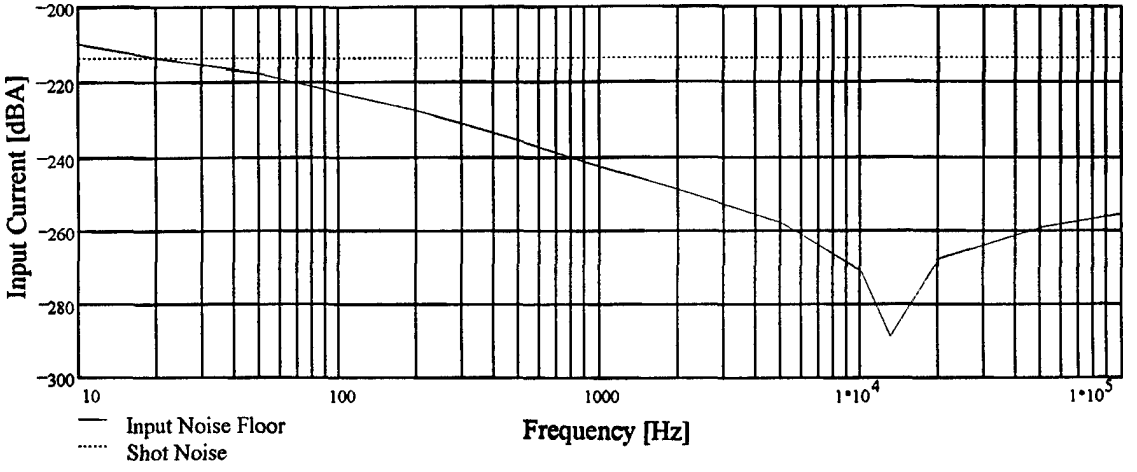


Figure 7: Measurement System Input Noise Floor

Measurement of Current Noise in Battery Bias Current

Measurement of chemical batteries in the system in Fig. 5 confirmed shot noise suppression. Data from typical battery measurements were identical to data from measurements of the system's input noise floor within the accuracy of the measurements. That is, a 1.2 V battery provided 1 mA of bias current with no noise to the resolution of our measurement system, which was over 40 dB below shot noise at resonance. The calculation of current noise from measured voltage noise suggested that the actual current noise was up to 95 dB below shot noise.

V. Conclusion

Very high resolution voltage measurement systems were constructed using two similar amplifiers and a cross-correlation method. This operation is built into many modern high end digital spectrum analyzers, and can resolve up to 25 dB. This approach, however, requires approximately 10^5averages.

Using this low noise measurement system we have characterized the voltage noise of 5 different battery types and compared them to a popular voltage regulator and a high performance power supply. We found that the voltage noise of the chemical batteries measured was many decades smaller than that of traditional power supplies. The lowest noise battery tested was a AA Ni-Cd with Vn = -205 dBV/Hz at 1 kHz. Different battery types exhibit vastly different noise voltage. In the batteries measured the broad band noise voltage appears to be approximately equal to the Johnson noise of their internal resistance.

We have made measurements of the current noise in battery driven bias currents. We show that chemical batteries correlate the arrival times of charge carriers at their terminals, thus suppressing current noise to well below shot noise. For the AA Ni-Cd batteries tested, the current noise in a 1 mA bias current was more than 50 dB below shot noise. The fundamental relationships between voltage and current noise were explored, and a simple model for circuit elements with noise was demonstrated.

The unique characteristics and superior noise performance of chemical batteries suggest interesting applications. A steady state arrangement may be possible in which the battery is not drained. The long term stability of other components might, through the charging circuitry, be able to enforce long term stability upon batteries, thus providing a low noise voltage reference that will last for years. In some applications such as biasing networks it may be possible to replace noisy capacitor/resistor chains with Ni-Cd batteries. The use of chemical batteries is already widespread in low noise applications and more precise knowledge of their performance will aid design and analysis.

Acknowledgments

We thank Hugh Robinson for the original suggestion to investigate the absence of shot noise in bias currents from chemical batteries. We also thank Franklin Ascarrunz for suggestions on low noise measurement circuits.

References

[1] H. Nyquist, Thermal Agitation of Electric Charge in Conductors, Phys. Rev. 32, pp. 110-113 (1928).

[2] J. B. Johnson, Thermal Agitation of Electricity in Conductors, Phys. Rev. 32, 97-109 (1928).

[3] F. L. Walls, S. R. Stein, J. E. Gray, and D. J. Glaze, "Design Considerations in State-of-the-Art Signal Processing and Phase Noise Measurement Systems", Proc. 30th Ann. SFC (1976), pp. 271

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

TRANSPOSED GAIN MICROWAVE OSCILLATORS WITH LOW RESIDUAL FLICKER NOISE

J.K.A. Everard and M.A. Page-Jones
Department of Electronics
University of York
Heslington, York, Y01-5DD, England

ABSTRACT

This paper describes the design of two low noise microwave oscillators operating at 7.6 GHz. These oscillators use room temperature Sapphire resonators operating in the $TE_{01\delta}$ mode which demonstrate unloaded Q's of 44,000. Silicon transposed gain amplifiers are used to produce near-flicker-free gain. Oscillators with flicker noise corners around 1 KHz and a phase noise of -131dBc/Hz at 10 KHz offset are demonstrated. This is within 5.5 dB of the theoretically predicted minimum noise. Further improvements of 30 dB are expected using related techniques.

INTRODUCTION

Microwave Oscillators are usually built using GaAs active components. These GaAs devices demonstrate high flicker noise corners which greatly degrade oscillator noise performance close to carrier. Silicon transistors have low flicker noise corners but also have low gain at microwave/mm wave frequencies. It is however fairly easy to fabricate silicon Schottky barrier mixer diodes operating at microwave frequencies.

This paper describes the design of two low noise 7.6 GHz oscillators with flicker noise corners of around 1KHz and a noise performance better than -131 dBc/Hz at 10 KHz offset. This noise is within 5.5 dB of the theoretical minimum for a Sapphire dielectric resonator oscillator with an unloaded Q_0 of 44,000 and a power available at the output of 1 mW. The noise changes at a rate of $1/\Delta f^2$ down to around 1KHz. These oscillators use the transposed gain of a low frequency silicon amplifier and demonstrate a suppression greater than 45 dB of the reference LO phase noise.

TRANSPOSED GAIN

Transposed gain is achieved using two mixers, a low frequency amplifier and a local oscillator as shown in Figure 1. This upconverts the LF gain to microwave frequencies on both sides of the reference LO. The frequency of the local oscillator (which can be noisy) is therefore used at a frequency close to that required for the gain. The noise from the local oscillator is significantly suppressed (>45 dB) in an oscillator using transposed gain.

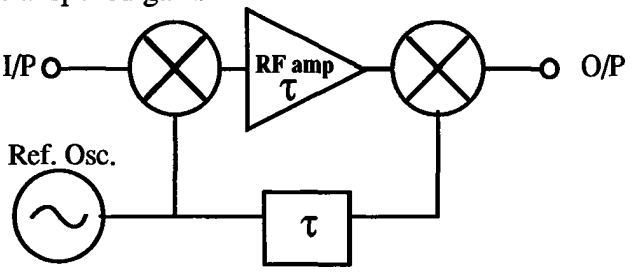


Figure 1. Transposed Gain Amplifier (TGA)

NOISE CONVERSION

It is important to ensure that the noise performance of the LO does not degrade either the AM or PM noise of a signal passing through the Transposed Gain Amplifier (TGA).

Phase noise on the local oscillator can cause an output phase modulation because of the group delay of the amplifier. This is corrected by introducing a delay line with the same delay (and dispersion) characteristics of the amplifier in between the LO drive of the two mixers. In these experiments the delay of the low frequency (inverting) amplifier is around 1.3 ns. The delay of the amplifier and mixers matches the delay line to less than 20 ps.

0-7803-2500-1/95 $4.00 © 1995 IEEE

AM on the reference oscillator has two effects on the output of the TGA. The first is to vary the mixer conversion losses, effectively transferring AM to the output. The second effect of LO drive variation is to alter the phase shift of the TGA. This is due to the mixer impedance changing with drive level. This is less serious with a broadband mixer. The LO should therefore have low AM by using an oscillator which incorporates 'hard limiting'.

OSCILLATORS

With the addition of a resonator a TGA can be made into a transposed gain oscillator (TGO) as shown in Figure 2. Because of the cancellation of FM and AM described above, such an oscillator is shown to have much lower phase noise than the original local oscillator.

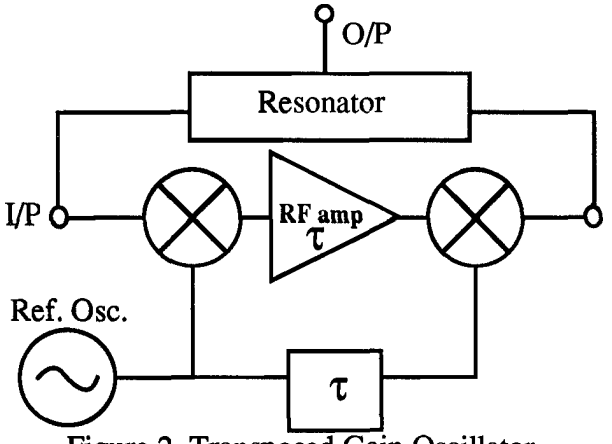


Figure 2, Transposed Gain Oscillator.

Theories have been developed which show how minimum noise can be obtained in oscillators operating in the thermal/additive noise regime[1,2,3,4]. These theories predict that for minimum additive noise the insertion loss of the resonator should be between 6 and 9 dB, in other words the ratio of loaded to unloaded Q should be between 1/2 to 2/3.

The oscillator phase noise should therefore roll off at $1/\Delta f^2$ rate if thermal noise predominates (20dB/decade). Close to carrier this becomes $1/\Delta f^3$ (30dB/decade of offset), as flicker noise is transposed into the oscillator. If the flicker noise corner occurs a few MHz from carrier (typical GaAs) this is much more serious than a few hundred Hz (typical LF silicon). At frequencies where this advantage exceeds the effects of a noise floor increase due to the mixer, then such a system could be advantageous.

The measured flicker noise corners of the amplifier and mixers were 30KHz and less than 3 KHz respectively. The flicker noise corner of the oscillator was around 1 KHz. It should be noted that although the baseband and transposed corners are related, their ratio depends on the transposition (modulation) mechanism. In these oscillators the amplifier is not non-linear because the limiting occurs in the output mixer. Both ports of the output mixer are driven at powers greater than +7dBm to obtain an output power of around 0dBm.

Systems have been produced (by this group[5]) for measuring the cross correlation between baseband gate and drain noise and the AM and PM components that these transpose onto a carrier. These measurements enable prediction of the equivalent circuit noise sources of the FET and the cross correlation coefficients between them.

The final system has a residual noise floor of -180 dBc for offsets greater than 3 KHz.

EXPERIMENTAL MEASUREMENTS

To evaluate the transposed gain method two Transposed Gain Oscillators have been built at 7.6 GHz operating 13 MHz apart. The phase noise of the beat signal was then measured using the phase detector method (in conjunction with the low noise, HP8662A, signal source) as shown in Figure 3.

The oscillators use room temperature Sapphire resonators operating in the fundamental $TE_{01}\delta$ mode. These resonators are operated substantially as a two port with a lightly coupled third port to provide a filtered output. The unloaded Q (including the O/P port) is measured to be 44,000. To obtain minimum noise performance using Thermal/additive noise the insertion loss of the resonator is set to 6 dB[1,2,3,4], producing a loaded Q of 22,000. The insertion loss of the mixers is around 7 dB therefore the open loop gain of the silicon

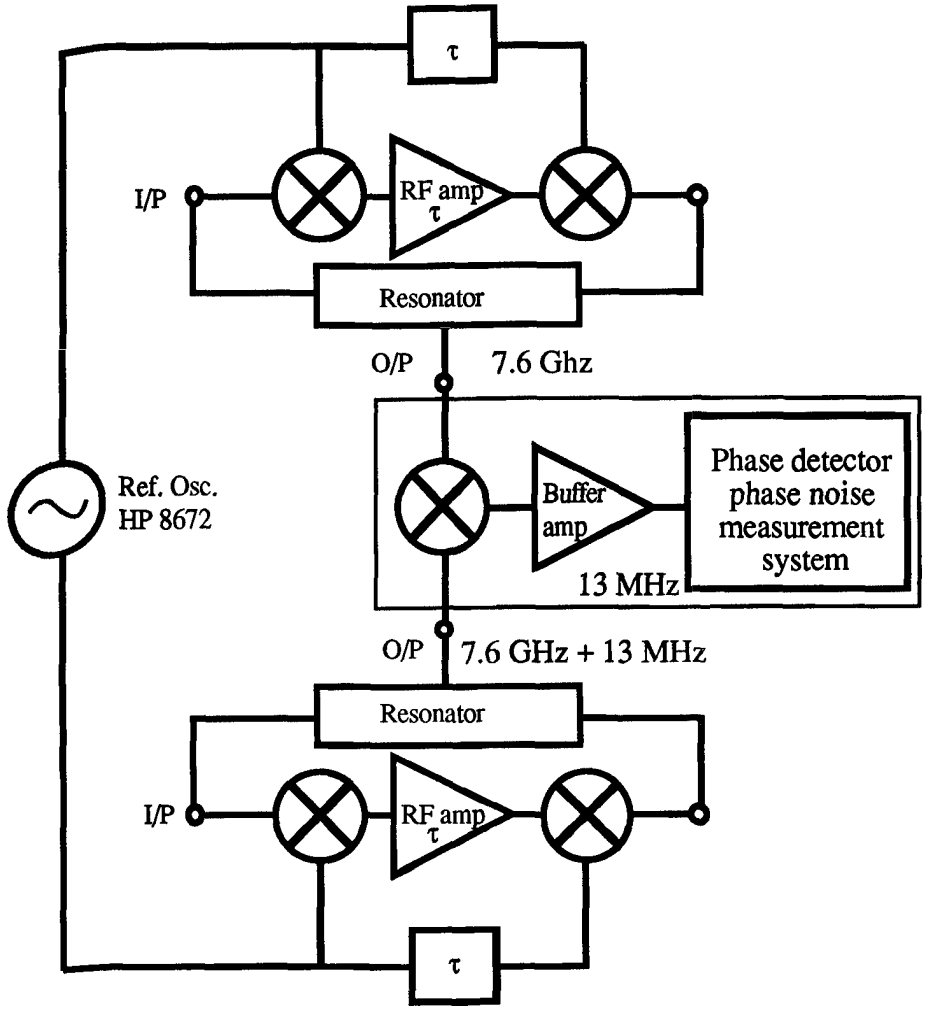


Figure 3. Experimental Oscillators and Phase Noise Measurement System

amplifier should exceed 20 dB. The total noise figure of the TGA which incorporated the two mixers and the low frequency amplifier was around 9.5 dB. The single sideband output power available from the O/P mixer was around 0.0dBm. The local oscillator used for each oscillator was an HP 8672A synthesised signal generator. A GaAs based oscillator using similar dielectric resonators could also have been used.

The noise performance is shown and tabulated in Figure 4. These measurements demonstrate a noise performance better than -131 dBc/Hz at 10 KHz offset (assuming that both oscillators are similar and uncorrelated, 128dBc/Hz -3dB) and a flicker noise corner of around 1 KHz. This noise is within 5.5 dB of the theoretical minimum for a Sapphire DR oscillator with an unloaded Q_0 of 44,000 and an available amplifier power of 1 milliwatt. The equation used to describe the noise performance L_{fm} under optimum operating conditions for thermal noise is [2,3,4].

$$L_{fm} = (2FkT/Q_0^2P)(f_0/\Delta F)^2$$

where F (9.5dB) is the noise figure under the correct operating conditions, k is Boltzmanns constant, Q_0 (44,000) is the unloaded Q, P (0dBm) is the power available at the output of the

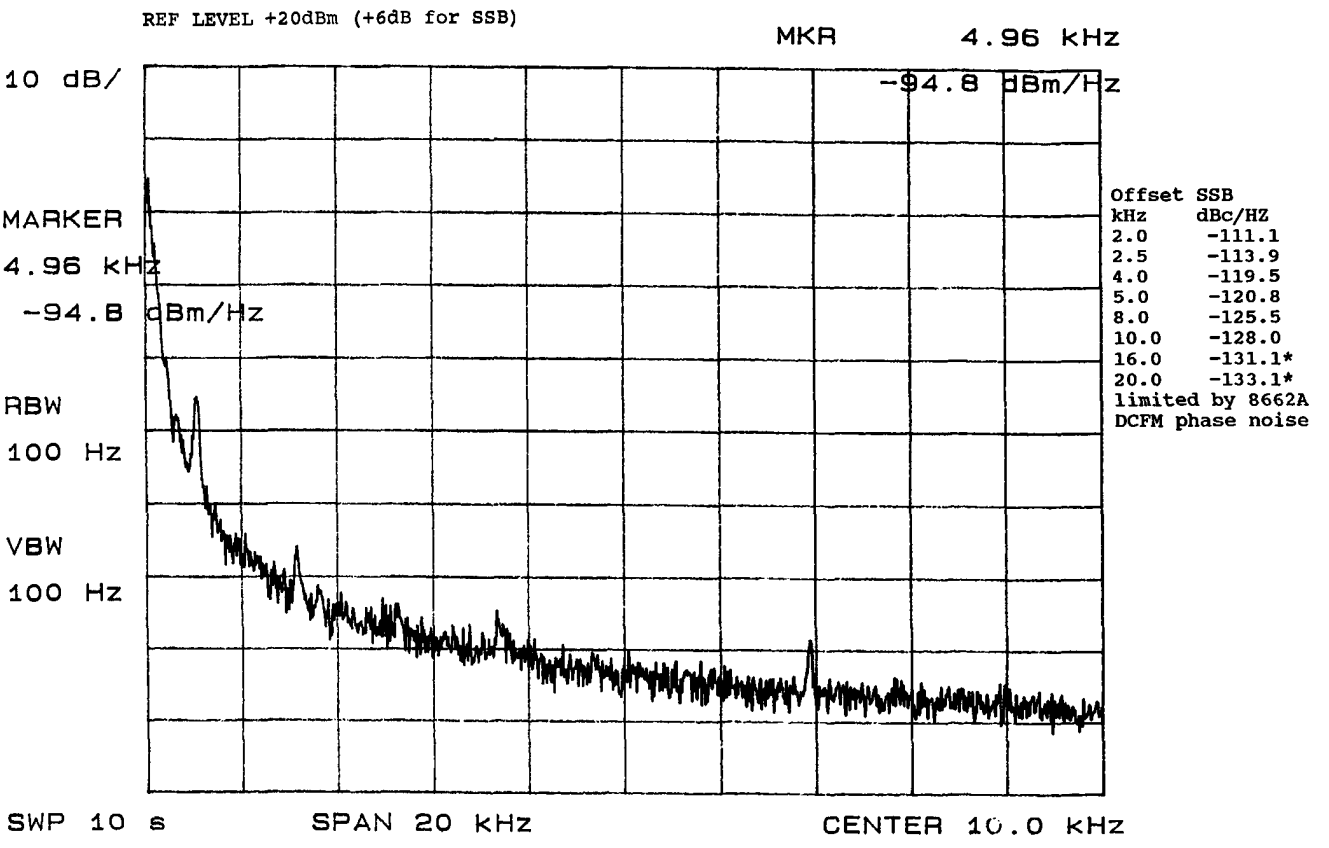


Figure 4. Phase Noise of the beat signal of
two Transposed Gain Oscillators operating around 7.6 GHz

TGA, f_0 is the centre frequency (7.6GHz) and Δf is the offset frequency. This equation incorporates the optimum operating condition that the ratio of $Q_L/Q_0 = 1/2$ and therefore the gain of the amplifier is 6 dB. The measured noise may be limited by the performance of the HP 8662A at 13 MHz (spec. -132dBc/Hz @ 10 KHz offset)

The typical noise performance of the HP8672A signal generator at 10 KHz offset is -85 dBc/Hz. The transposed gain oscillator is therefore suppressing this noise by more than 45 dB.

FUTURE WORK

Higher unloaded Q, Higher power operation and zero IF operation are currently under investigation. It is expected that a further noise improvement of >30 dB can be achieved using these methods.

REFERENCES

1. T.E. Parker, " Current Developments in SAW Oscillator Stability". 31st Annual Frequency Control Symposium 1977, pp.359-364.

2. J.K.A. Everard. "Low Noise Oscillators" IEEE Microwave Theory and Techniques Conference, Albuquerque New Mexico, 1-5th June 1992, pp.1077-1080.

3. Editors D. Haigh and J.K.A. Everard. "Gallium Arsenide Technology and its Impact on Circuits and Systems". Published by Peter Peregrinus Ltd., August 31st 1989. Chapter 8 (J.K.A. Everard) on Low noise oscillators, pp. 237-280.

4. J.K.A. Everard "Low Noise Power Efficient Oscillators: Theory and Design" IEE Proceedings Part G, Vol. 133, No.4, August 1986, pp. 172-180.

5. P.A. Dallas and J.K.A. Everard. "Measurement of the Cross Correlation between Baseband and Transposed Flicker Noises in a GaAs MESFET". IEEE Microwave Theory and Techniques Conference, Dallas USA, May 1990, pp. 1261-1264.

6. T.E. Parker, " Characteristics and sources of phase noise in stable oscillators". 41st Annual frequency control symposium 1987.pp.99-110.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Designing SAW Resonators and DRO Oscillators Using Nonlinear CAD Tools

Ulrich L. Rohde
Compact Software, Inc.
201 McLean Boulevard
Paterson, New Jersey 07504

Abstract - *In recent years, these types of high stability oscillators have been empirically designed. The key quality factors one seeks are both short- and long-term stability. When using CAD tools, the analysis of such oscillators has been done looking for the negative resistance approach indicating that there is potential for oscillation. Under large signal conditions, however, the semiconductor parameters change and the loading by the transistor imposes changes on the circuit, specifically the resonator, and therefore, the linearized assumptions will fail. By taking advantage of a) precise design modeling for time domain nonlinear equivalent circuits and its noise contribution, b) using the noise correlation matrix as a vehicle to put the noise sources together, and c) by using the modulation matrix calculation for close-in phase noise and the conversion matrix calculation for far-out noise, we gain accurate insight into the phase noise. By combining all of these techniques and applying optimization techniques, we can take circuits of arbitrary topologies at microwave frequencies which exhibit high Q and optimize them for performance. This method eliminates the time consuming trial and error approach and provides accurate results as will be demonstrated using simulation and published measured results.*

Introduction

SAW oscillators and dielectric resonator oscillators (DROs) have found intensive use in a number of microwave circuit-related applications. These types of oscillators are built to obtain very low phase noise. SAW oscillators are generally used up to 1 GHz and DROs used at frequencies of 12 GHz and above. It appears that the leading manufacturers/users of SAW oscillators are Westinghouse and Raytheon. Due to fairly low frequency applications, silicon bipolar transistors are used with an f_t for 5 GHz and below. Therefore, it is possible to build these oscillators using silicon MIMICs as offered through Avantek, Synergy Microwave and others.

DROs cover a much wider frequency range. There is a fairly wide variation in the phase noise versus complexity of circuitry. The insertion loss of the DRO depends on the properties of the resonator itself and its coupling through a microstrip circuit. There is, however, a trade-off between the operating Q and loop gain, which requires careful consideration in the configuration selection and the number of transistors used. In essence, two or more stage oscillators provide a higher gain bandwidth product and can compensate for its losses. There is a larger number of combinations one can use to build a DRO. Figure 1 shows the frequency stabilization methods [1].

0-7803-2500-1/95 $4.00 © 1995 IEEE

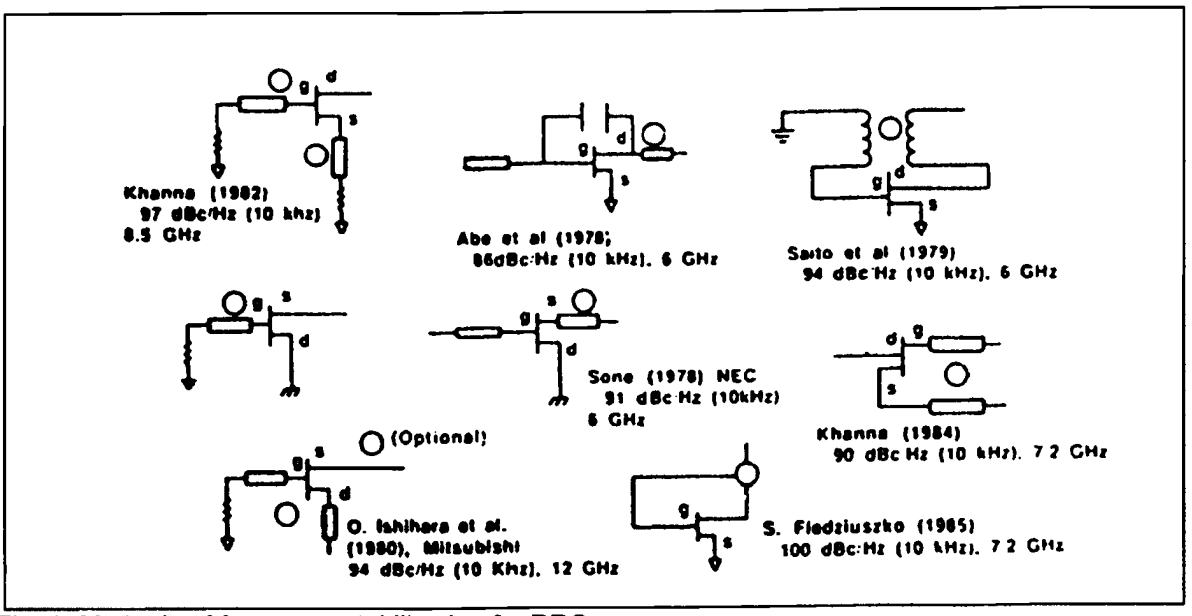


Fig. 1- Methods of frequency stabilization for DROs

Resonators/Models

The SAW resonator can be modeled as a either one- or two-port configuration. Figure 2 shows the electronic equivalent circuits [2].

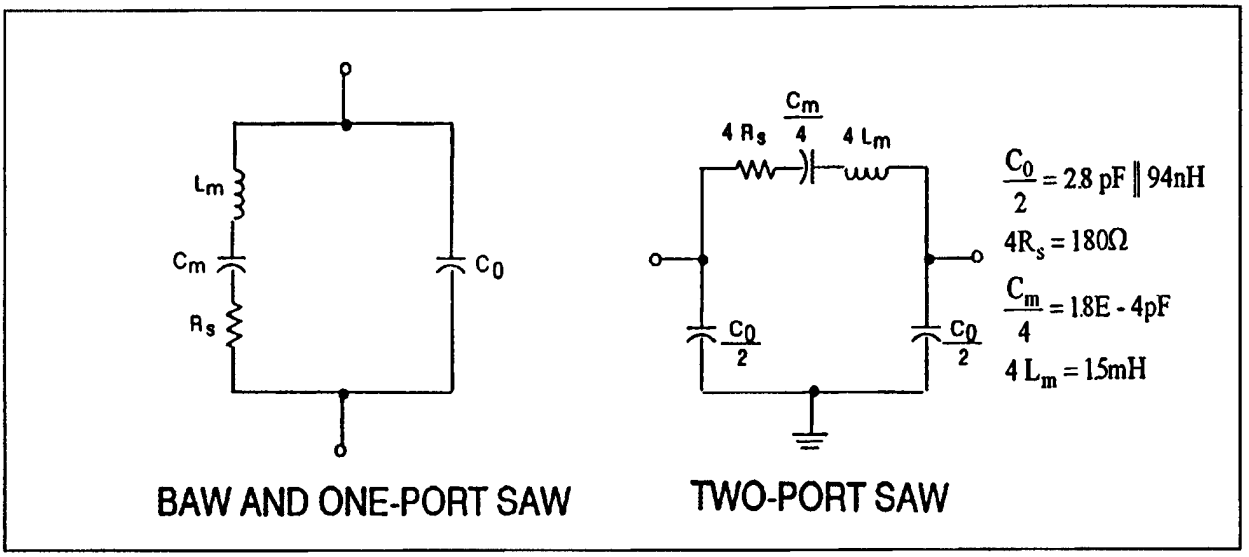


Fig. 2-Electronic equivalent circuits of a one- or two-port configuration for a one- or two-port SAW resonator (Equivalent circuit parameters supplied by T.E. Parker of Raytheon, now with NIST.)

DROs are described by their mechanical properties and are available as BandStop and BandPass structures. Figure 3a shows a sketch of the mechanical arrangements of the dielectric resonator and microstrip (DRMS) BandStop structure, Figure 3b shows the sketch of the DRMS BandPass structure and Figure 3c shows the field interaction between the microstrip and the resonator. [3-8].

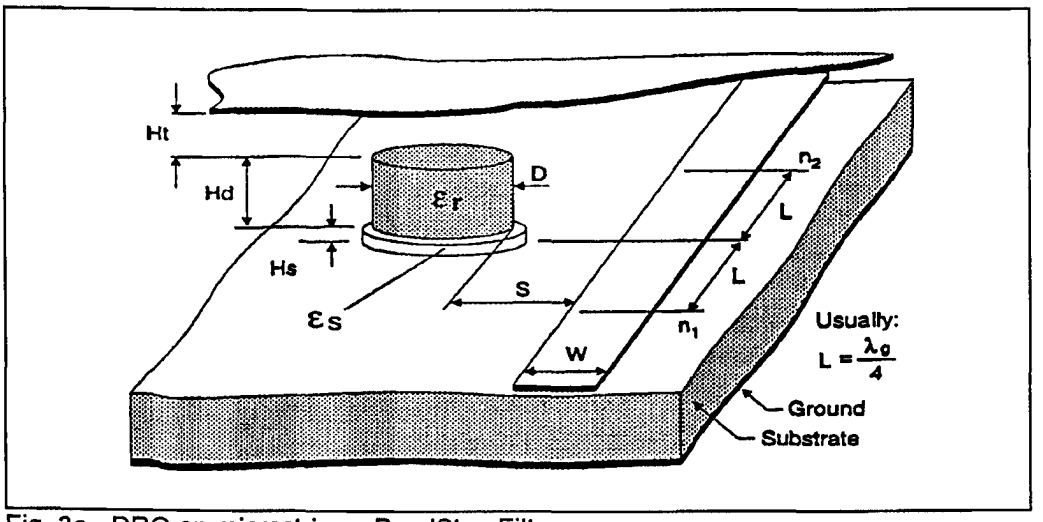


Fig. 3a - DRO on microstrip as BandStop Filter

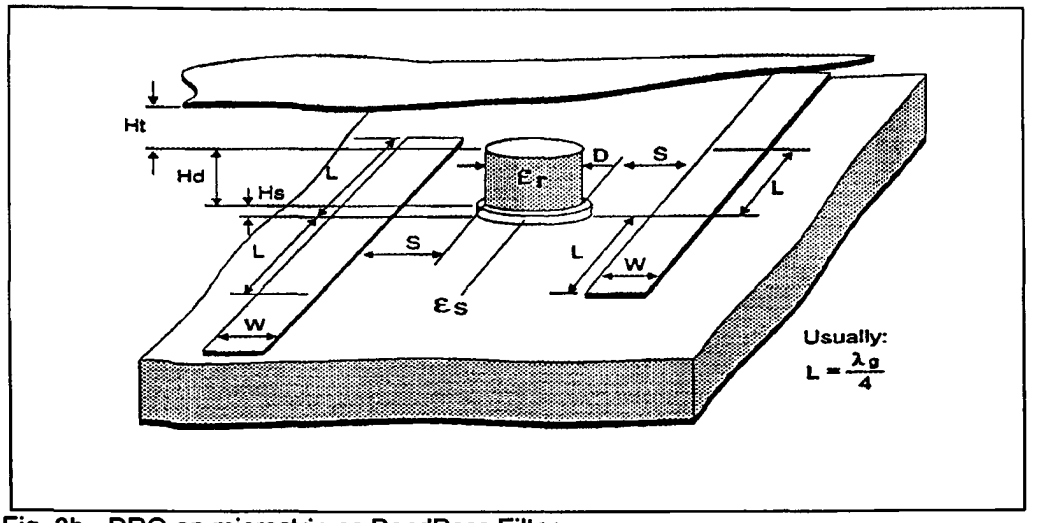


Fig. 3b - DRO on microstrip as BandPass Filter

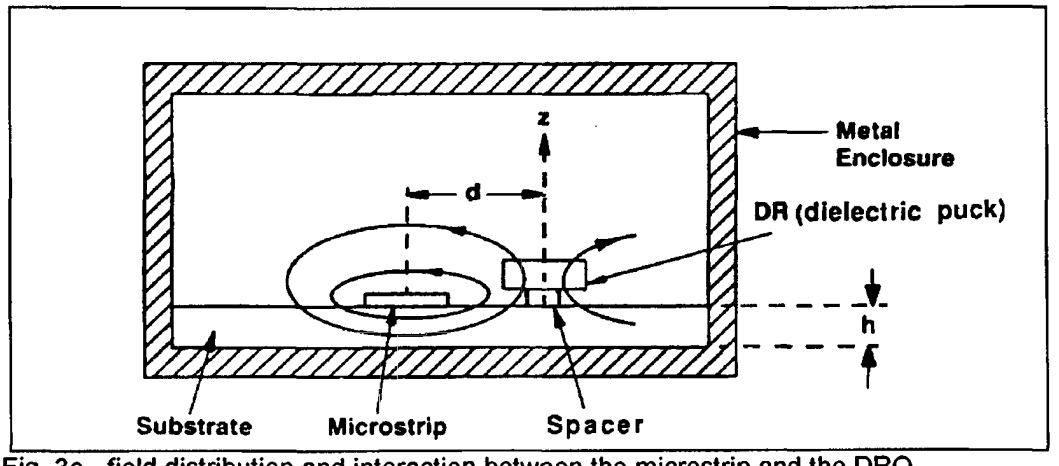


Fig. 3c - field distribution and interaction between the microstrip and the DRO

The actual Q is determined by the conductivity of the dielectric resonator (DR) and the coupling to the circuit, which is a function of the distance between the DR and the microstrip line. As previously indicated, the insertion loss of such a filter depends highly upon the coupling.

Figure 4 shows the frequency response of a SAW resonator.

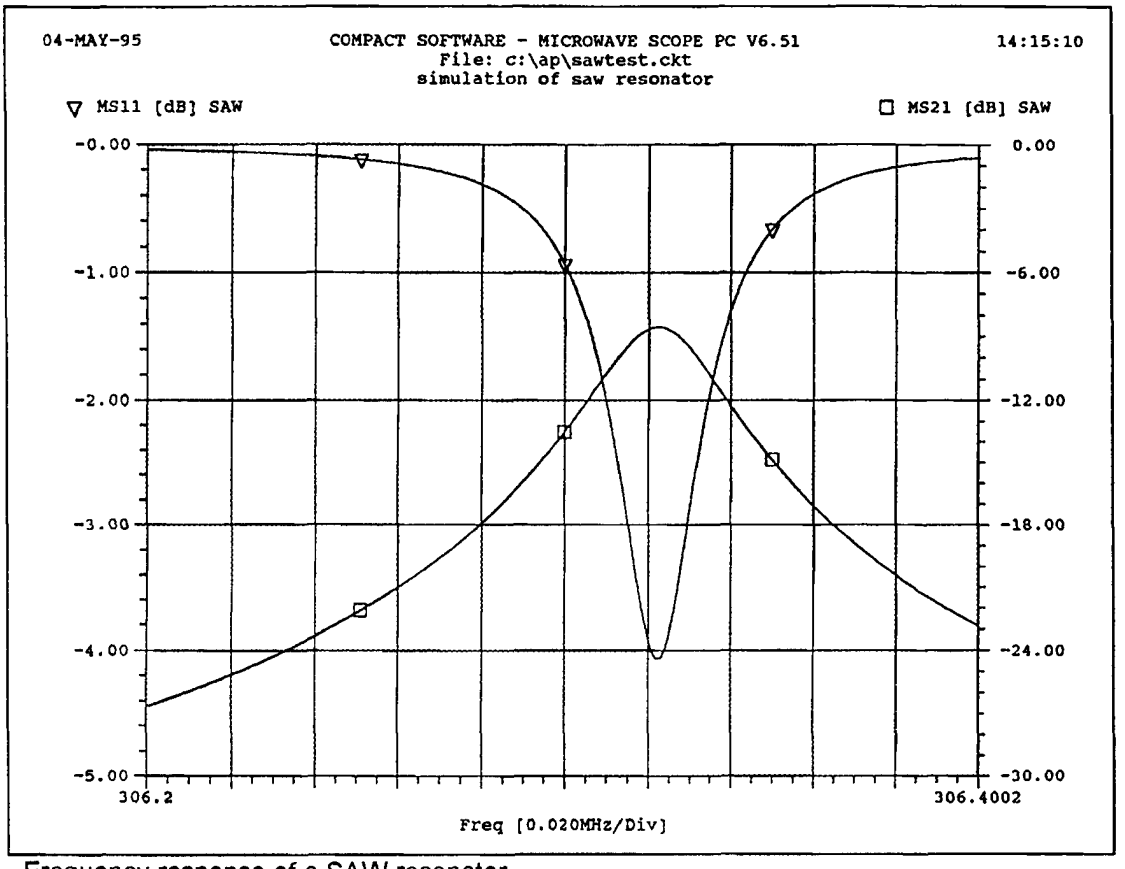


Fig. 4 - Frequency response of a SAW resonator

Table 1 shows the physical dimensions of the DRO used in Figure 5 in Super-Compact/Microwave Harmonica format.

TABLE 1

```
BLK
        DRMS 1  2 D=6.12e-3  HD-2.45e-3  ER=38  HT=1.5e-3  S=.5e-3;
        + W=1.1e-3 L=4e-3  SRD=1e-4 BPF SUB;
        trf  2 0  0 3  N=1
        pug: 2POR 1  3
END
DATA
        SUB: MS  er=2.4  h=0.380e-3 met1=cu  3.175e-6  tand=0.0001
END
```

Figure 5 shows the frequency response of a DR wherein one can also observe the "Q" as a function of coupling.

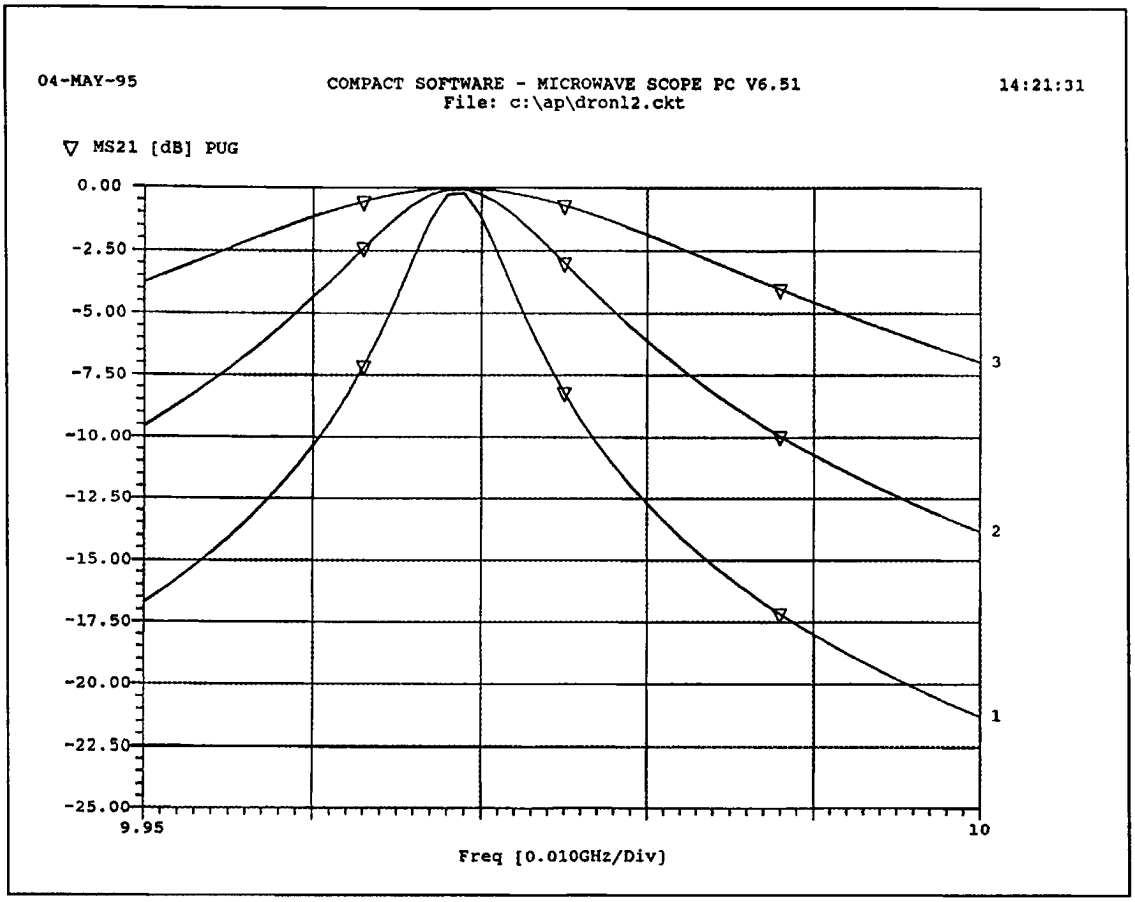


Fig. 5 - shows the frequency response and insertion loss of DRO as a function of the distance or spacing between the resonator and the microstrip. The three values are .5, 1.0 and 1.5 millimeters.

Selection of Active Devices

For applications below 1 GHz, either discrete silicon microwave transistors or silicon MIMICs can serve as an active device to provide the necessary loop gain. At frequencies above 4 or 5 GHz, both silicon transistors and heterojunction bipolar transistors have been used. At frequencies above 4 - 6 GHz, GaAs FETs are available. The major drawback in using GaAs FETs is the fairly high flicker frequency corner which varies between 10 and 100 MHz. Once commercially available heterojunction bipolar transistors are available, one can expect to see them replace GaAs FETs more frequently.

Basic Oscillator Configurations

VHF/microwave oscillators generally can be viewed as either two-terminal negative resistance generators or an oscillator seen as feed-forward amplifier with positive feedback. Figure 6a and 6b shows such examples.

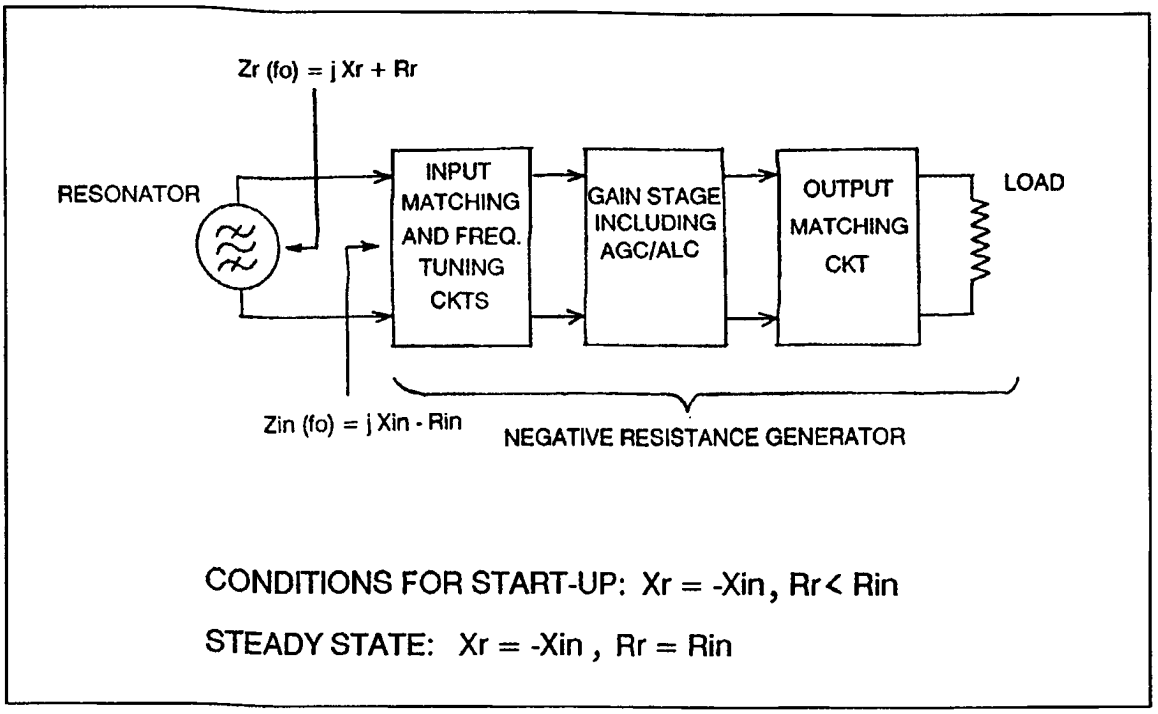


Fig. 6a - Oscillator viewed as a two terminal negative resistance generator

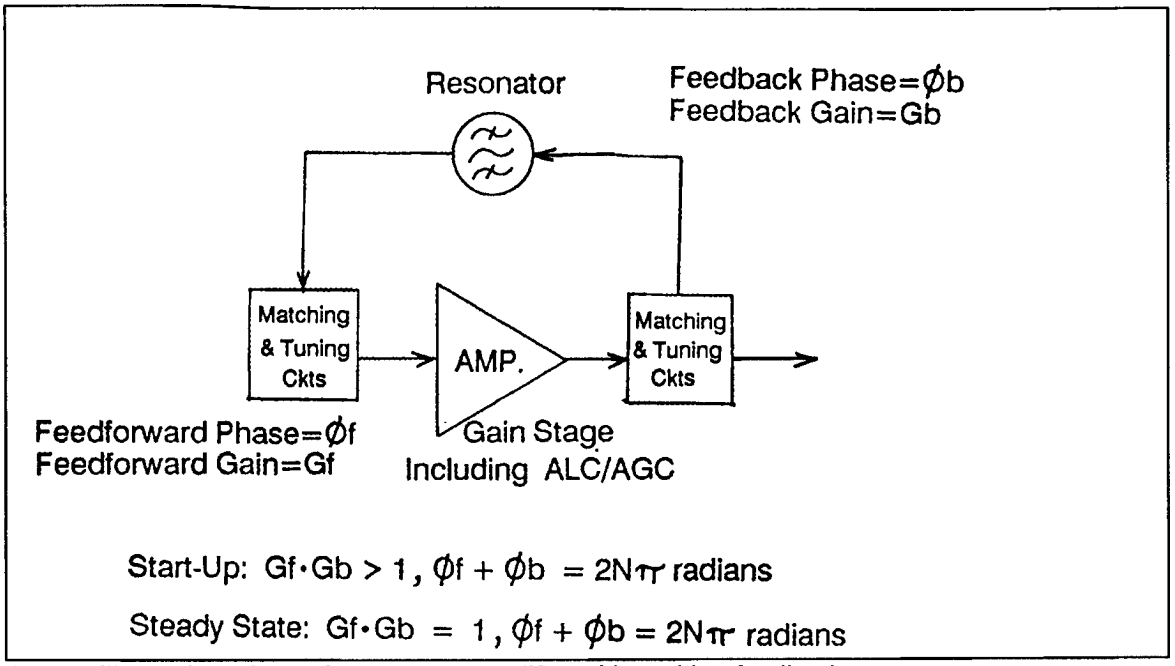


Fig. 6b - Oscillator viewed as a feedforward amplifier with positive feedback

In the case of a SAW oscillator, this translates into a circuit with a phase shifter. These circuits generally adjust properly to compensate for the internal phase shift of the transistor or transistor configurations operating at higher frequencies close to their cut-off frequencies.

Figure 7 shows the circuit configuration used for very low phase noise SAW oscillator, which also includes electric frequency controlled elements, which result in an SAW resonator-based UHF voltage controlled oscillator (VCO) [8]. The actual circuits are the result of many years of research and the purpose of modern CAD is to reduce this level of effort, as will be shown in this presentation.

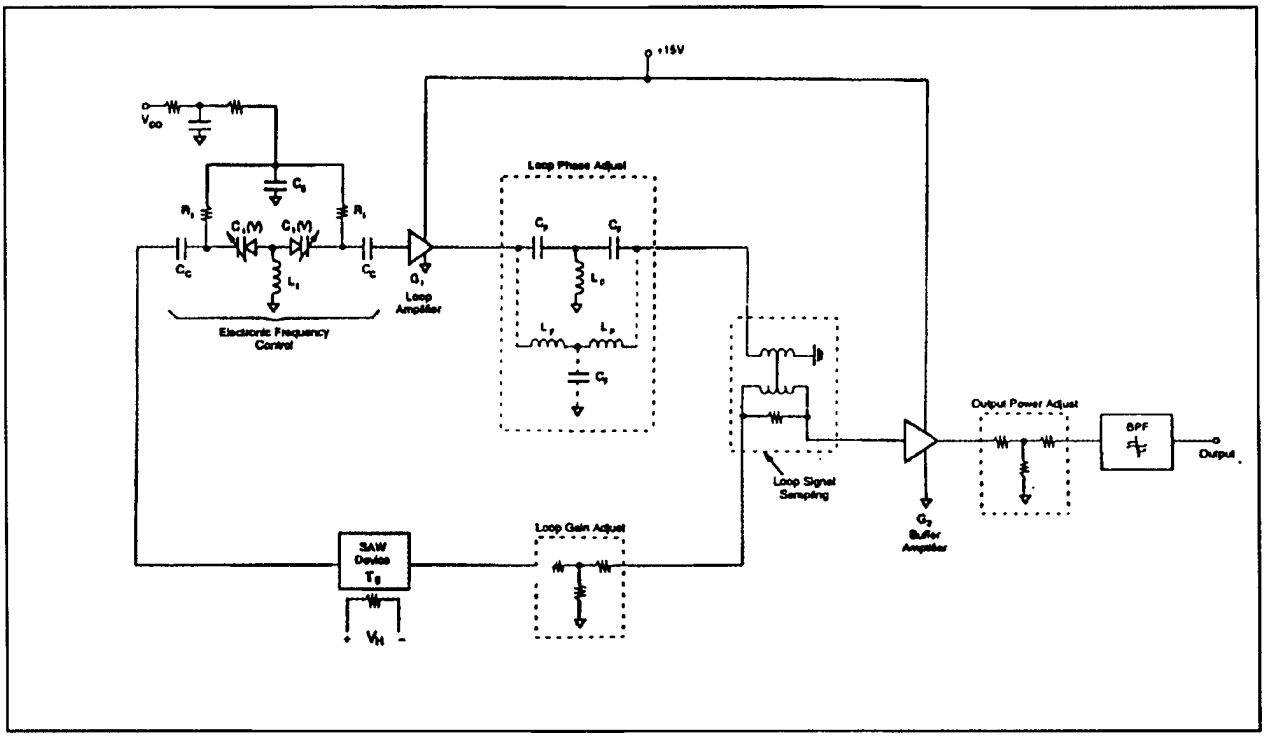


Fig. 7 - SAW resonator-based UHF voltage controlled oscillator

Phase Noise Contribution

In the case of very high "*Q*" circuits, the phase noise contribution can be expected from the active device and the resonator. In normal oscillator theory, the high "*Q*" circuit is considered noiseless or noise-free; however, because of the vibration in the mechanical resonator, unlike LC oscillators, it shows a noise contribution.

The active device has two important noise contributions. One is the flicker noise component and the other a wideband noise, which is also frequency dependent. As outlined in my *Design and Optimization of Low-Noise Oscillators using Nonlinear CAD Tools*, which was presented at the 1994 Frequency Control Symposium, the noise from the active device is either modulating (flicker noise) or up-converted (wideband noise) and appears as single sideband phase noise on top of the oscillator output. While both the flicker noise and the wideband noise are bias-dependent, there are certain contradictions between output power and the best close-in noise. In the past, researchers would try to maximize the output for best signal-to-noise ratio, which is generally applicable only for distance off the carrier, but the close-in phase noise suffers from high DC currents. [9]

Such bias dependency exists for both silicon transistors and members of the GaAs FET family. Generally, low noise operation requires smaller currents.

Wideband VCOs

Here is a quick reminder that at wide tuning ranges, the output noise is mainly determined by the noise contribution of the tuning diode. The noise voltage modulates the VCO. It is possible to define an equivalent noise $R_{\alpha eq}$ that, inserted in Nyquist's equation

$$V_n = \sqrt{4KT_0R\Delta f} \qquad (1)$$

where KT=4.2 x 10^{-21} at about 300°K, R is the equivalent noise resistor, and f is the bandwidth, determines a noise voltage across the tuning diode. Practical values of $R_{\alpha eq}$ for carefully selected tuning diodes are in the vicinity of 1000Ω to 50kΩ. For $R_{\alpha eq}$ = 10k:

$$V_n = \sqrt{4x4.2x10^{-21}x10{,}000}\,. \qquad (2)$$

The resulting noise voltage is $1.265 \times 10^{-8}\ V\sqrt{Hz}$. This noise voltage, generated from the tuning diode, is now multiplied by the VCO gain, resulting in the rms frequency deviation $(\Delta f_{rms}) = K_0 \times (1.256 \times 10^{-8}\ V)$ in a 1-Hz bandwidth.

In order to translate this into the equivalent peak phase deviation,

$$\theta_d = \frac{K_0\sqrt{2}}{f_m}(1.265 \times 10^{-8}\ \text{rad}) \qquad (3)$$

in 1-Hz bandwidth or, for a typical oscillator gain of 100 kHz/V,

$$\theta_d = \frac{0.00179}{f_m}\ \text{rad in 1-Hz bandwidth.} \qquad (4)$$

For f_m =25 kHz (typical spacing for adjacent-channel measurements for FM mobile radios), θ_c=7.17 x 10^{-8}. This can be converted now into the SSB signal-to-noise radio

$$L(f_m) = 20\log_{10}\frac{\theta_c}{2} \qquad (5)$$

$$= -149\ \text{dBc / Hz.}$$

Figure 8 shows the single sideband noise of an oscillator with tuning diodes attached (VCO). As the tuning range increases and the tuning diode has more influence, it becomes the dominant factor for the noise similar to that of the flicker noise contribution. Please note that this noise contribution has to be added to the Leeson equation and will dominate the noise.

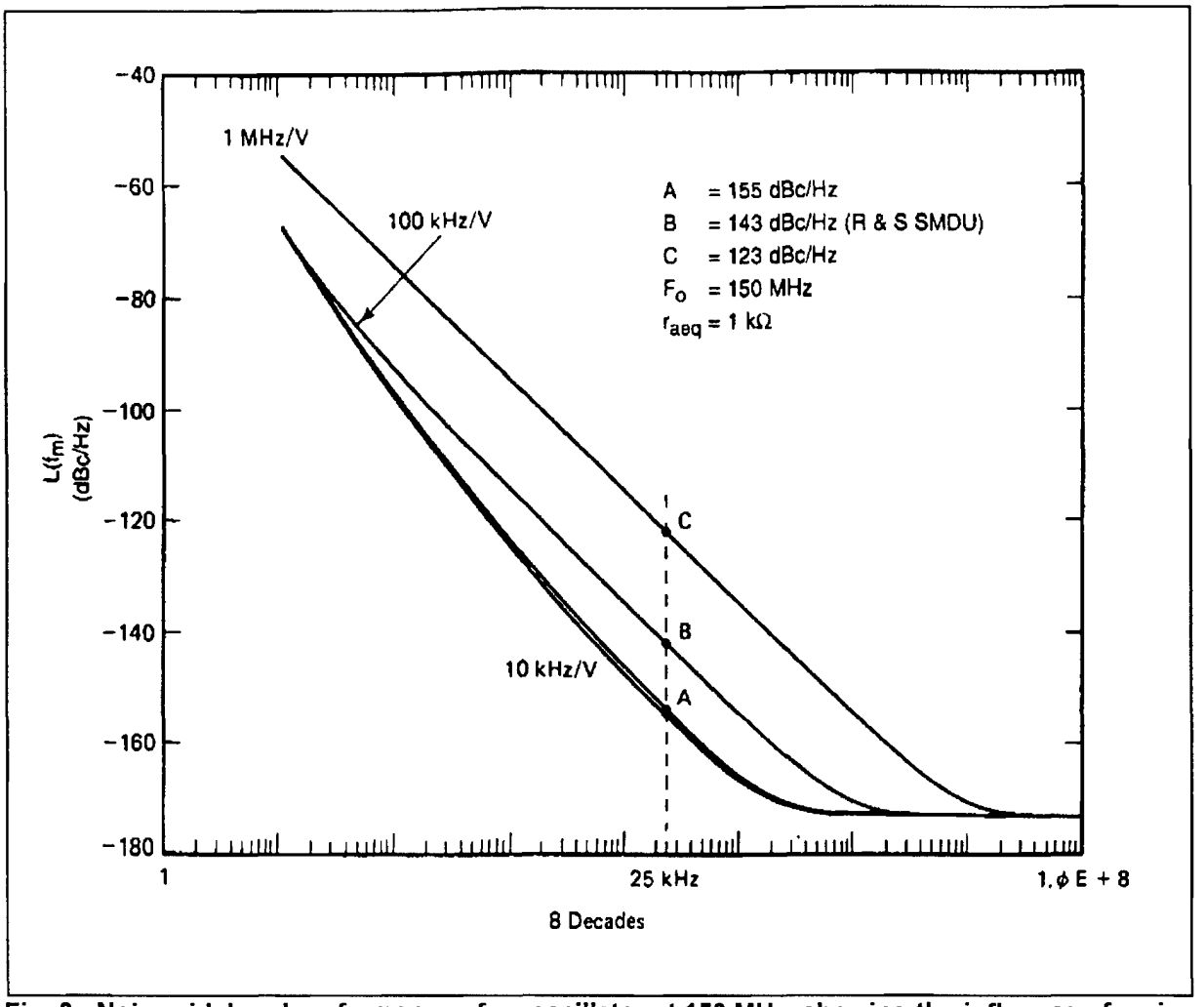


Fig. 8 - Noise sideband performance of an oscillator at 150 MHz, showing the influence of various tuning diodes and increased noise as a function of modulation sensitivity.

Practical Examples

A. SAW Oscillator

Figure 9 shows the schematic of the SAW oscillator, which has many applications including use as a reference for auxiliary frequencies in multi-loop synthesizers. The fact that the tuning diode permits a certain amount of pulling, allows for its use as part of a phase-locked loop. In beginning a design of a SAW oscillator with CAD applications, one needs to select values which provide sufficient feedback to meet start-up conditions. This is easily achieved by using a tight coupling between the resonator and the input and output. Next, it is critical to use a reasonable and well-characterized transistor. Table 2 shows a printout of the nonlinear parameters required for the analysis. The nonlinear model used is consistent with the modeling approaches done in Spice. It is more difficult to obtain the noise parameters for the flicker noise, which had been

obtained as part of the parameter extraction process using a test oscillator for calibration purposes. This is an indirect method but has proven to be very reliable. Extracted parameters for devices which were done solely for Spice applications should be viewed cautiously because the level of accuracy required for RF and microwave applications frequently exceeds the level of effort expended in standard Spice parameters.

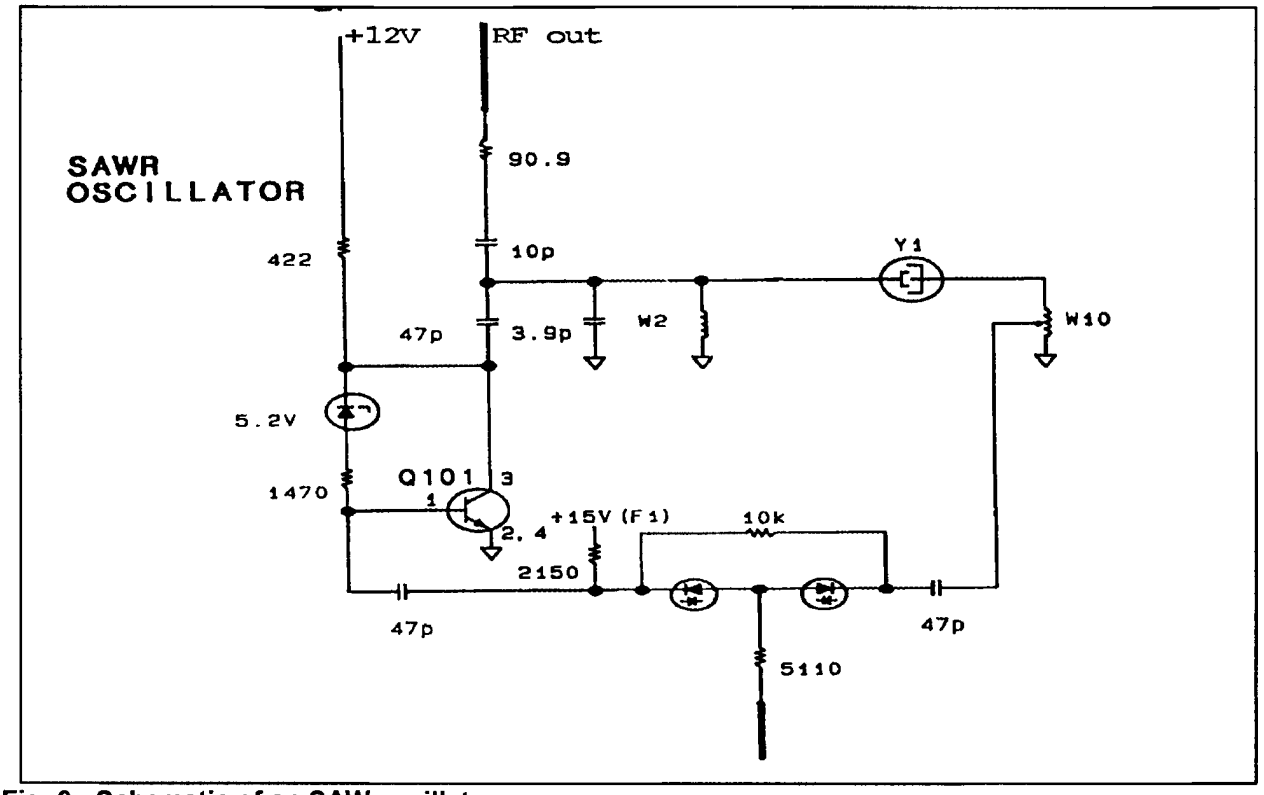


Fig. 9 - Schematic of an SAW oscillator

Table 2

```
* NONLINEAR EQUIVALENT MODEL OF C:\SCOUT\DEVICES\BJTS\NEC\NE68530\NE68530_.MOD

* BIP model parameters for Microwave Harmonica
BIP n1 n2 n3   NAME=SCOUT
+   IS  =  .1105E-15  BF  =  .1325E+03  NF  =  .9866E+00  VA  =  .2122E+02
+   IKF =  .2608E+00  ISE =  .4111E-10  NE  =  .8094E+01  BR  =  .4130E+02
+   NR  =  .1098E+01  VB  =  .3630E+01  IKR =  .1200E-01  ISC =  .1020E-15
+   NC  =  .9110E+00  RBM =  .3293E+01  RB  =  .5236E+01  IRB =  .1078E-03
+   CJE =  .1687E-13  VJE =  .1000E+01  MJE =  .5000E+00  CJC =  .5569E-12
+   VJC =  .1000E+01  MJC =  .5000E+00  XCJC=  .2323E+00  FC  =  .7766E+00
+   TF  =  .1130E-10  XTF =  .2534E+01  VTF =  .1051E+01  PTF =  .4458E+01
+   ITF =  .2500E-01  TR  =  .5000E-08  RE1 =  .1277E+01  RC2 =  .6428E+00
+   LB  =  .9110E-09  LE  =  .9532E-09  LC  =  .1145E-09  CBC =  .4781E-13
+   RB2 =  .3403E+01  CBEP=  .7433E-13  CBCP=  .1827E-13  CCEP=  .4376E-13
+   ZBT =  .5877E+02  LBT =  .3204E-03  ZCT =  .4943E+02  LCT =  .1659E-02
+   ZET =  .6939E+02  LET =  .4274E-03
```

Table 2 - parameter extraction from the NEC NE68530 bipolar microwave transistor. Parameter extraction was done by Compact Software, Inc.

The first step is to analyze the circuit in a suitable harmonic balance simulator such as Compact Software's Microwave Harmonica Workstation product. The software will determine the actual frequency of the oscillator, its output power and SSB phase noise. The mathematical background of the application had been outlined in *Design and Optimization of Low-Noise Oscillators using Nonlinear CAD Tools*, which was presented at the 1994 Frequency Control Symposium [9].

Figure 10 shows the predicted phase noise as determined by the initial conditions or start-up values chosen by the designer.

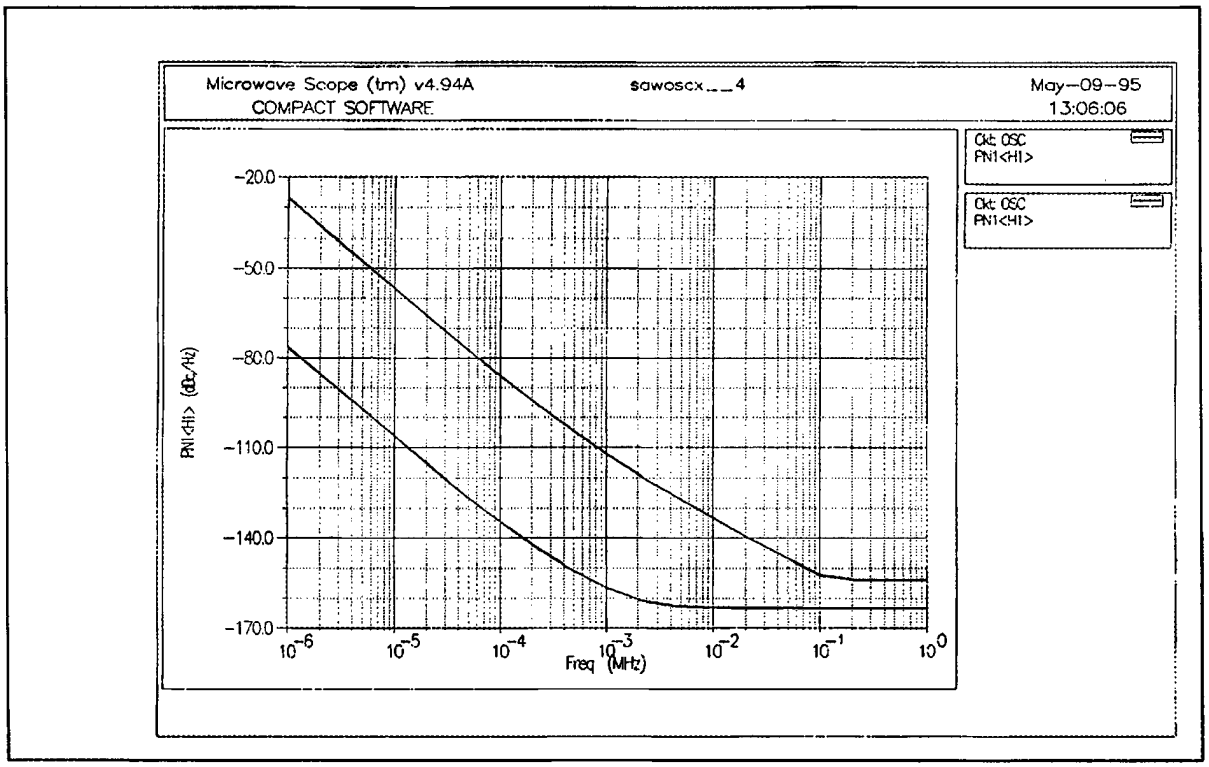


Fig. 10 -Phase noise as determined by the initial start-up values and after optimization.

After invoking optimization, which simultaneously tried to optimize the phase noise, output power and reduction of harmonics at the output signal, Figure 10 also shows the final result following optimization. The values changed by the optimizer were the DC current of the transistor, the series pulling capacitor for the SAW device and the transformation ratio which couples the SAW device between the output tuned circuit and the input. As this value changes, it has influence on both the insertion loss and the operating Q. It is interesting to note that in reviewing the literature, it took approximately 10 years to optimize the phase noise to a performance level, while this program achieved the same result in less than one hour -- including entering all the necessary parameters. Figure 11 shows the measured phase noise as published by Parker, Montress and Driscoll in their various Frequency Control Symposium contributions.

It is also possible to enter the phase noise contribution of the resonator and show its modeling approach, which will be published at a later time. We need to obtain more data from experiments and measurement predictions to calibrate the accuracy of this approach. We can obtain additional

insight into the oscillator's behavior by inspecting the currents and voltages. Figure 12 shows the voltages at the various transistor terminals.

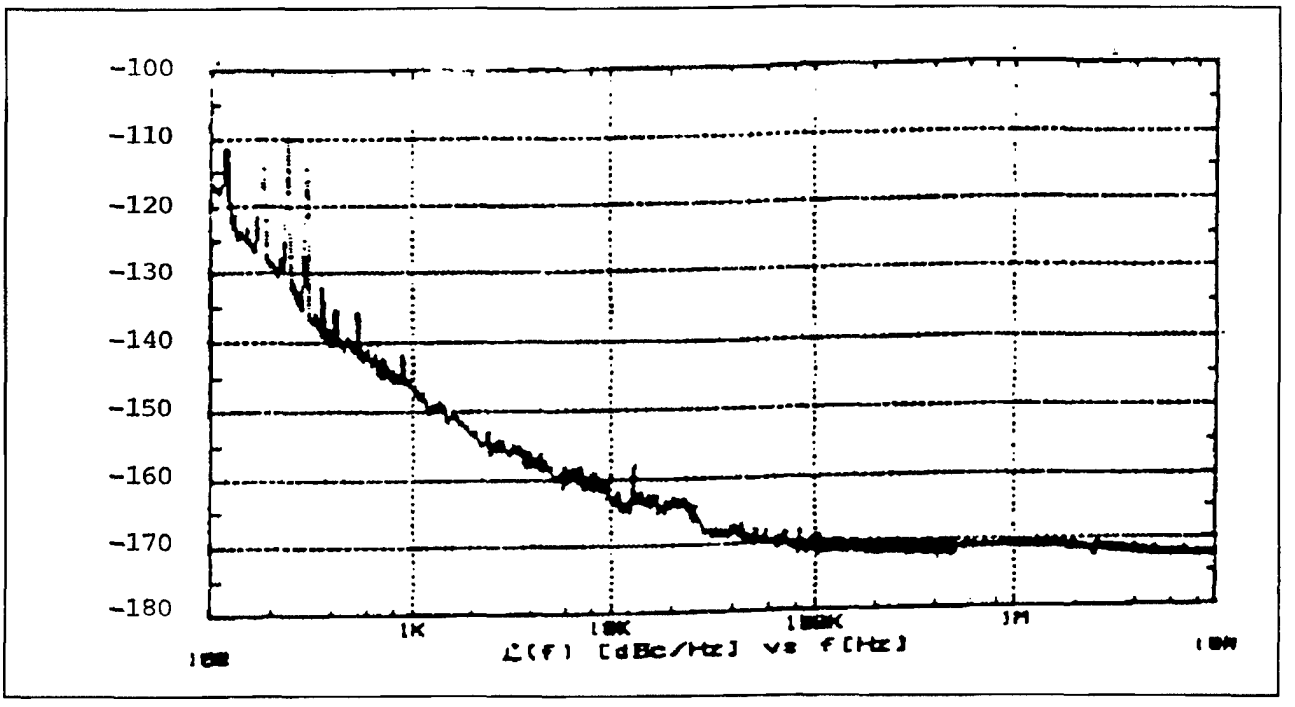


Figure 11 - measured phase noise of the optimized circuit shown in Figure 9 after optimization. [10]

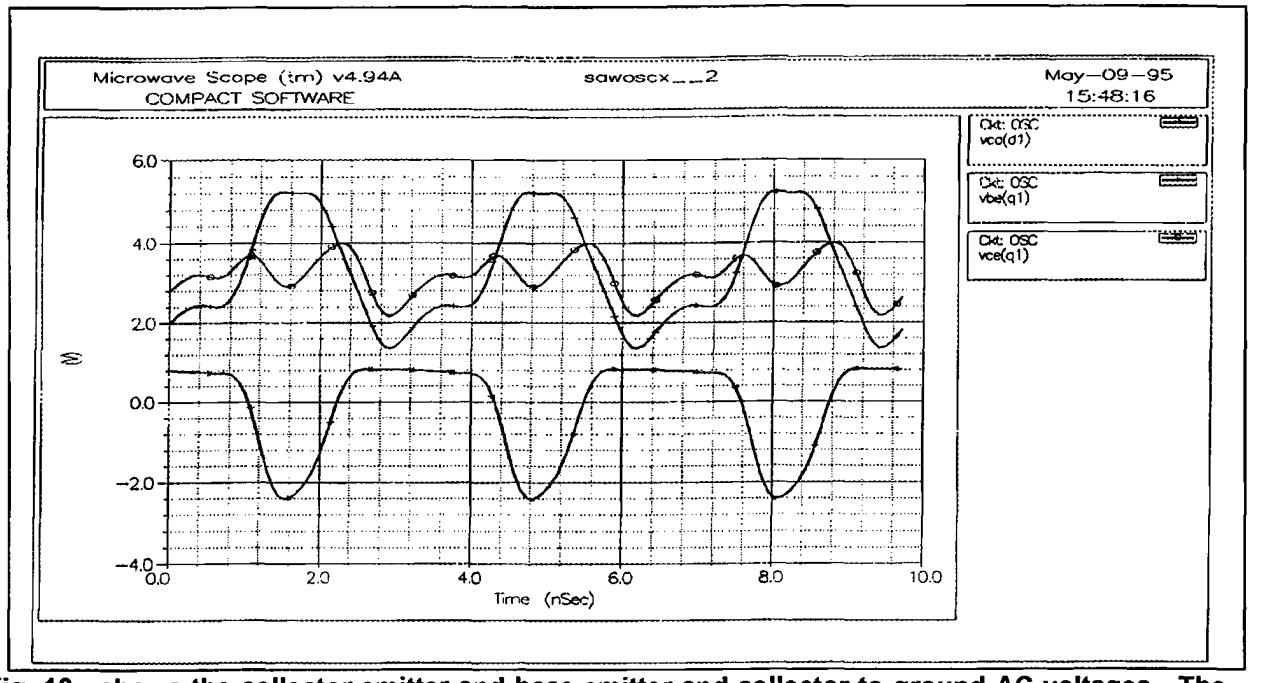


Fig. 12 - shows the collector-emitter and base-emitter and collector to ground AC voltages. The transistor operates in a highly nonlinear mode as can be seen from the negative peaks of the base emitter voltage. This requires the selection of a transistor with an appropriate high breakdown voltage.

B. DR-Based Oscillators

In Figure 1, the state-of-the-art DRO, as a function of different configurations, has been plotted. The publications mentioned do not address the comparison between bipolar and FET performance. Figure 13 shows the schematic of the 6 GHz DRO and Figure 14 shows the predicted phase noise of the 6 GHz DRO. For reasons of noise analysis, it uses the Materka model for which the CAD simulator has a bias-dependent noise model. The SSB phase noise of the simple oscillator using the DRO as a BandStop filter achieves 115 dB/Hz after optimization, which is consistent with the results previously achieved by the researchers.

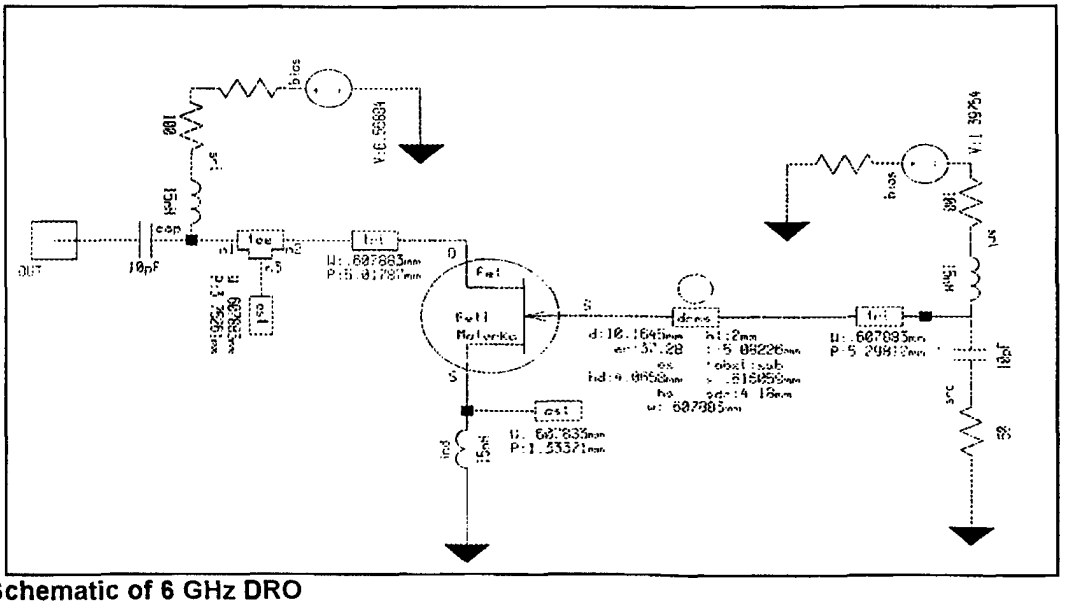


Fig. 13 - Schematic of 6 GHz DRO

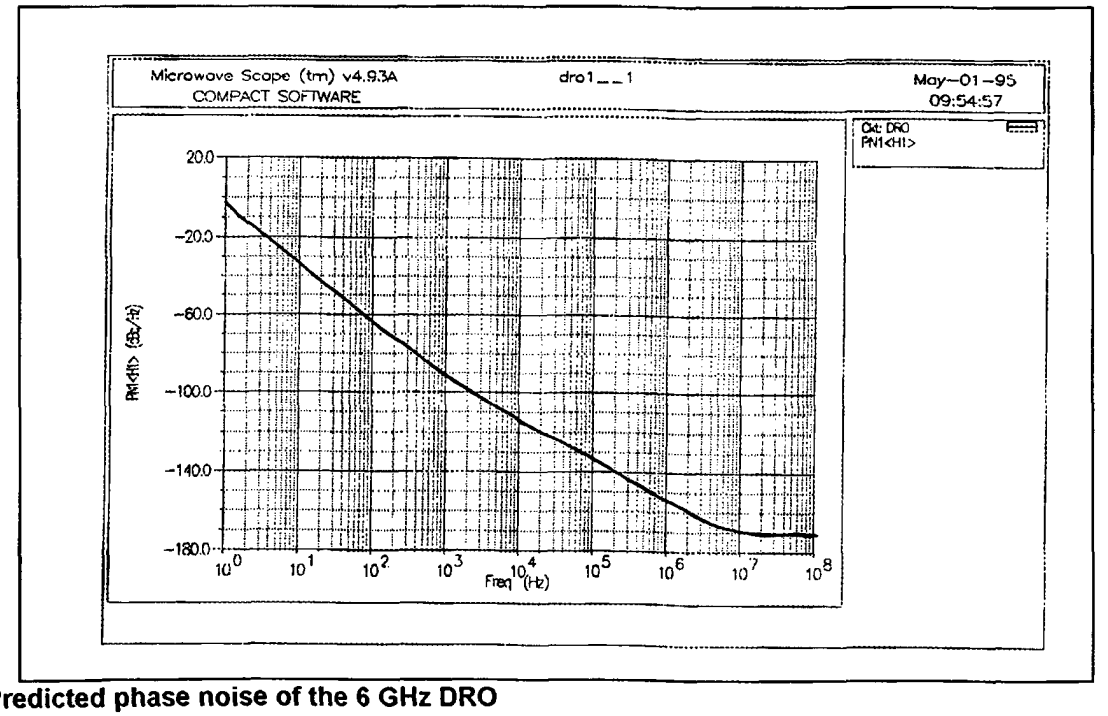


Fig. 14 - Predicted phase noise of the 6 GHz DRO

C. Bipolar Transistor-Based Oscillator

Figure 15 shows the schematic of the 10 GHz bipolar transistor. As silicon bipolar transistors and heterojunction bipolar transistors increase in frequency, it is possible to build these types of oscillators with bipolar transistors and, therefore, for better close-in phase noise at lower frequencies.

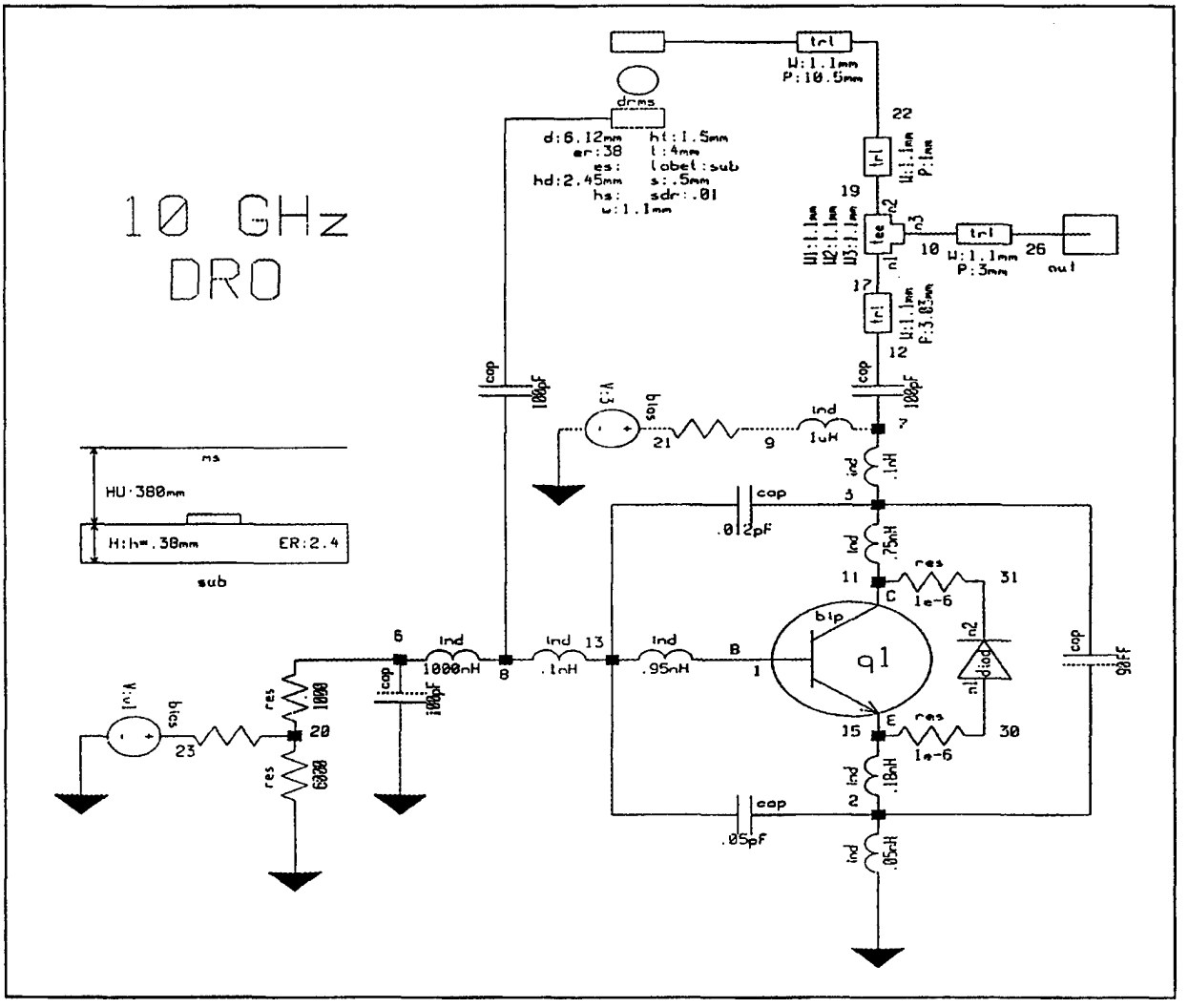


Fig. 15 - schematic of the 10 GHZ bipolar transistor-based oscillator

The bipolar transistor oscillator as shown in Figure 16 has better close-in phase noise and similar far-out phase noise to that of the GaAs FET. Such performance would not have been possible until recently.

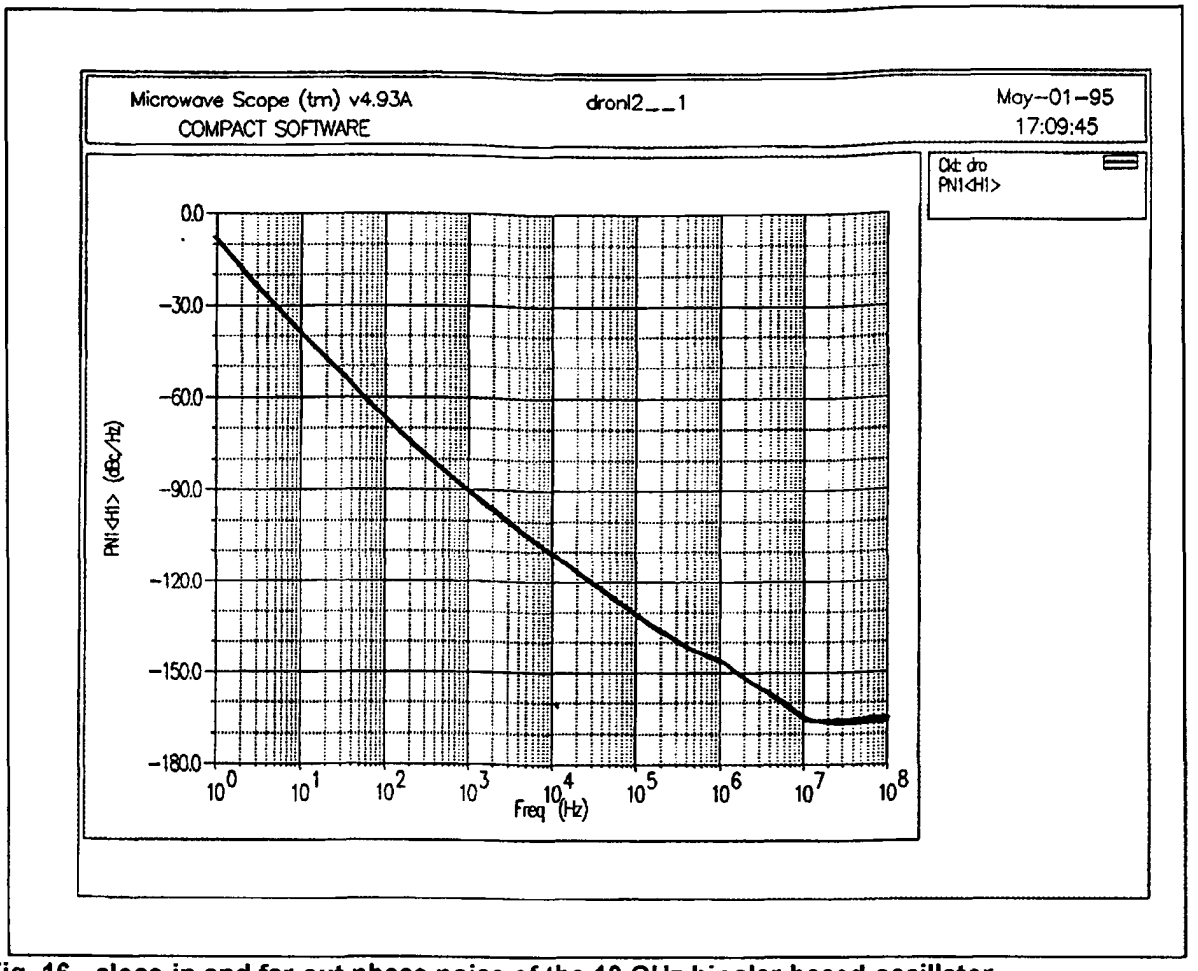


Fig. 16 - close-in and far-out phase noise of the 10 GHz bipolar-based oscillator

Time Domain Analysis

Figures 17 and 18 show the time domain response of the DRO. This analysis can only be done in a true time domain circuit analysis program which also provides discontinuity models. Compact's Super-Spice is ideal for this application.

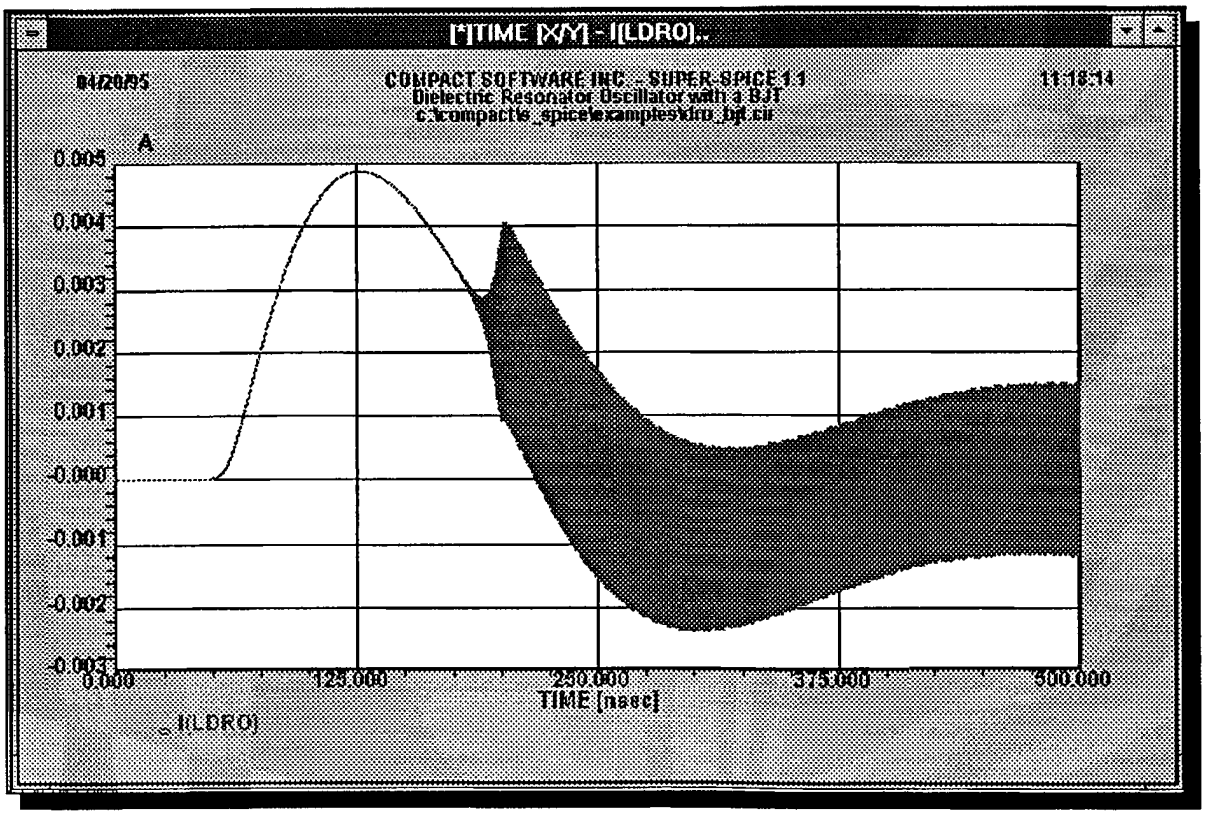


Fig. 17 - Initial current surge 50 nsec after switch-on time

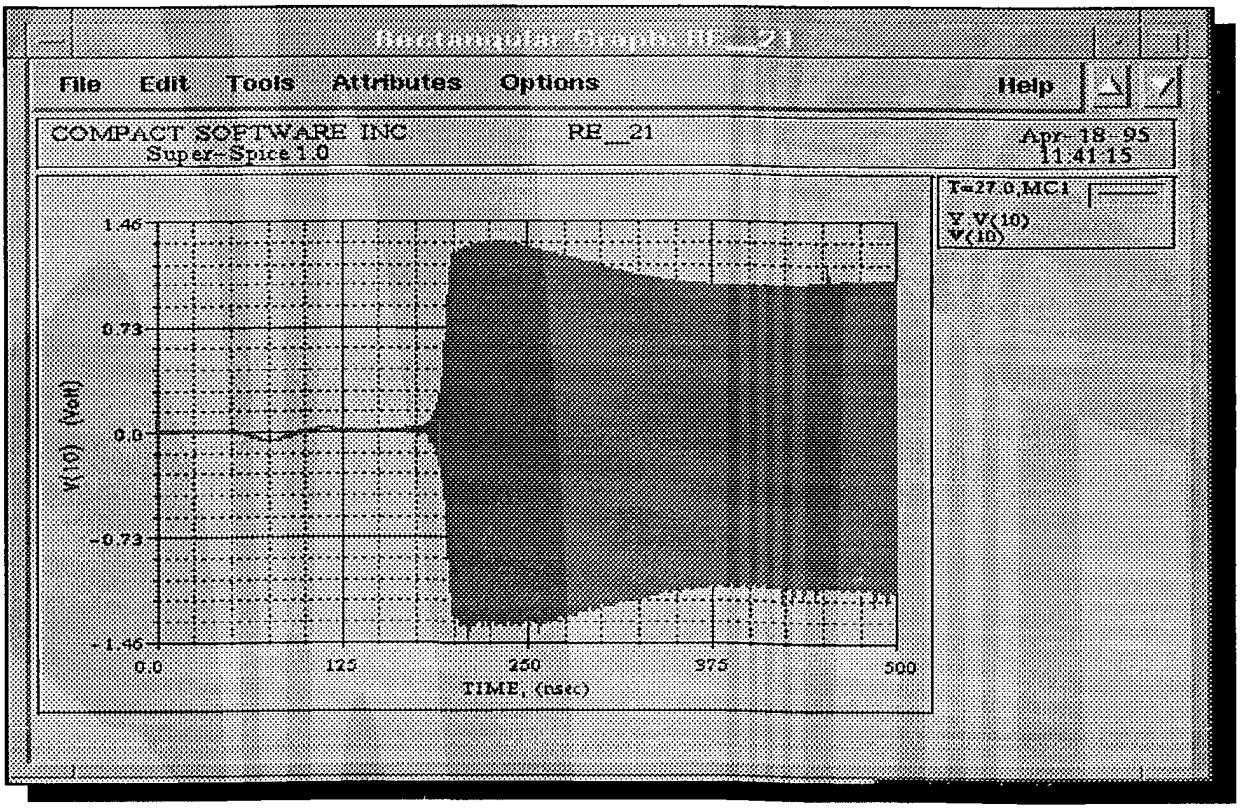


Fig. 18 - Start-up condition of the DRO with BJT

SUMMARY

The purpose of this presentation was to demonstrate that high performance oscillators, such as SAW or DROs, can be designed and optimized using modern CAD tools, which have phase noise optimization capabilities. The basis for this optimization process is the availability of a good nonlinear model, which also encompasses bias-dependent noise and compared to the state-of-the art oscillators, achieves good agreement. This CAD-based approach also drastically reduces the length of the design cycle.

ACKNOWLEDGEMENTS

I would again like to acknowledge the contributions of various individuals to include M.M. Driscoll, T.E. Parker, Professor Vittorio Rizzoli of the University of Bologna and the Compact Software Engineering Team.

References

[1] R.A. Pucel, "The GaAs FET Oscillator, Its Signal and Noise Performance," *40th Annual Frequency Control Symposium*, 1986. pp. 385-391.

[2] M.M. Driscoll, *"Low Noise Oscillator Design using Acoustic and Other High Q Resonators"*, 44th Annual Symposium on Frequency Control, Baltimore, MD, May 1990.

[3] T. Itoh and R.S. Rukodas, "New method for computing the resonant frequencies of dielectric resonators," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-25, pp. 52-54, Jan. 1977.

[4] Y. Komatsu and Y. Murakami, "Coupling coefficient between microstrip line and dielectric resonator," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-31, pp. 34-40, Jan. 1983.

[5] D. Kajfez and P. Guillon, eds., *Diectric Resonators*, Artech House Books, 1986.

[6] T. Bhattacharjee and X. Tu, "Investigations on EM field coupling in dielectric resonator, " *IEEE Trans.* on Magnetics, vol. 25, no. 4, pp.2935-2937, July 1989.

[7] A. Abaramowitz, "Letters to the editor," Microwave Journal, pp. 165, Oct. 1990.

[8] G.K. Montress, Raytheon Company, *"Low Phase Noise VHF and Microwave Sources Using SAW Frequency Controlling Elements"*, IEEE MTT-S International Microwave Symposium, June 14, 1993.

[8] G. D. Vendelin, A.M. Pavio, U.L. Rohde, *Microwave Circuit Design: Using Linear and Nonlinear Techniques*, John Wiley & Sons, New York, 1990.

[9] U.L. Rohde, C.R. Chang, J. Gerber , *Design and Optimization of Low-Noise Oscillators using Nonlinear CAD Tools*, presented at the Frequency Control Symposium, Boston, MA June 1-2, 1994

[10] S.P. Caldwell, M.M. Driscoll, S.D. Stansberry, D.S. Bailey, H.L. Salvo, "*High-Overtone, Bulk Acoustic Resonator Frequency Stability Improvements*", 1993 IEEE International Frequency Control Symposium, pp. 744 -748.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

IMPROVED PERFORMANCE OF A TEMPERATURE COMPENSATED LN_2 COOLED SAPPHIRE OSCILLATOR*

David G. Santiago, Rabi T. Wang, and G. John Dick
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109-8099, U.S.A.

Abstract

We report on improved stability in a whispering gallery sapphire resonator for which the dominant WGH_{n11} microwave mode family shows frequency-stable, compensated operation for temperatures above 77K. Several modifications during the past year have led to significant improvements in performance. Current tests with improved thermal stability provide Allan Deviation of frequency of $2.6 - 4 \cdot 10^{-13}$ for measurement times of $1 \le \tau \le 100$ seconds. We project a frequency stability of 10^{-14} for this resonator with stabilized housing temperature and with a mode Q of 10^7.

Introduction

Improved ultra-stable frequency standards such as JPL's Linear Ion Trap Standard (LITS) are necessary to meet radio science requirements for future NASA/JPL missions [1]. To meet all mission stability requirements, the local oscillator steered by the trapped ion discriminator must have an Allan Deviation of better than 5×10^{-14} for $1 \le \tau \le 100$ seconds. The best available crystal quartz oscillator performance is no better than 1×10^{-13} for $1 \le \tau \le 100$ seconds.

Cryogenic superconducting or sapphire oscillators [2-5] achieve the desired performance, but are complex and expensive to maintain due to their operation at liquid helium temperature (~4.2K). High stability is achieved because the temperature coefficients of expansion and dielectric constant are 'frozen out' at this temperature. A cryogenic oscillator operating at liquid nitrogen (LN_2) temperature or above could be a simpler, less expensive and more compact solution. A mechanical compensation of the temperature coefficients of expansion and dielectric constant has the capability to provide the stability necessary for the LITS's future local oscillator.

Methodology

A traditional whispering-gallery mode resonator, as in Figure 1 (top), has an intrinsic quality factor (Q) of about 30 million at 77K and provides excellent phase noise performance when used to stabilize a crystal quartz oscillator [6-7]. The frequency stability, however, is poor due to the temperature coefficients of expansion and dielectric constant in the sapphire.

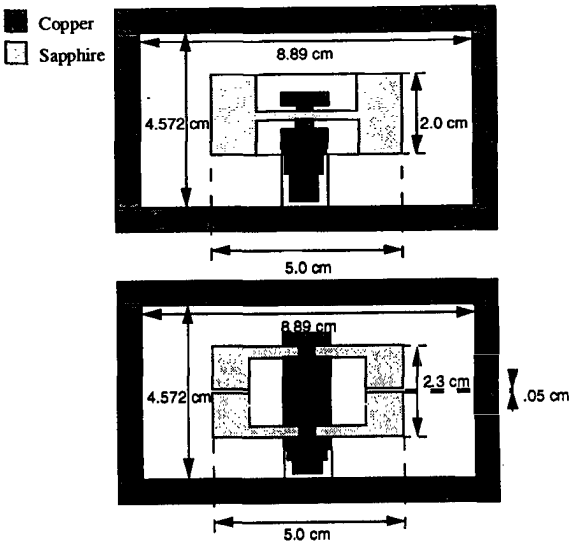


Figure 1. Top resonator is a typical whispering-gallery mode resonator which has poor frequency stability at LN_2 temperature due to temperature fluctuations. The bottom resonator's frequency is temperature compensated at ~87K - the expansion of the copper increases the gap spacing between the sapphire elements which counteracts the effect of the expansion and increase in dielectric constant in the sapphire.

The solution for mechanical compensation is a sapphire-copper composite structure, as in Figure 1 (bottom), in which a split sapphire resonator is separated by a copper post [7-8]. Increasing temperature, which would tend to decrease resonant frequency, causes the copper post to expand separating

* This work was sponsored by and carried out in part at Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

0-7803-2500-1/95 $4.00 © 1995 IEEE

the sapphire elements, thereby increasing the vacuum gap and causing an increase in frequency. At a certain operating temperature these effects completely cancel, and therefore compensate the resonator frequency for temperature variation.

Sources of Frequency Instability

We have identified several significant sources of frequency instability during the development of our compensated sapphire resonator stabilized oscillator. Vibration sensitivity is a concern because of the multiple element structure of our resonator. No deliberate quantitative testing and analysis was performed to determine vibration sensitivity, but during initial tests the system was subjected to vibration and impulses. Several mechanical resonances were observed in the range 1 kHz<f<10 kHz with ringing times of less than a second. These are not expected to degrade frequency stability performance as the resonator's response and vibration induced noise level were no greater than those seen in the traditional type sapphire resonator or the crystal quartz oscillator.

Flicker noise in the rf system components are limiting factors in the completed system. We therefore employ the lowest noise components available and design for the shortest microwave path lengths possible.

Temperature fluctuations and gradients in the system are the largest contributors to poor stability performance. We achieve a thermal time constant in the composite resonator of less than five seconds (unit responds as a single thermal mass) for operation of the compensation mechanism.. A thermal time constant of 300 seconds between cavity and resonator allows the compensated resonator to provide temperature compensated ultra-stable performance if not impaired by the rest of the system.

A frequency lock loop gain of 100 is required to achieve a stability of 10^{-13} with a VCO of stability 10^{-11}. Insufficient loop gain is seen as a turn up in frequency stability at short averaging times in the upper two curves of Figure 2..

Previous Status

Previously reported performance results were obtained using a compensated resonator as in Figure 1. In experimental tests, the WGH_{811} mode showed a frequency turn-over temperature of 87K in agreement with finite element calculations. Tests of oscillator operation show an Allan Deviation of frequency variation of $1.4 - 6 \cdot 10^{-12}$ for measuring times $1 \le \tau \le 100$ seconds with unstabilized resonator housing temperature and a mode Q of $2 \cdot 10^6$. These first stability results for such a compensated sapphire resonator stabilized oscillator are seen in the first (top) graph of Figure 2. Further experiments have discovered several limitations in the original system design. Over the past year several modifications have been made to the system which have steadily improved stability performance.

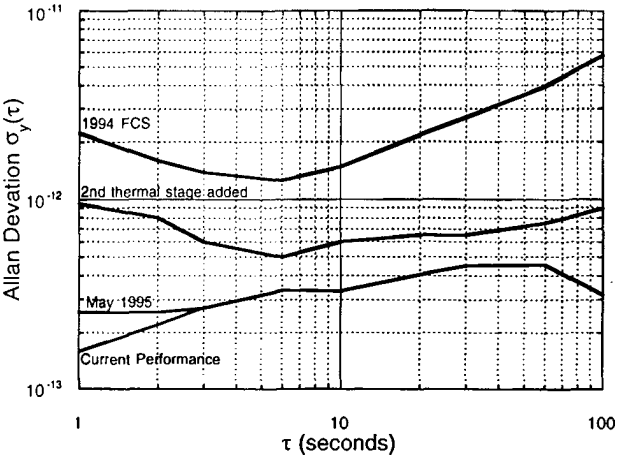


Figure 2. Allan Deviation of the compensated sapphire resonator oscillator presented at last years symposium, after addition of a thermal isolation stage, after Pound circuit and various other modifications (May 1995), and current performance.

Modifications for Improvement

Resonator and Cavity Thermal Stability

The performance limit reported in last year's proceedings (Figure 2) was found to be limited by thermal stability of the resonator containment can. The copper can cavity which houses the resonator was thermally well anchored to the LN_2 bath. Changes in room temperature and pressure as well as the LN_2 level effect the temperature of the liquid nitrogen bath. These temperature fluctuations, which are well followed by the resonator housing, limit Allan Deviation performance to parts in 10^{-12}. An isolation stage with a long thermal time constant was added to dampen the effect of the LN_2 temperature fluctuations on the resonator cavity.

The design, as shown in Figure 3, is re-entrant to minimally effect the resonator's placement in the cryostat. The original bottom plate with the copper center that sits in the LN_2 bath is spaced approximately 8mm from the copper can, but the thermal path length is approximately 6.5cm. The thermal isolation stage is composed of a stainless steel 'deep dish' in which a copper cylinder is attached. On top of the copper

cylinder is a stainless steel plate which only makes contact to the copper can with a ~0.5cm width ring at its outer radius. The copper cylinder has thermistors and a heater element which allow the temperature of the stage to be controlled.

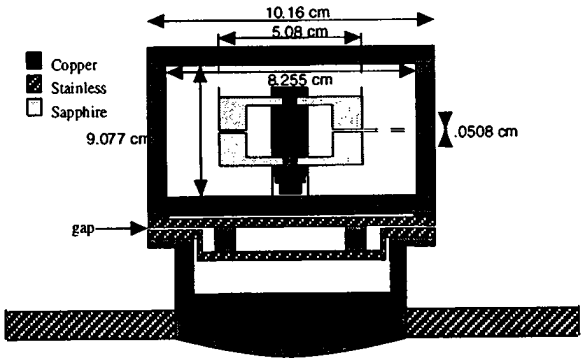


Figure 3. Resonator can with added thermal isolation stage.

First tests with the stabilized resonator housing temperature achieved Allan Deviation of frequency variation of $5 - 6 \cdot 10^{-13}$ for measuring times $5 \leq \tau \leq 30$ seconds, and better than 10^{-12} for all measurement times between 1 and 100 seconds. The resonator operated at its turnover temperature (~87K), the new isolation stage was a few degrees above the LN_2 temperature, and there was no servo control of either of the heaters during the experiment resulting in the second graph of Figure 2. The stability of the current sources used to drive the heaters was found to be poor, so they were replaced with more stable supplies for subsequent experiments.

Pound Circuit Modifications

The first two graphs in Figure 2 show a degradation for $\tau < 5$ seconds due to insufficient loop gain in the frequency lock circuitry. The resonator is frequency locked to a 100 MHz crystal quartz voltage controlled oscillator using a Pound circuit as in Figure 4. Previous experiments used a 50-200 kHz phase modulation frequency injected into the VCO input. The frequency multiplier frequency lock circuitry loses its lock to the VCO at modulation frequencies greater than 200 kHz. Therefore we modified the multiplier to allow injection of a 2 MHz modulation frequency into its phase modulation input. This increase in loop gain in the frequency lock circuitry greatly improved stability performance and, as can be seen in the third graph of Figure 2, the slope from τ = 1-5 seconds was eliminated.

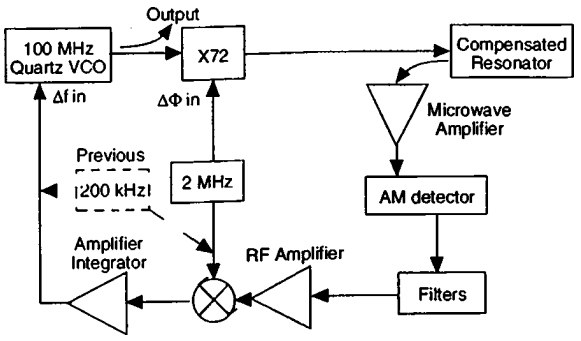


Figure 4. Previous Pound (frequency lock) circuit limited to a 200 kHz modulation frequency. To increase loop gain, circuit modified to allow a 2 MHz modulation frequency.

Other Improvements

Increasing the loop gain proves to be the significant factor in achieving the May 1995 performance, but several other improvements contribute to the oscillator's stability. The resonator is the only component of the system contained in the cryostat. The external microwave components are exposed to the environment of our open laboratory. These mixers, hybrids, amplifiers, connectors, and cable lengths are sensitive to temperature fluctuations and in combination greatly contribute to the instability of the system. By insulating the microwave components, including the Pound circuitry, with foam, the system reaches a more stable operating temperature.

Another effect is temperature fluctuations and gradients on the relatively long coaxial line which feeds the resonator in the cryostat. It is at LN_2 temperature at the resonator and room temperature at the cryostat's input making it sensitive to LN_2 and room temperature fluctuations. This instability is reduced by better isolating the line from the cryostat wall and maintaining a more stable LN_2 surface temperature and level. This improvement noticeably helps provide the May 1995 stability in Figure 2.

Monitoring various temperatures in the system, we found that the sapphire temperature followed the outer can as well as showing its own temperature fluctuations. This indicated a thermal 'leak' in the resonator. The thermal path was a combination of a small vacuum leak and thermal radiation. Sealing the leak and adding radiation shielding significantly improved short term stability as seen in the current performance in Figure 2.

Future Work

Several additional improvements have been identified as necessary to achieve the desired oscillator stability. The resonator's thermal environment is currently free running and not actively controlled. Addition of feedback control electronics to the heater elements of the system with better than milliKelvin temperature resolution is required for the desired ultra-stable performance.

Temperature sensitivity of the microwave components continues to be a concern, and so it may be advantageous to move the microwave components closer to the resonator (perhaps in the cryostat) to reduce the path length of the microwave signals.

A significant improvement in stability will come with increased resonator Q. The current resonator Q for the compensated mode used is 2-3 million, while other modes in the resonator show Q's of 20 million. This is due to losses from poor surface cleanliness of the sapphire, especially in the gap where the compensated mode is most sensitive. An improved sapphire cleaning and assembly procedure should allow us to achieve resonator Q's of 7-8 million. The calculated noise limit performance at a Q of 8 million is 10^{-14} compared to $3 \cdot 10^{-14}$ for our current Q.

Conclusion

We have demonstrated a new USO capability with a 6-10x improvement over the past year. Stabilities of $2.6 \cdot 10^{-13}$ for $1 \le \tau \le 200$ seconds have been demonstrated in our laboratory. We predict stabilities of $7 \cdot 10^{-14}$ for $1 \le \tau < 200$ seconds in the near future and $\sim 1 \cdot 10^{-14}$ for $1 \le \tau < 200$ seconds with a resonator Q of 8 million.

References

[1] J. D. Prestage, R. L. Tjoelker, R. T. Wang, G. J. Dick, and L. Maleki, "Hg^+ Trapped Ion Standard with the Superconducting Cavity Maser Oscillator," *IEEE Trans. Instrumentation and Measurement*, vol 42, no. 2, pp. 200-205 (1993).

[2] A.G. Mann, A.N. Luiten, D.G. Blair and M.J. Buckingham, "Ultra-Stable Cryogenic Sapphire Dielectric Microwave Resonators", *Proc. 1992 IEEE Frequency Control Symposium*, pp. 167--171 (1992).

[3] A.J. Giles, S.K. Jones, D.G. Blair and M.J. Buckingham, "A High Stability Microwave Oscillator based on a Sapphire Loaded Superconducting Cavity," *Proc. 43rd Symposium on Freq. Control*, 83--89 (1989).

[4] R. T. Wang and G. J. Dick, "Improved performance of the Superconducting Cavity Maser at short measuring time," Proceedings of the 44th Annual Frequency Control Symposium, 89--93 (1990).

[6] D.G. Santiago and G.J. Dick, "Closed Loop Tests of the NASA Sapphire Phase Stabilizer," *Proc. 1993 IEEE International Frequency Control Symposium*, pp. 774--778 (1993).

[7] D.G. Santiago and G.J. Dick, "Microwave Frequency Discriminator with a Cooled Sapphire Resonator for Ultra--Low Phase Noise," *Proc. 1992 IEEE Frequency Control Symposium*, 176--182 (1992).

[8] G. J. Dick, D. G. Santiago and R. T. Wang, "Temperature-Compensated Microwave Sapphire Oscillator at Temperatures Above 80 Kelvin," *Proc. 1994 IEEE International Frequency Control Symposium*, 421-432 (1994).

[9] G. J. Dick, D. G. Santiago and R. T. Wang, "Temperature-Compensated Microwave Sapphire Oscillator at Temperatures Above 77 Kelvin," *IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control*, Sept. (1995).

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SPECTRAL PERFORMANCE OF SAPPHIRE DIELECTRIC RESONATOR-CONTROLLED OSCILLATORS OPERATING IN THE 80K TO 275K TEMPERATURE RANGE

Michael M. Driscoll* and Robert W. Weinert**
*Westinghouse Electronics Systems Group, Baltimore, MD
**Westinghouse Science & Technology Ctr., Pittsburgh, PA

Abstract

This paper reports on the phase noise performance obtained for X-band oscillators using cooled, sapphire dielectric resonators as the frequency-determining element. We report on results obtained using: (1) a TE-cooled, high-order mode resonator purchased from Poseidon Industries exhibiting a loaded Q of 140,000 at an operating temperature of 275K, and (2) a low-order (TE02) mode resonator fabricated at Westinghouse exhibiting a loaded Q of 350,000 at an operating temperature of 77K. The oscillator sustaining stage designs incorporate GaAs amplifier flicker-of-phase noise feedback reduction techniques as well as a technique that avoids the need for X-band signal amplification altogether. Sustaining stage open loop, flicker of phase noise levels obtained are typically 20dB below those normally exhibited by GaAs X-band amplifiers.

In the TE-cooled resonator-controlled oscillator, flicker-of-frequency noise performance was obtained that is characterized by $Sy(f)=6X10^{-23}/f$ [£(1KHz)=-115dBc/Hz]. For the TE02 mode resonator (T=80K), flicker-of-frequency noise levels are near-identical to those of the crystal oscillator-based reference source used in the phase noise measurements [$Sy(f)=3X10^{-24}/f$]. In addition, single axis measurement of the vibration sensitivity of each resonator has been made indicating frequency sensitivities ($\delta f/f$) on the order of 0.5 to $1.0X10^{-8}$ per g. Based on the results obtained to date, reasonably accurate predictions of oscillator spectral performance as a function of resonator loaded Q (operating frequency, mode number, and temperature) are possible.

Introduction

In a low noise, microwave signal generator, the characteristics of the generator reference oscillator resonator play an important role with regard to resultant output signal short-term as well as long-term frequency stability. Achievement of lowest phase noise in the output signal requires an "ideal" oscillator resonator that simultaneously exhibits high Q, high operating frequency, high short-term frequency stability, low insertion loss, high drive capability, and low frequency sensitivity to environmental stress-especially to vibration [1-3]. Unfortunately, these characteristics are not provided in entirety by any single resonator technology. Table 1 shows a comparison of key resonator characteristics for some commonly used resonators.

Microwave carrier signal spectral performance tradeoffs are associated with the use of different resonator/oscillator technologies. For example, lowest phase noise floor levels are obtained using microwave oscillators employing dielectric resonators (DRs) operated at high drive level. Lowest near-carrier noise, on the other hand, is achieved via oscillator implementation in the VHF band using much higher Q, less vibration sensitive, but lower frequency quartz crystal resonators having relatively low drive level capability. Oscillator implementation at lower frequency results in higher resultant microwave signal phase noise floor level due to the signal phase (noise) expansion associated with requisite signal frequency multiplication.

Extremely high, X-band unloaded Q values ($2X10^7$ to $2X10^5$) have been reported for high-order mode, sapphire DRs operated in the 77K to 300K temperature range [4-8]. In addition, X-band unloaded Q values of $3X10^5$ have been obtained at 77K using smaller relative volume, low-order mode sapphire DRs using high temperature superconducting (HTS) films [9]. State-of-the-art phase noise performance has been demonstrated using sapphire DR oscillator technology in at-rest hardware using resonators operated at cryogenic temperatures [7]. Oscillator signal near-carrier phase noise levels obtained using sapphire DR technology operated at and near room temperature have not been competitive with VHF quartz crystal oscillator technology in spite of the fact that similar resonator unloaded Q values are obtainable in each case. The reason is that the flicker-of-phase noise level exhibited by the oscillator sustaining stage amplifier is much higher at X-band (using GaAs transistors) than at VHF (using silicon bipolar transistors). In addition, the

0-7803-2500-1/95 $4.00 © 1995 IEEE

vibration sensitivity reported for microwave frequency, ceramic DRs specifically designed for vibration immunity is typically 100 to 1000 times higher than that of VHF quartz crystals [10]. The higher vibration sensitivity of the sapphire DR, coupled with the necessary use (at cryogenic temperatures) of relatively high vibration level, closed-cycle cryocoolers also tend to limit resultant oscillator signal near-carrier spectral performance.

This paper reports on attempts at Westinghouse to develop and evaluate methods for reducing X-band oscillator sustaining stage flicker-of-phase noise and to characterize typically obtained, sapphire DR frequency sensitivity to vibration in an effort to better assess the potential offered by sapphire DR technology with regard to low noise signal generation.

Prototype Sapphire DR Characteristics

X-band oscillators were evaluated using two, prototype resonators. They were: (1) A high-order mode resonator operated at 275K was purchased from Poseidon Industries. The self-contained resonator assembly includes adjustable (TE-cooler) resonator frequency/temperature control circuitry. Resonator operating frequency, loaded Q, and insertion loss were 10.24GHz, $1.4X10^5$, and 10dB, respectively.
(2) A lower order mode (TE02) resonator operated at 77K. Resonator operating frequency, loaded Q, and insertion loss for the TE02 device were 10.08GHz, $3.5X10^5$, and 10dB.

Oscillator Sustaining Stage Design

GaAs Amplifier with Reduced Flicker-of-Phase Noise

X-band GaAs amplifiers exhibit flicker-of-phase noise levels typically 40dB higher than that obtained at VHF using silicon bipolar amplifiers and 20dB higher than that obtained in X-band, schottky diode mixers [11]. Oscillator sustaining stages have been designed that successfully incorporate feedback techniques for sensing and reducing GaAs amplifier flicker-of-phase noise [12]. Those techniques employ a (double balanced mixer) phase detector operating in conjunction with the oscillator resonator, as a frequency discriminator. In these schemes, the detector sensitivity is enhanced by including the high Q oscillator resonator in the detection circuitry. As a result, the detection circuit frequency response is affected by the resonator bandwidth.

Figure 1 shows a method for sensing and reducing GaAs flicker-of-phase noise that functions independently of the oscillator resonator and allows implementation of a relatively wide bandwidth feedback control loop. Figure 2 shows the measured results obtained via utilization of the figure 1 circuit. In figure 2, the limitation in feedback loop bandwidth (200KHz) was associated with use of on-hand, low noise op-amps having smaller than desired, gain-bandwidth product. As can be seen from the figure, 20dB in GaAs amplifier flicker-of-phase noise has been achieved. The 1/f noise levels exhibited in the loop phase detector (double-balanced mixer) constitutes the primary limitation in noise reduction capability. As shown in figure 3, double-balanced mixer phase detector 1/f noise level tends to increase with decreasing sensitivity (decreasing RF drive level). The appearance of increasing phase noise sideband level at carrier offset frequencies above 1MHz in figure 2 is due to the combined effects of signal generator FM noise and unequal group delay in the measurement set-up phase detector input signal paths.

The figure 1 circuit was used as the sustaining stage amplifier in an oscillator incorporating the TE-cooled resonator as the frequency control element. The oscillator block diagram is shown in figure 4. As shown in the figure, the feedback loop op-amp was AC-coupled to the phase shifter for the initial experiments in order to avoid interaction between the oscillator positive feedback loop and the noise reduction negative feedback loop. As a result, the noise reduction loop gain drops below unity at carrier offset frequencies below 10Hz.

Oscillator output signal flicker-of-frequency noise measurements were accomplished via comparison to a 10.24GHz reference signal generator whose output signal is derived from an ultra-low noise, 80MHz crystal oscillator. The phase noise characteristics of the reference signal generator are shown in figure 5. The oscillator noise measurement test set-up is shown in figure 6, and the measurement results are given in figure 7. As shown in figure 7, a flicker-of-frequency noise level given by $Sy(f)=6X10^{-23}/f$ or -115dBc/Hz at 1KHz carrier offset frequency was achieved. Since the resonator half-bandwidth is approximately 37KHz, this performance is consistent with a sustaining stage open loop, flicker-of-phase noise level given by £(1KHz)=-147dBc/Hz. As shown in figure 2(b), this level is within 2dB of that measured for the sustaining stage circuit. At carrier offset frequencies above 1KHz, the figure 7 of data is dominated by the phase noise of the crystal oscillator-derived reference source signal. The increase in noise level at low carrier offset frequencies is a result previously mentioned, low frequency loop gain rolloff associated with AC-coupling of the phase shifter

feedback control signal.

Double Frequency Conversion Sustaining Stage

Figure 8 shows the microwave oscillator circuit that was used for evaluation of the low order (TE02) mode sapphire DR designed for operation at 10.08GHz at a temperature of 77K. As shown in the figure, the sustaining stage is configured to avoid the use of (high flicker-of-phase) GaAs amplifiers altogether. A common microwave frequency, local oscillator signal is used to downconvert the resonator carrier signal to a VHF frequency where signal amplification can be provided using low flicker-of-phase noise, silicon bipolar amplifiers. The downconverted, VHF signal is then upconverted back to the resonator frequency using the same local oscillator signal. The amount of sustaining stage open loop phase noise that results from the FM noise of the common, local oscillator signal depends on the VHF signal path group delay. The degree of L.O. signal FM noise cancellation occurring in the sustaining stage feedback loop is related to the ratio of the group delay in the microwave signal path (i.e. sapphire DR delay) to that in the VHF signal path. For the small delays (on the order of several nanoseconds) typically exhibited by broadband, VHF modular amplifiers, a great degree of local oscillator FM noise cancellation occurs. Oscillation at undesired sapphire DR resonant frequency responses is prevented by virtue of the 200MHz VHF gain rolloff exhibited by the VHF amplifiers used, and additional suppression could have been provided via use of an additional bandpass filter installed in either the microwave or VHF portion of the sustaining stage circuit. The figure 8 arrangement is similar to a scheme described by Pan, et al., except that we have intentionally eliminated both microwave signal amplification and IF signal bandpass filtering [13].

Figure 9 shows the result of measurement of (figure 8) oscillator sustaining stage open loop noise. The measurement was made by substituting an on-hand, 200MHz bandwidth, 6 pole bandpass filter and attenuator for the sapphire DR (figure 10). Bandpass filter inclusion was necessary in order to suppress the unwanted sideband signal generated in the upconverting mixer. The 1/f noise of the (low R port drive level) downconverting mixer is the dominant source of sustaining stage pen loop, flicker-of-phase noise.

Figure 11 shows the oscillator flicker-of-frequency noise measurement test set-up, with the measurement results shown in figure 12. As shown in the figure, the measurement results indicate that the flicker-of-frequency noise obtained via comparison of the reference signal generator to the oscillator are virtually identical to that measured (figure 5) for two, identical, reference signal generators. The implication of the figure 12 data is that the TE02 sapphire DR oscillator flicker-of-frequency noise is near-identical to that of the reference generator and is described by $Sy(f)=3X10^{-24}/f$ or -98dBc/Hz at 100Hz carrier offset frequency. The TE02 resonator loaded Q was on the order of 350,000 (a half-bandwidth of approximately 15KHz). Therefore the figure 12 data is consistent with an oscillator sustaining stage open loop, flicker-of-phase noise level at fm=100Hz of: -98dBc + 20log[100Hz/15KHz] or -142dBc/Hz. This value is consistent with that measured and shown in figure 9.

Sapphire DR Vibration Sensitivity Measurements

Computation of sapphire DR resonant frequency sensitivity to vibration-induced, relative dimensional change between the sapphire "puck" and the enclosure covers indicated that resonators designed for operation on a higher relative mode number should be substantially less sensitive to vibration. In earlier-reported phase noise experiments conducted for signal generators incorporating TE01 mode sapphire DRs, severe output signal spectral degradation due to low-level vibration was noted [9].

Vibration sensitivity experiments were performed for each of the two sapphire DR designs described in this paper. Vibration measurements were confined to a single axis normal to the resonator enclosure end covers, and were conducted for sinusoidal excitation at low frequency in order to assure unity transmissibility (excitation frequency below that of any potential, mechanical resonances). The TE02 resonator measurements were made with the resonator installed in oscillator circuitry using a rather crude arrangement (figure 13) complicated by the fact that the resonator was submersed in a liquid nitrogen bath. The TE-cooled resonator measurements were made with the resonator employed in a passive phase bridge mounted on a shake table. The measurement results are shown in figure 14 and compared with those reported in reference [10] for a TE01 mode, ceramic DR. The TE01 data in figure 14 is consistent with measurements made at westinghouse for a TE01 mode, sapphire DR employing HTS films [9]. As shown in the figure, both prototype resonator designs exhibit substantially less vibration sensitivity than that reported for a (TE01 mode) resonator. Nevertheless, the maximum allowable vibration levels consistent with no degradation in at-rest oscillator signal spectral performance are quite low.

Conclusions

The use of ultra-high Q, sapphire dielectric resonator technology, especially at cryogenic operating temperatures, can provide a means for generation of microwave signals exhibiting exceptionally superior short-term frequency stability.

Reduction of flicker-of-phase noise level in the sustaining stage circuitry is important and results in corresponding reduction in oscillator signal flicker-of-frequency noise.

Reduction in both resonator frequency sensitivity to vibration and cryocooler vibration levels is necessary to fully exploit the use of sapphire DR technology in low noise signal generators.

Acknowledgements

The authors wish to acknowledge the support of ARPA (contract 972-88-C-0050) in connection with companion work not reported here but listed in reference [9] for sapphire dielectric resonator and resonator-stabilized frequency source development using high temperature superconducting films.

References

1. M. M. Driscoll, "Low Noise Microwave Signal Generation Using Bulk- and Surface-Acoustic-Wave Resonators," IEEE Trans. UFFC, Vol 35, No 3, May, 1988, pp. 426-434.

2. M. M. Driscoll, "Measured vs. Volume Model-Predicted Flicker-of-Frequency Instability in VHF Quartz Crystal Resonators," Proc. 1993 IEEE Freq Contr Symp, June, 1993, pp. 186-192.

3. T. E. Parker and D. Andres, "Dependence of SAW Resonator 1/f Noise on Device Size," Proc. 1993 IEEE Freq Contr Symp, June, 1993, pp. 178-185.

4. J. Dick and J. Saunders, "Measurement and Analysis of a Microwave Oscillator Stabilized by a Sapphire Dielectric Ring Resonator for Ultra-Low Noise," Proc. 43rd Freq Contr Symp, May 1989, pp. 107-114.

5. M. E. Tobar and D. G. Blair, "Analysis of a Low Noise Tunable Oscillator Based on a Tunable Sapphire Loaded Superconducting Cavity," Proc. 45th Freq Contr Symp, May 1991, pp. 495-499.

6. K. A. Zaki and X. P. Liang, "Higher Order Modes, High Q Dielectric Resonators for Oscillator Applications, Proc. 1992 IEEE Freq Contr Symp, May 1992, pp. 163-166.

7. C. A. Flory and R. C. Taber, "Microwave Oscillators Incorporating Cryogenic Sapphire Dielectric Resonators," Proc. 1993 Freq Contr Symp, June 1992, pp. 763-773.

8. D. P. Tsarapkin, "An Uncooled Microwave Oscillator with 1 Million Effective Q Factor," Proc. 1993 Freq Contr Symp, June 1993, pp. 779-783.

9. M. M. Driscoll and R. W. Weinert, "Low Noise, Microwave Signal Generation Using Cryogenic, Sapphire Dielectric Resonators: An Update," Proc. 1992 Freq Contr Symp, May 1992, pp. 157-161.

10. S. W. Sparagna, "L-Band Dielectric Resonator Filters and Oscillators with Low Vibration Sensitivity and Ultra Low Noise," Proc. 43rd Freq Contr Symp, May 1989, pp. 94-106.

11. F. G. Ascarrunz, et al., "Investigations of AM and PM Noise in X-Band Devices," Proc. 1993 Freq Contr Symp, June 1993, pp. 303-311.

12. Z. Galani, et al., "Analysis and Design of a Single Resonator GaAs FET Oscillator with Noise Degeneration," IEEE Trans. MTT, Vol MTT-32, Vol 12, Dec, 1984, pp. 1556-1564.

13. Pan, et al., "High Q Multi-Mode Resonator Controlled Source," U.S. Patent 4,245,193, January 1981.

Table 1. Comparison of Key Characteristics of Resonators for Use in Low Noise Oscillators

	"Ideal" Resonator	Quartz BAWR	Quartz SAWR	Ceramic DR	Sapphire DR
Q Factor i.e., Qu•f	HIGH	$1X10^{13}$	$5X10^{12}$	$5X10^{13}$	$2X10^{15}$ (T=300K) $2X10^{17}$ (T=77K)
Operating Frequency	HIGH	200MHz (max)	1GHz (max)	1GHz (min)	5GHz (min)
Drive Level Capability	HIGH	0.01W	0.1W	1W	1W(T=300K) 0.2W(T=77K)
Resonator Sy(f)	NIL	$\frac{5X10^{-24}}{f}$	$\frac{5X10^{-23}}{f}$	NIL	NIL
Vibration Sensitivity	LOW	$5X10^{-10}$	$5X10^{-9}$	$10^{-7}/g$*	$10^{-7}/g$*

* Typically obtained sensitivity for DRs operated on a low order (TE01) mode. Higher order mode designs should be less sensitive to vibration (i.e., to DR-to-enclosure wall dimensional change).

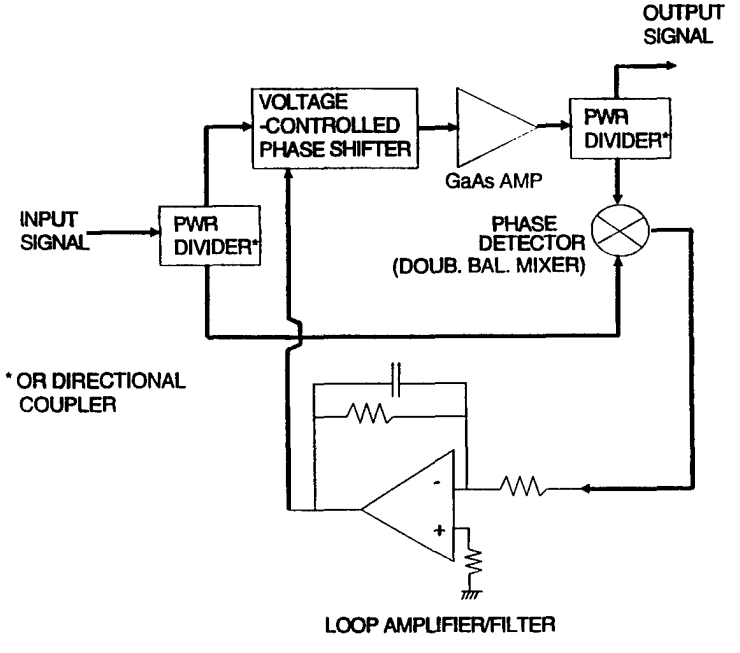


Figure 1. Wide Bandwidth, GaAs Amplifier Phase Noise Reduction Feedback Technique

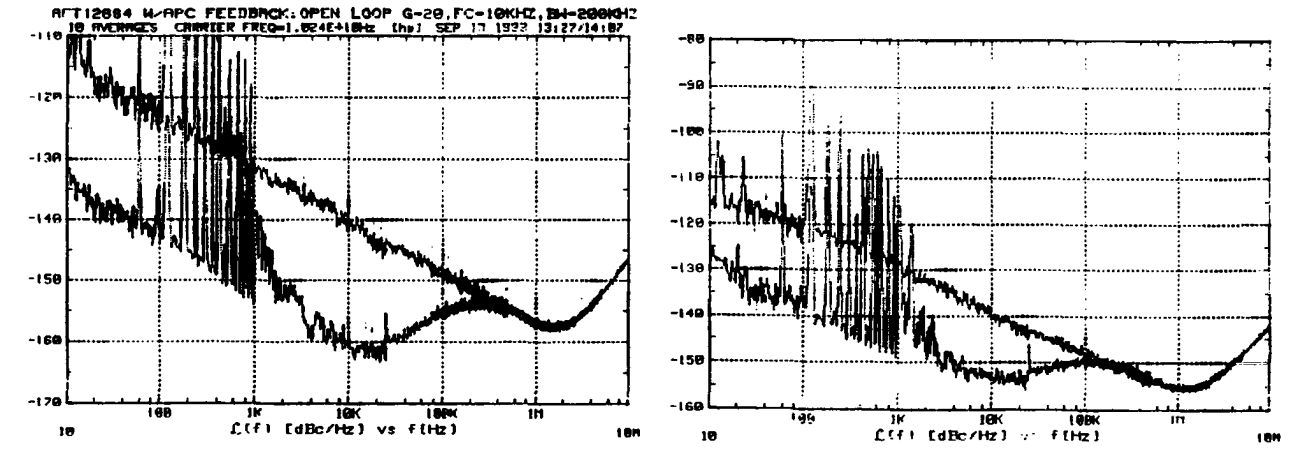


Figure 2. Measured Results: GaAs Amplifier Flicker-of-Phase Noise Reduction Using Feedback

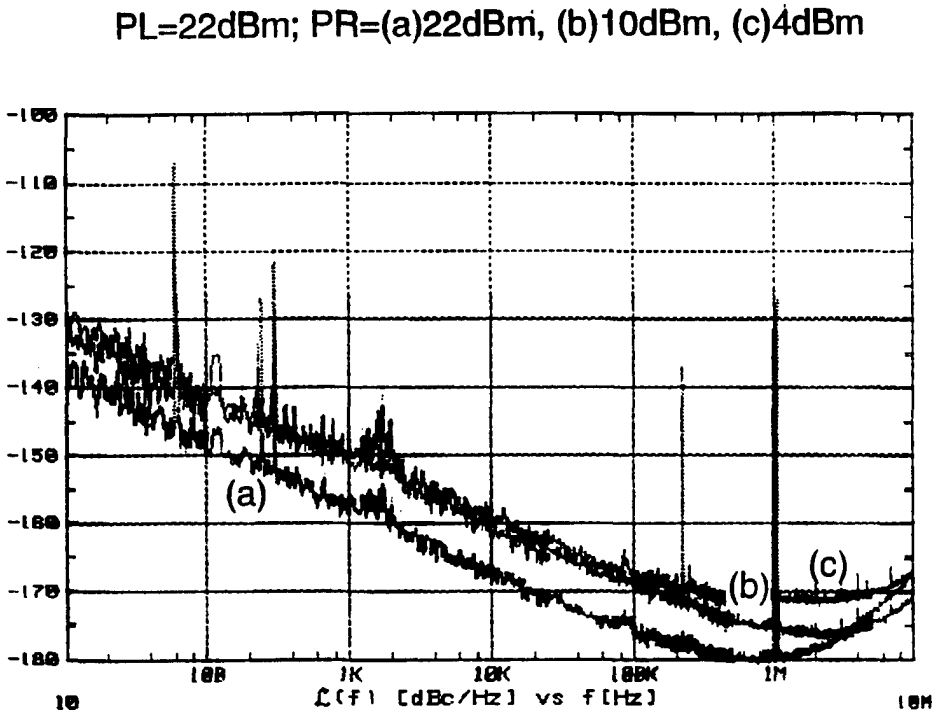


Figure 3. Measurement Results: WJM79HC, X-Band, Schottky Diode Mixer 1/f Noise vs. R Port Drive Level

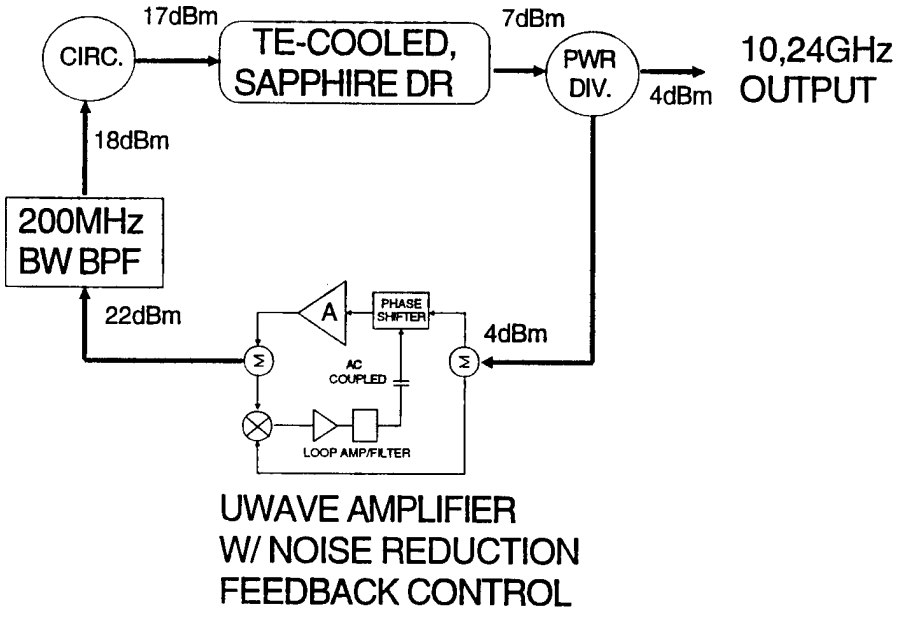


Figure 4. TE-Cooled, Sapphire DR-Controlled, X-Band Oscillator Block Diagram

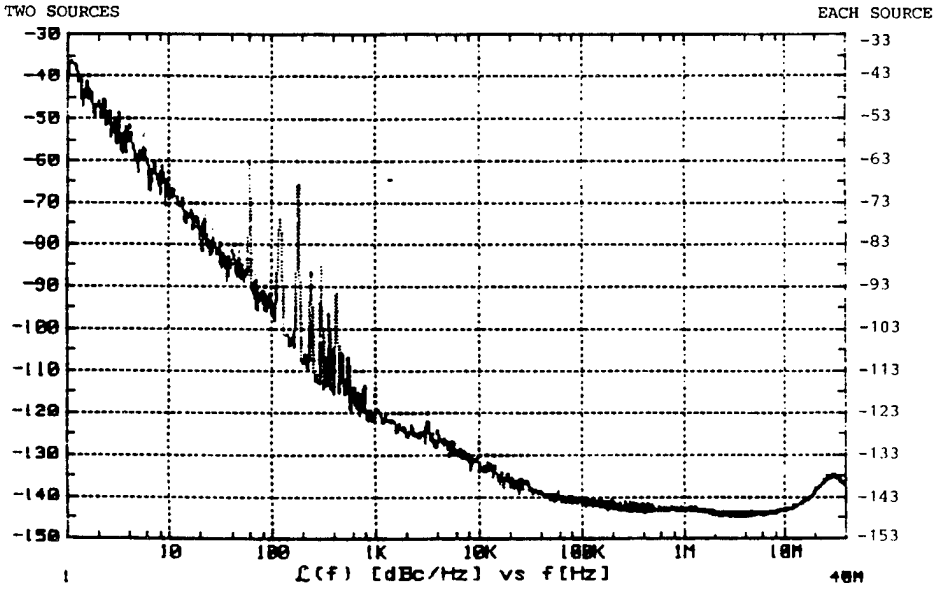


Figure 5. Measurement Results: Reference Signal Generator (10.24GHz) Signal Phase Noise Spectrum

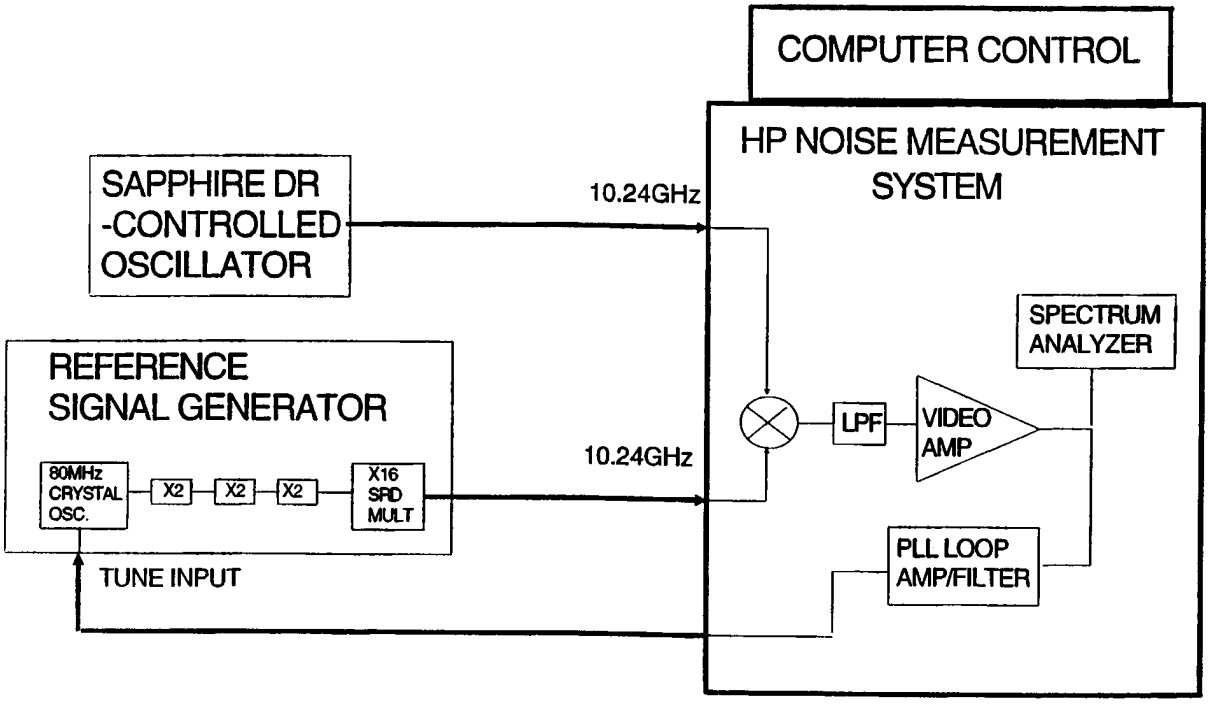


Figure 6. Oscillator Signal Flicker-of-Frequency Noise Measurement Test Set-Up

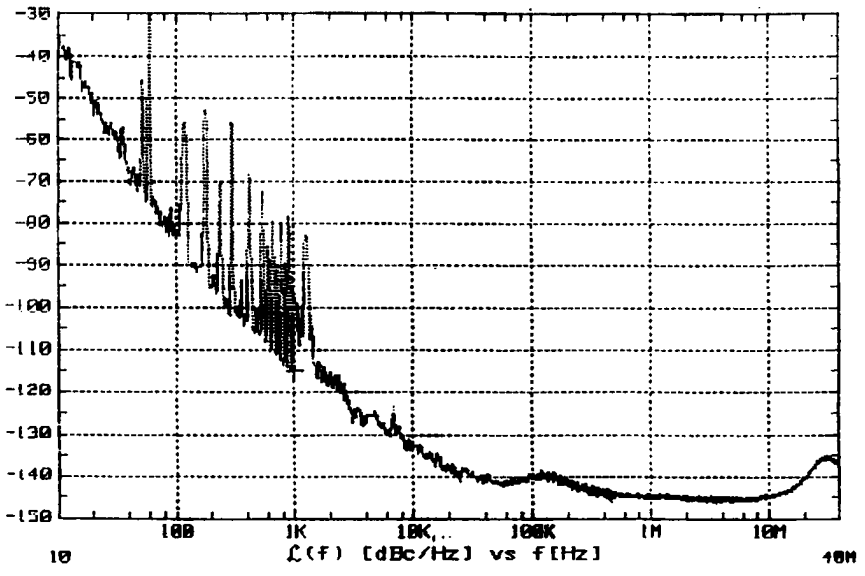


Figure 7. Measured Results: TE-Cooled, Sapphire DR-Controlled Oscillator Signal Flicker-of-Frequency Noise

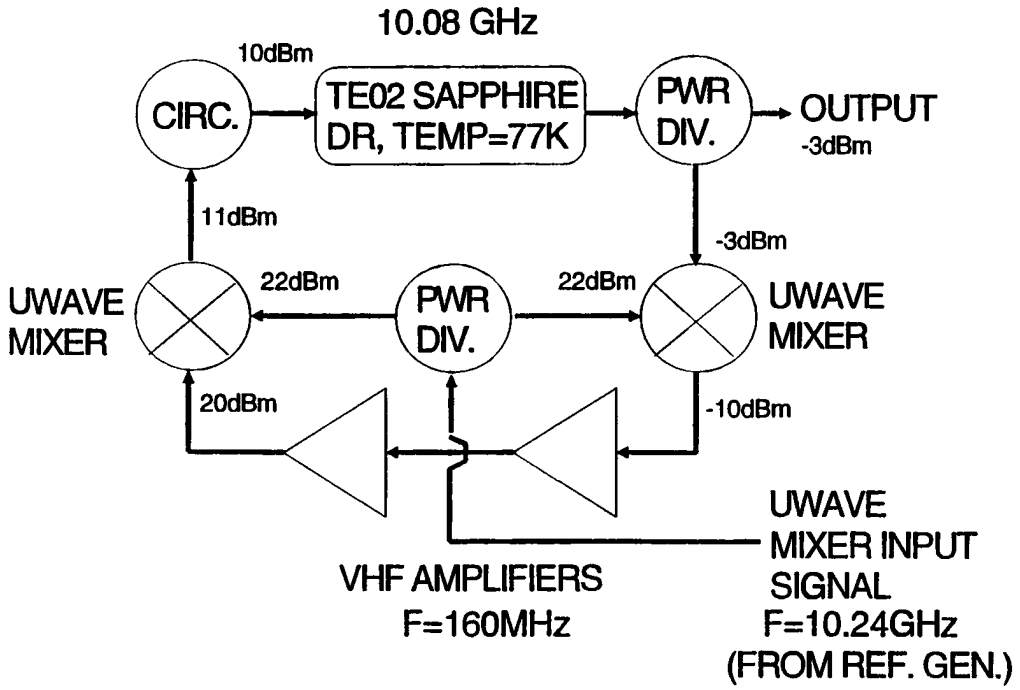


Figure 8. X-Band, Sapphire DR-Controlled Oscillator Incorporating a Double Frequency Conversion Type Sustaining Stage

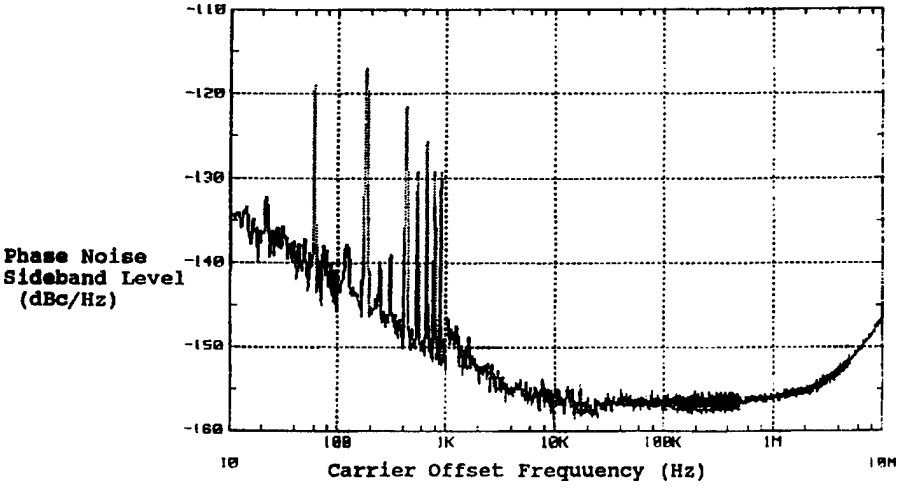


Figure 9. Measurement Results: Double Frequency Conversion, Oscillator Sustaining Stage Open Loop Flicker-of-Phase Noise

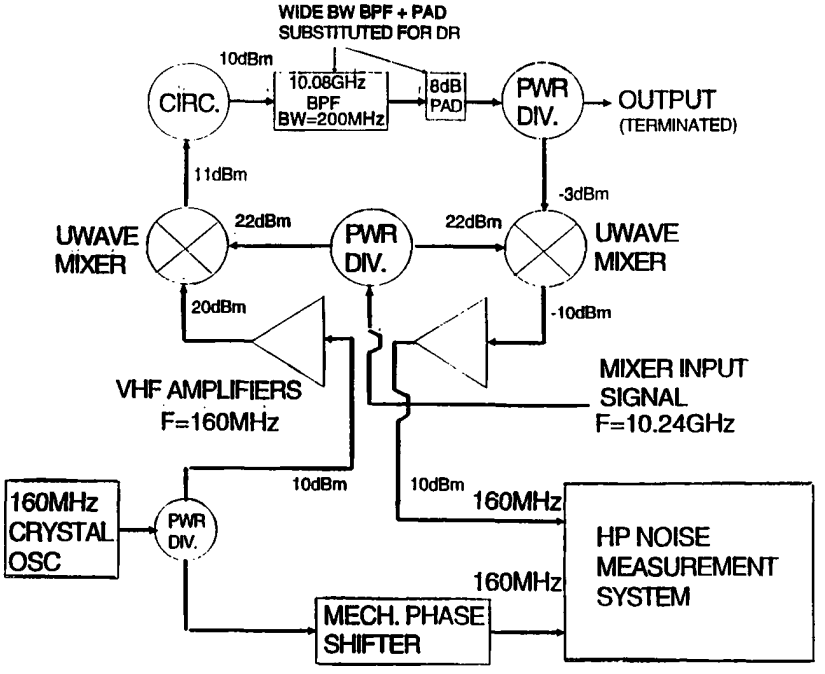


Figure 10. Test Set-Up for Measurement of Double Frequency Conversion Oscillator Sustaining Stage Open Loop Flicker-of-Phase Noise

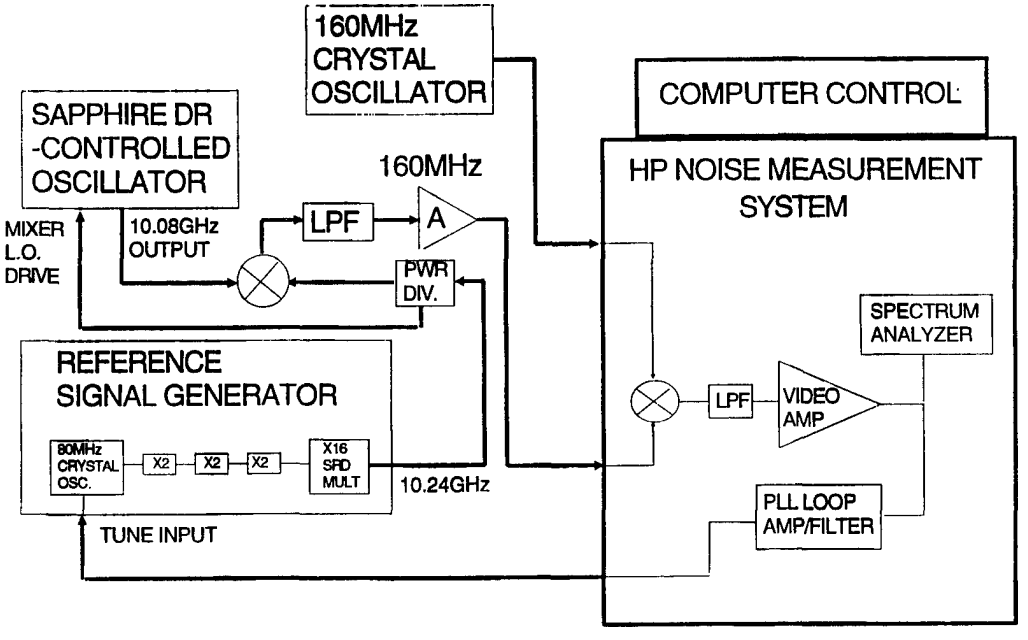


Figure 11. Flicker-of-Frequency Noise Measurement Test Set-Up for 10.08GHz (Double Frequency Conversion Sustaining Stage) Oscillator

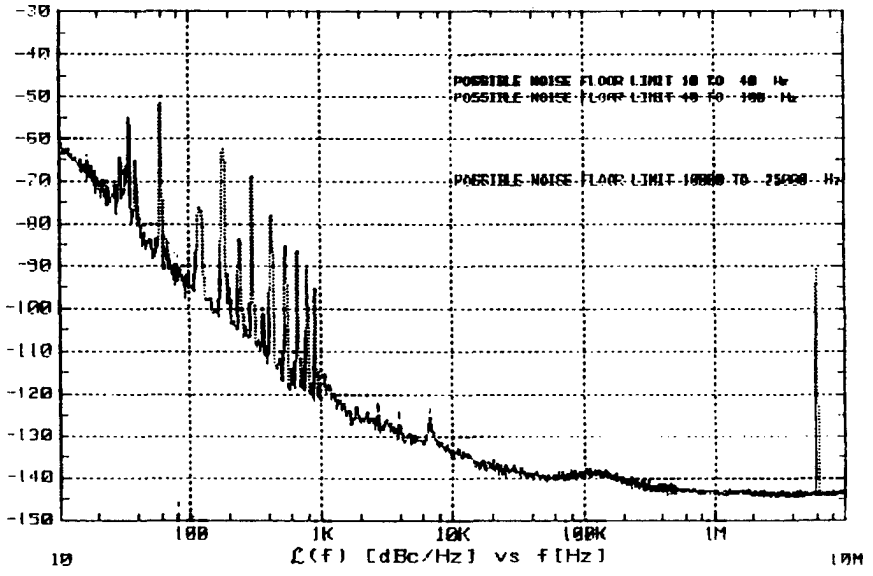


Figure 12. Measurement Results: 10.08 GHz Oscillator Signal Flicker-of-Frequency Noise

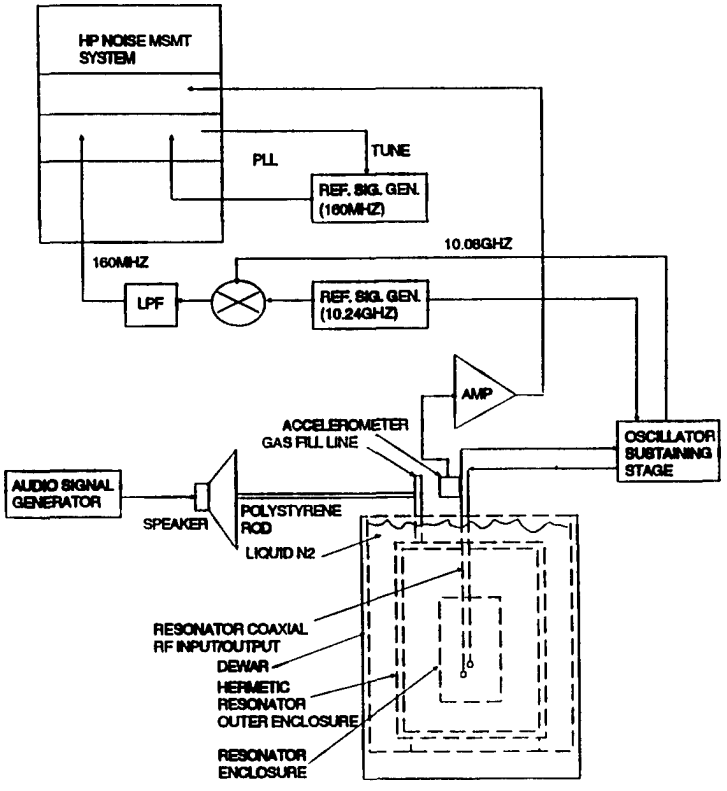


Figure 13. Vibration Sensitivity Test Set-Up with Resonator Immersed in Liquid Nitrogen Bath

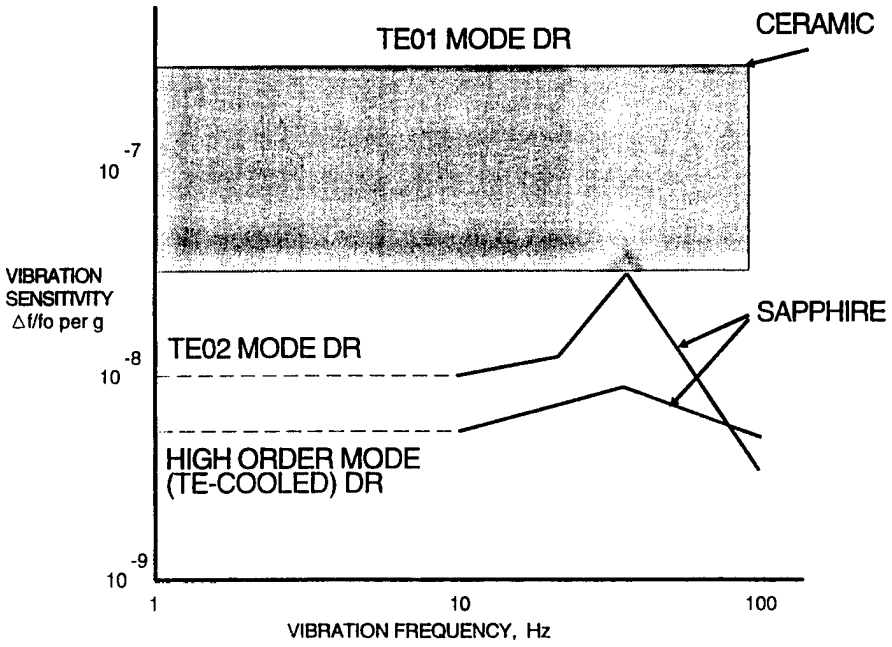


Figure 14. Sapphire DR Vibration Sensitivity Measurement Results

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A NEW METHOD OF FREQUENCY STABILIZATION

Dmitry P.Tsarapkin

Radio Engineering faculty, Moscow Power Engineering Institute
(Technical University), 14 Krasnokazarmennaya, Moscow 111250, Russia
fax: (7-095)361-1620; E-mail: rpdu@mpei-rt.msk.su

Abstract: This paper is devoted to a new method of achieving of higher spectral purity in an oscillator output signal, based on features of harmonic interactions in two-frequency oscillators. To attain this goal, the impact of output reactance instability of an active device is neutralized for a generated frequency. This new approach can be used together with traditional methods of frequency stabilization.

Introduction

Self-oscillation period instability posses severe limitations when determining technical characteristics of many modern electronic systems, prevents a number of precise physical experiments and restricts our ability to measure time. That is why for many years this topic has stayed an object of intensive investigations.

The known frequency stabilization methods exploit two basic ideas: improving stability of active device parameters and the increasing of circuit stability and Q-factor. Today's methods, are quantum sources and microwave oscillators with cryogenic "whispering gallery" mode sapphire resonators [1]. The paper proposed is devoted to another principle of achieving higher spectral purity in an oscillator output signal and is based on features of harmonic interactions in some multifrequency oscillators.

The classic nonlinear oscillation theory tends to consider the presence of higher harmonics of the fundamental frequency as an additional source of instability [2-4]. However, this statement, true in general, may have an exception when dealing with a specially designed biharmonic (two-frequency) oscillator, where an active device is loaded by two tuned circuits with resonant frequencies which are close to multiples of each other. With the correct tuning of the biharmonic oscillator, it will be shown it is possible to reduce, neutralize or even reverse generated frequency dependence on the output capacitance of an active device. For neutralization we suppress the basic source of 1/f noise, and for the reverse case the output frequency rises when the shunting capacitance increases. It goes without saying such behaviour would be an absolute "heresy" in respect of our usual professional practice.

Main equations

Initial analysis of the biharmonic oscillator under consideration was performed by G.M. Utkin [5]. Following him, let us keep in mind an oscillator equivalent network, of a one-pole active device (AD) loaded with two series connected tuned circuits, each tuned circuit comprising of a parallel L-C-R element (figure 1).

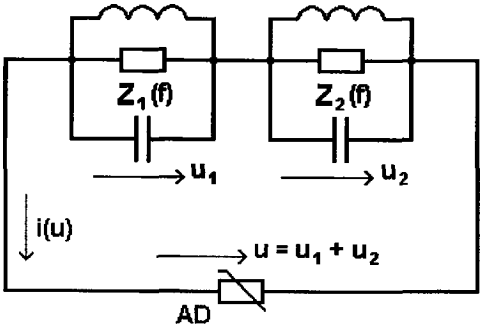


Fig. 1. Equivalent network of a biharmonic oscillator.

In the steady-state regime an alternating voltage u(t) at AD is a sum of two sinusoidal components, $u_1(t)$, $u_2(t)$, with multiple frequencies, f_1 , $f_2 = nf_1$:

$$u(t) = u_1 + u_2 \ . \qquad (1)$$

Usually the second harmonic of the AD current is the most pronounced one. So, put $n = 2$, $f_2 = 2f_1$.

The volt-ampere characteristic (VAC) of the active device around an operating point, $i = i(u)$, can now be approximated with the third-order polynomial

0-7803-2500-1/95 $4.00 © 1995 IEEE

$$i(u) = -[\alpha u + \beta u^2 - \gamma u^3], \quad (2)$$

where α, $\gamma > 0$, and β can have any sign. Furthermore, we put $\beta > 0$. The coefficient α in this expansion is a small signal transconductance.

Resonators' shortened impedances, $\mathbf{Z}_1(f)$, $\mathbf{Z}_2(f)$ have the form

$$\mathbf{Z}_i = R_i/(\delta_i + T_i\mathbf{p}), \quad (3)$$

where $R_i = \rho_i/Q_i$ – a resonant impedance, $\rho_i = (L_i/C_i)^{1/2}$ – a characteristic impedance, Q_i – a loaded Q–factor, $\delta_i = 1/Q_i$ – a damping factor, $T_i = 2Q_i/\omega_{oi}$ – a resonator time constant, $\omega_{oi} = 2\pi f_{oi} = (L_iC_i)^{-1/2}$ – a self–resonant angular frequency, $\mathbf{p} = d/dt$ – a differential operator with respect to time; in a steady–state regime $\mathbf{p} = j\Omega$, where $\Omega = 2\pi F$ – an angular offset frequency, $F = f - f_{oi}$ – a frequency shift in hertz with respect to current frequency f, $i = 1, 2$.

The self–resonant frequencies, f_{o1}, f_{o2}, are about multiple each other, i. e. $f_{o2} \approx 2f_{o1}$. Each of the two partial oscillations may be written as

$$u_i = U_i \cos\Psi_i,$$

where its instantaneous phase

$$\Psi_i = \omega_{oi}t + \psi_i = 2\pi f_{oi}t + \psi_i \quad (i = 1, 2). \quad (4)$$

To characterize a mutual phase shift between the harmonics' vectors it is suitable to define the *extended phase difference* (EPD), φ, putting

$$\varphi = \Psi_2 - 2\Psi_1 = \Delta\Omega_{12}t + \psi_2 - 2\psi_1, \quad (5)$$

where $\Delta\Omega_{12} = 2\pi\Delta f_{12} = 2\pi(f_{o2} - 2f_{o1})$ is an initial detuning of the resonators.

With these designations the shortened differential equations of the biharmonic oscillator take the form

$$\tau_1\frac{dX_1}{dt} = \left(A_1 - X_1^2 - 2X_2^2 + BX_2\cos\varphi\right)X_1,$$

$$\tau_2\frac{dX_2}{dt} = \left(A_2 - X_2^2 - 2X_1^2 + \frac{B}{2}\cdot\frac{X_1^2}{X_2}\cos\varphi\right)X_2, \quad (6)$$

$$\frac{d\varphi}{dt} = \Delta\Omega_{12}t + \left(\frac{d\psi_2}{dt} - 2\frac{d\psi_1}{dt}\right)\cdot\sin\varphi,$$

$$\frac{d\psi_1}{dt} = \frac{BX_2}{\tau_1}\sin\varphi, \qquad \frac{d\psi_2}{dt} = \frac{-B}{2\tau_2}\cdot\frac{X_1^2}{X_2}\sin\varphi.$$

Here $X_1 = U_1/U_m$, $X_2 = U_2/U_m$ – the amplitudes of the two spectral components normalized with respect of the maximal amplitude of oscillations in a single frequency regime, $U_m = (4\alpha/3\gamma)^{1/2}$;

$B = 2\beta(3\alpha\gamma)^{-1/2}$ – the *extended parameter of non–linearity*;

$A_1 = 1 - 1/\alpha R_1$, $A_2 = 1 - 1/\alpha R_2$ – the modified coefficients of regeneration; $\tau_1 = T_1/\alpha R_1$, $\tau_2 = T_2/\alpha R_2$ – the normalized time constants;

αR_1 and αR_2 – the regeneration factors.

All eqs. (6) include terms with the coefficient B. Hence, it follows that a quadratic term of VAC produces an essential effect on the setting processes and on the steady amplitudes and frequencies of the biharmonic oscillator under consideration.

Substituting the normalized instantaneous RF voltage, $x = u/U_m$, into eq. (2), we arrive at the normalized VAC

$$i_n(x) = i(u)/\alpha U_m = -[x + Bx^2 - (4/3)x^3], \quad (7)$$

Figure 2 illustrates typical positions of the operating point at the N–type VAC for $B > 0$.

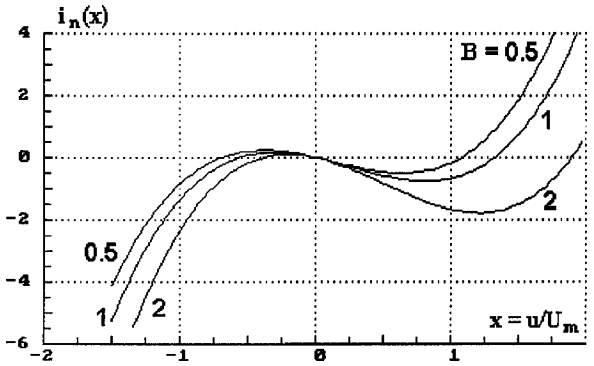


Fig. 2. Influence of the extended parameter of nonlinearity B at a form of the volt-ampere characteristic.

Steady−state regimes

In steady−state condition

$$dX_1/dt = dX_2/dt = d\varphi/dt = 0, \qquad (8)$$

and the derivatives $d\psi_1/dt$, $d\psi_2/dt$ represent the deviations of the generated frequencies from their resonant values. Namely,

$$d\psi_1/dt = \omega_{os} - \omega_{o1} = 2\pi(f_{os} - f_{o1}) \equiv 2\pi\Delta f_1 ,$$
$$d\psi_2/dt = 2\omega_{os} - \omega_{o2} = 2\pi(2f_{os} - f_{o2}) \equiv 2\pi\Delta f_2 . \qquad (9)$$

From eqs. (6), (8), rejecting trivial solutions $X_{1,2} = 0$, we obtain the system of equations describing steady−state regimes of the oscillator:

$$A_1 - X_1^2 - 2X_2^2 + BX_2\cos\varphi = 0,$$
$$A_2X_2 - X_2^3 - 2X_1^2 + \frac{B}{2}X_1^2\cos\varphi = 0, \qquad (10)$$
$$\Delta\Omega_{12}t - \left(2\frac{BX_2}{\tau_1} + \frac{B}{2\tau_2}\cdot\frac{X_1^2}{X_2}\right)\cdot\sin\varphi = 0.$$

The system (10) has different consistent decisions. Some typical situations for the case of $\varphi = 0$ are shown in figure 3.

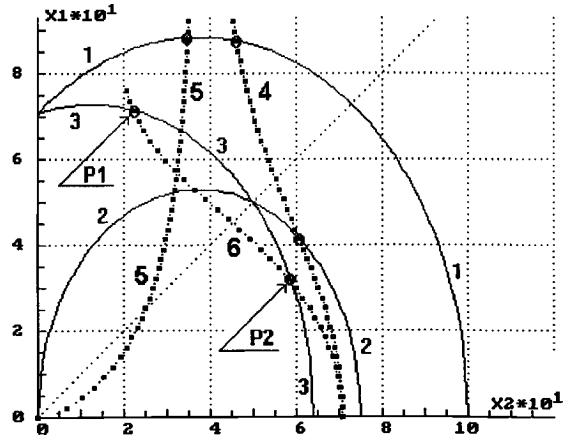


Fig. 3. Steady-state regimes of the biharmonic oscillator.
Curves 1...3 - $X_1(X_2, \alpha R_1, B)$, 4...6 - $X_2(X_1, \alpha R_2, B)$:
1 - $\alpha R_1 = 2$, $B = 1.5$; 2 - $\alpha R_1 = 1$, $B = 1.5$;
3 - $\alpha R_1 = 2$, $B = 0.5$;
4 - $\alpha R_2 = 2$, $B = 1.5$; 5 - $\alpha R_2 = 1$, $B = 1.5$;
6 - $\alpha R_2 = 2$, $B = 0.5$. $\varphi = 0$ for all curves.

Stationary points in figure 3 are marked by circles. For some particular sets of the parameters there is a single steady−state regime (represented, for example, by intersections of the curve 1 with the curves 4, 5). Some others (look at the curves 3, 6) may give two of them (points P1, P2). At last, the situation is possible when the curves $X_1(X_2)$ and $X_2(X_1)$ have no intersections at all. This takes place usually at $\alpha R_1 < 1$ or $\alpha R_1 \ll \alpha R_2$ and means mainly the generation of the only frequency f_2. Further we concentrate on the case $\alpha R_1, \alpha R_2 > 1$ as the most promising.

The investigation of stability of the steady states can be performed by usual means. Taking each variable Y_i as a sum of a steady value Y^o_i and a small variation δY_i from it, we arrive at a system of linear equations of the first approximation for disturbances:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \times \begin{pmatrix} \delta X_1 \\ \delta X_2 \\ \delta\varphi \end{pmatrix} = 0 . \qquad (11)$$

The elements a_{ij} of the matrix (A_{ij}) are first−order partial derivatives of the functions (10) with respect to the variables X_1, X_2, φ evaluated at the point of equilibrium. One can find:

$$a_{11} = -2X^2_1 ; \quad a_{12} = X_1(-4X_2 + B\cos\varphi);$$
$$a_{13} = -BX_1X_2\sin\varphi ; \quad a_{21} = a_{12} ;$$
$$a_{22} = A_2 - 3X^2_2 + a_{11} ; \quad a_{23} = -0.5BX^2_1\sin\varphi ;$$
$$a_{31} = -[BX_1/(\tau_2 X_2)]\sin\varphi ; \qquad (12)$$
$$a_{32} = -B[2/\tau_1 - X^2_1/(2\tau_2 X^2_2)]\sin\varphi ;$$
$$a_{33} = -B[2X_2/\tau_1 + X^2_1/(2\tau_2 X_2)]\cos\varphi .$$

The steady state is stable if all roots of a characteristic equation

$$\det(A_{ij}) = 0 \qquad (13)$$

have negative real parts. Calculations reveal that point P1 in figure 3 is stable while P2 one is not.

Principle of frequency stabilization

Let us study now the dependence of a generated frequency f_{os} on AD and resonator parameters. Some

data, obtained by solving eqs. (10), are shown in figure 4.

For our purpose, the most important feature of the curves in figure 4, is that they are not monotone. The origin of this is quite clear. A steady−state regime of

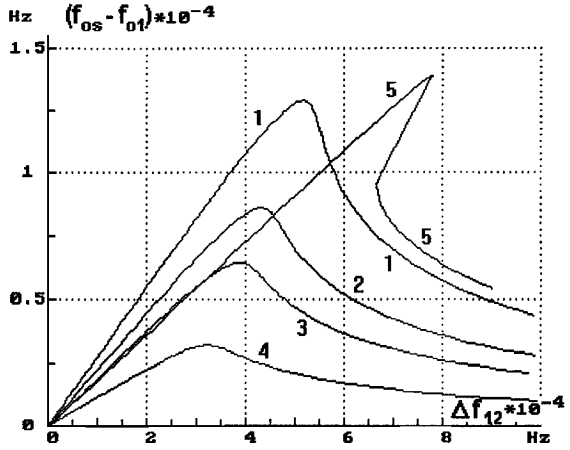


Fig. 4. Dependencies of a generated frequency on resonators' detuning. f_{o1} = 1 MHz. Curves 1...4: Q_2 = 100, αR_1 = αR_2 = 2, B =2; values of Q_1: curve 1 - 50; 2 - 75; 3 - 100; 4 - 200. Curve 5: αR_1 = 2, αR_2 = 2.5, B = 4, $Q_1 = Q_2$ = 150.

the device under consideration is characterized by mutual sychronization of two partial oscillations. An oscillation period depends on both resonant frequencies. The closer these frequencies to exact multiplicity the more intense their interaction. In result we see the strongest pulling at small values of modulus Δf_{12}. On the other hand, with large detuning the harmonic interaction becomes weaker due to decreasing resonator impedances. This fact is a consequence of necessity to keep the phase balance at both frequencies while the EPD value is going to ±90°. Curve 5 in figure 4 illustrates a hysteresis obtained by excessive influence of nonlinearity.

The case depicted is rather general and one can expect to meet the same phenomena in other types of biharmonic oscillators as well.

The curves in figure 4 give samples of the function $(f_{os} - f_{o1})$ with respect to the difference $(f_{o2} - 2f_{o1})$. Let us suppose now that there is some influence, say χ, which disturbs simultaneously both resonant frequencies, f_{o1} and f_{o2}. One obvious example of such influence is 1/f AD reactance fluctuations. Then the full change in generated frequency, $\Delta f_{os}(\chi)$, may be written as

$$\Delta f_{os} = \left[-\frac{\partial(f_{os} - f_{o1})}{\partial \Delta f_{12}} \frac{\partial f_{o2}}{\partial \chi} + \left(2\frac{\partial(f_{os} - f_{o1})}{\partial \Delta f_{12}} + 1 \right) \frac{\partial f_{o1}}{\partial \chi} \right] \Delta\chi . \quad (14)$$

It is easy to see that changes of f_{o1} and f_{o2} of the same sign produce just opposite impacts on Δf_{os}. Our goal is to minimize frequency instability. Thus setting Δf_{os} equal to zero, we obtain the *criterion of exact compensation* via a value of the parameter D_c :

$$D_c \equiv \left(\frac{\partial(f_{os} - f_{o1})}{\partial \Delta f_{12}} \right)_{comp} = \frac{1}{2 - æ} , \quad (15)$$

where $æ = \frac{\partial f_{o2} / \partial \chi}{\partial f_{o1} / \partial \chi}$ is a relationship of the sensitivities of self−resonant frequencies with respect to the disturbance. From figure 4 it follows that we may obtain any negative slope of the curve. So in principle, the mutual compensation of correlated frequency change is always possible if æ >2.

This statement is illustrated in figure 5, in which the curves repeat the conditions used in figure 4. The solid and dashed horizontal lines in figure 5 mark the levels of exact compensation for two particular values of æ.

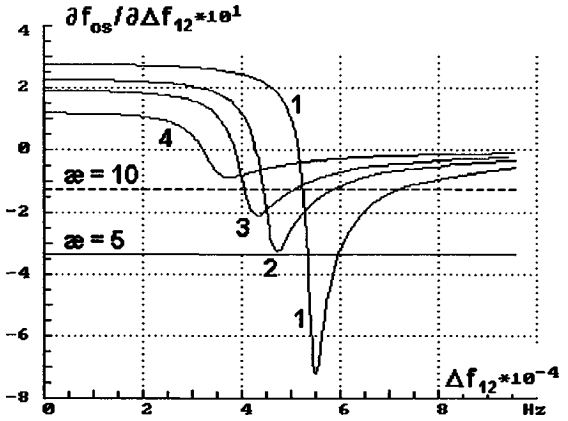


Fig. 5. Influence of a relative value of Q_1 on a slope of the curves in figure 4. Curve 1 - Q_1/Q_2 = 0.5; 2 - 0.75; 3 - 1; 4 - 2. Q_2 = 100.

Choosing the oscillator parameters, we can weaken, neutralize or even reverse a generated frequency dependence on tuning factor χ. Who's role, for example, could be AD output or mount capacitances. The corresponding data for conditions of figure 4 and æ = 7 are presented in figure 6 where for generality EPD was chosen as an argument of the curves.

The curve 3 demonstrates first−order suppression of the discussed frequency instability for EPD ranging between 60...73°. Curves 1, 2 have two extrema each spaced apart by an anomalous slope. When the os−cillator operates in this particular region its frequency increases with increasing shunting capacitance.

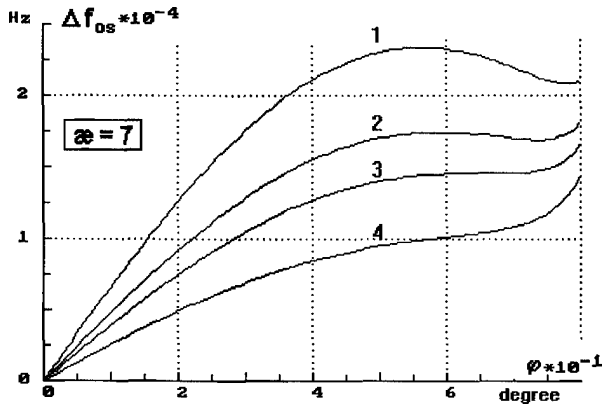


Fig. 6. Biharmonic oscillator frequency dependence on change of the enhanced phase difference due to simultaneous tuning of two resonators for conditions of figure 4.

Biharmonic oscillator realization

The equivalent network of a biharmonic oscillator in figure 1 is suitable for theoretical analysis but it is not convenient in practice. Primarily, this is due to difficulties in providing the needed values of æ. Therefore another oscillator circuit diagram, based on of a second−order resonant filter, was used instead (figure 7).

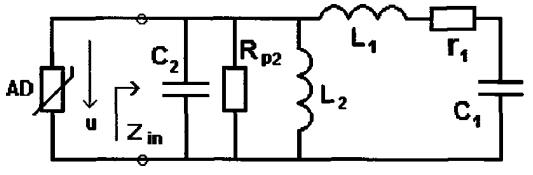


Fig. 7. Simplified biharmonic oscillator circuit diagram used in practice.

An input impedance of the filter, Z_{in}, has the form

$$Z_{in} = \frac{(f_{p2}/f)^2 \rho_{p2}(\delta_{p1} + j\nu_1)}{(\delta_{p1} + j\nu_1)(\delta_{p2} + j\nu_2) + (f^2/f_{p1}f_{p2})k_1k_2}, \qquad (16)$$

where: f_{p1}, f_{p2} − partial frequencies

$(f_{p1} = 1/2\pi\sqrt{(L_1+L_2)C_1}$, $f_{p2} = 1/2\pi\sqrt{L_2C_2})$;

δ_{p1}, δ_{p2} − partial damping factors;

$k_1 = L_1/L_2$, $k_2 = 1$ − coupling coefficients;

$\rho_{p2} = (L_2/C_2)^{1/2} = R_{p2}\,\delta_{p2}$;

$\nu_{1,2} = (f/f_{p1,2} - f_{p1,2}/f)$ − detuning.

A useful feature of the problem which helps to simplify many formulae consists in that two normal frequencies of the filter, f_{o1} and f_{o2}, which can be obtain from a biquadratic equation

$$\nu_1\nu_2 + (f^2/f_{p1}f_{p2})k_1k_2 = 0, \qquad (17)$$

must be almost multiple each other and f_{o2} twice as much f_{o1}. Bearing this additional condition in mind and denoting

$$\zeta = (f_{p2}/f_{p1})^2, \qquad (18)$$

it follows from eq. (17):

$$k_1 = \frac{17\zeta - 4(\zeta^2+1)}{25\zeta}, \qquad (19)$$

$$\frac{f_{o2}}{f_{p2}} = \sqrt{\frac{5}{\zeta+1}} \qquad (f_{o2} = 2f_{o1}). \qquad (20)$$

As the filter exploits reactive coupling its partial frequencies lie between natural (normal) frequencies. For our problem $\zeta \in [1, 4]$.

It is logical to take variations of capacitance C_2 as a cause of frequency instability. Then f_{p1} = const because it does not depend on C_2 and the normal frequencies are varied only due to changing of f_{p2}. Hence,

$$æ = \frac{\partial f_{o2}/\partial f_{p2}}{\partial f_{o1}/\partial f_{p2}} = \frac{8\zeta^2 + 6\zeta - 2}{-\zeta^2 + 3\zeta + 4}. \qquad (21)$$

To design an oscillator and evaluate its characteristics one needs to know equivalent resonator parameters at normal frequencies. Starting from eq. (16), we obtain

$$\frac{\delta_1}{\delta_{p1}} = \frac{\sqrt{5\zeta}}{6(\zeta+1)^{3/2}}\left[(4\zeta-1) + \frac{4-\zeta}{\sqrt{\zeta}}\cdot\frac{\delta_{p2}}{\delta_{p1}}\right], \qquad (22)$$

$$\frac{\delta_2}{\delta_{p2}} = \frac{\sqrt{5}}{3(\zeta+1)^{3/2}}\left[(4\zeta-1) + \sqrt{\zeta}(4-\zeta)\cdot\frac{\delta_{p1}}{\delta_{p2}}\right], \qquad (23)$$

$$R_1 = R_{p2}\frac{\delta_{p2}}{\delta_1}\frac{2(4-\zeta)}{3\sqrt{5(\zeta+1)}}, \qquad (24)$$

$$R_2 = R_{p2}\frac{\delta_{p2}}{\delta_2}\frac{4\zeta-1}{3\sqrt{5(\zeta+1)}}. \qquad (25)$$

Main properties of the filter as a part of the biharmonic oscillator are illustrated in figures 8, 9.

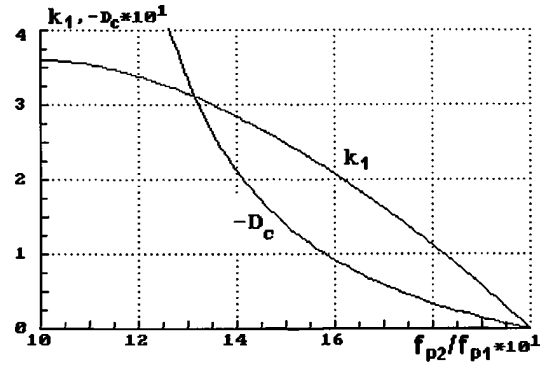


Fig. 8. Compensation parameter D_c and coupling coefficient k_1 vs partial frequencies' ratio.

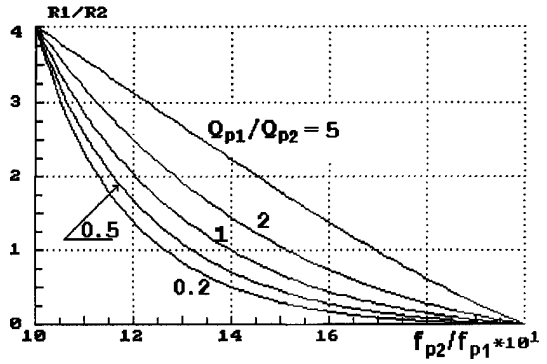


Fig. 9. Influence of a partial Q-factors' ratio on a resonant impedances' ratio.

One can see in figure 8 the parameter D_c goes to zero which means the compensation is getting easier when the f_{p2}, f_{p1} ratio tends to 2. It is a consequence of that partial circuits' interaction decreases while their coupling becomes weaker. The only snag here (figure 9) consist of parallel decrease of a resonant impedance at a fundamental frequency.

All principle positions of our theoretical analysis including the most important one – a possibility of the reversed oscillation frequency dependence on a shunting capacitance change – were confirmed experimentally. The RF model ($f_1 \approx 200$ kHz) repeats the circuit diagram in figure 7. To build a N–type AD the idea proposed in [6] was used. The detailed data will be publish later.

Noise analysis

The harmonic interaction modifies noise characteristics of a biharmonic oscillator in comparison with a single frequency one. To reveal main features of occurring phenomena an analysis for a case of fundamental noise was made.

An equivalent noise network (figure 10) differs from one drawn in figure 1 by presence of two noise sources describing white noise around corresponding resonant frequencies.

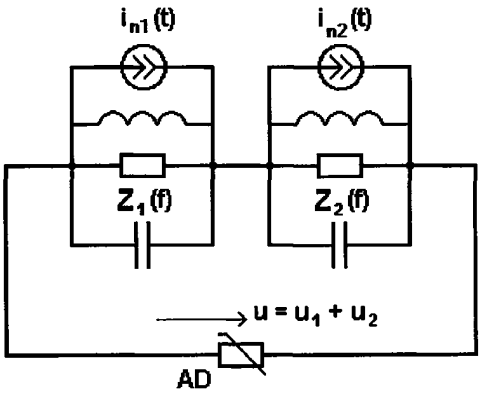


Fig. 10. Noise equivalent network of a biharmonic oscillator.

Each noise current in figure 10 can be written as a sum of in–phase (cosine) and quadrature (sine) components with respect to a voltage vector:

$$i_n(t) = I_{nc}\cos(\omega t + \psi) - I_{ns}\sin(\omega t + \psi). \qquad (26)$$

The noise currents play here the same role as virtual perturbations during a steady state stability analysis. Thus, basing on eqs. (6), one can obtain, for example, an expression for instantaneous fluctuations of the fundamental frequency, $\delta f_1(t)$:

$$\tau_1 \cdot 2\pi\delta f_1(t) = B\sin\varphi \cdot \delta X_2 + BX_2\cos\varphi \cdot \delta\varphi + n_{1s}/X_1, \quad (27)$$

where $\delta X_2(t)$, $\delta\varphi(t)$ – instantaneous noise deviations of corresponding variables, $n_{1s}(t) = I_{n1s}/\alpha U_m$ – a normalized value of this particular noise excitation.

Neglecting for simplicity by periodic non–stationarity of the noise sources, we can consider their power spectral densities as equal and put for normalized noise currents

$$S_{1c} = S_{1s} = S_{2c} = S_{2s} = S_n(F) = 2\cdot(2eI_0)/(\alpha U_m)^2. \qquad (28)$$

Dropping tedious intermediate transformations, we arrive at final formulae defining a relative value of the *frequency noise* in vicinity of the fundamental harmonic f_1:

$$\delta f_1(t) = (H_{11}n_{1c} + H_{21}n_{2c})\sin\varphi + H_{12}n_{1s} + H_{22}n_{2s}, \qquad (29)$$

where H_{11}, ... – appropriate coefficients. Hence, if EPD $\varphi \neq 0$, all noise sources together determine a total level of frequency noise.

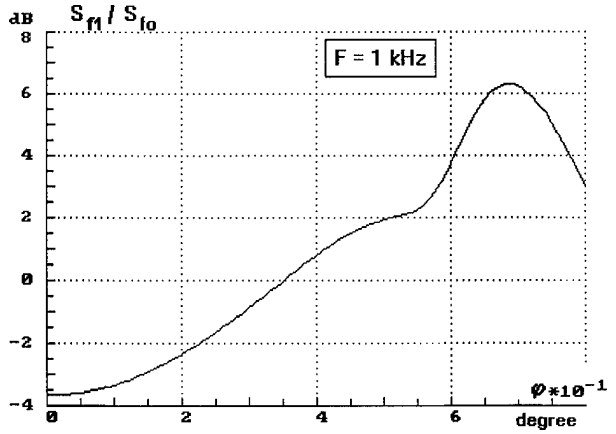


Fig. 11. Biharmonic oscillator frequency noise normalized with respect of single-frequency oscillator frequency noise vs EPD. Simulating conditions according curve 3 in fig. 6.

Figure 11 demonstrates a frequency noise power spectral density curve, scaled with respect to frequency noise S_{fo} of a single−frequency oscillator operating in the same conditions as a biharmonic one. Simulating conditions correspond to a case of almost ideal frequency instability suppression and offset 1 kHz. Comparing with figure 6, one can see that in the region of compensation ($\varphi \approx 60...73^{\circ}$) the white frequency noise level is about 10 dB greater than for synchronous resonator tuning.

The reason for such behaviour is getting clear from figure 12 where components of the resulting noise are presented. We can observe here rather strong amplitude noise influence through amplitude−phase conversion mechanism.

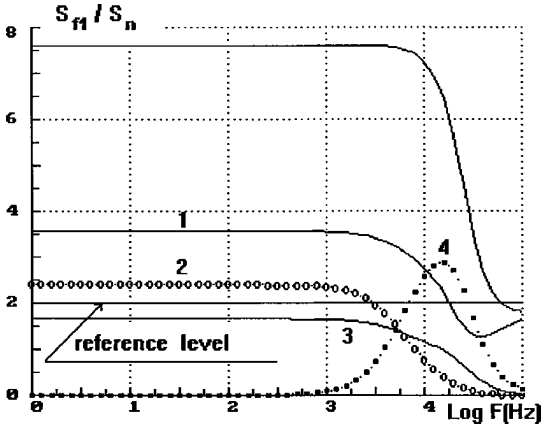


Fig. 12. Biharmonic oscillator frequency noise normalized with respect of noise excitation vs offset frequency. The reference level corresponds to single-frequency oscillator frequency noise. Simulating conditions according curve 3 in fig. 6 and $\varphi = 65^{\circ}$. 1 - a term due to I_{n1s}, 2 - due to I_{n1c}, 3 - due to I_{n2s}, 4 - due to I_{n2c}.

Conclusion

A new method of achieving of higher spectral purity in an oscillator output signal, based on features of harmonic interactions in biharmonic oscillators was developed. It is found that use of this new approach can decrease considerably influence of slow perturbations of an active device output reactance on a generated frequency.
The new approach discussed can be used together with traditional methods of frequency stabilization.

Acknowledgements

The author wish to thank Tom Hughes for his help on text editing.

References

[1] A.N. Luiten, A.G. Mann, M.E. Costa and D.G. Blair, "Cryogenic sapphire resonator-oscillator with exceptional stability: an update", in Proc. 48th Annual Freq. Contr. Symp., 1994, pp. 441-446.

[2] A. Blaquiere, Nonlinear system analysis. New York and London: Academic Press, 1966.

[3] V.V. Migulin, V.I. Medvedev, E.R. Mustel and V.N. Parygin, Basic theory of oscillations. Moscow: Mir Publishers, 1983 (translated from Russian).

[4] J.M.T. Thompson, Instabilities and catastrophes in science and engineering. Chichester and New York: John Wiley & Sons, 1982.

[5] G.M. Utkin, Self-oscillating systems and their wave analogues. Moscow: Soviet Radio, 1978 (in Russian).

[6] I.W. Stanley and D.Y. Ager, "Two-terminal negative dynamic resistance", Electron. Lett., vol. 6, N 1, pp. 1-2, 1970.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

AN ULTRA-LOW NOISE MICROWAVE OSCILLATOR BASED ON A HIGH-Q LIQUID NITROGEN COOLED SAPPHIRE RESONATOR

R.A. Woode, M.E. Tobar, E.N. Ivanov
Department of Physics, University of Western Australia, Nedlands, WA, 6009, Australia.

Abstract: Two liquid nitrogen cooled Sapphire Loaded Cavities (SLC) operating at about 80K have been successfully constructed. Both resonators were designed to operate on the Whispering Gallery (WG) $E_{12,1,\delta}$ mode at a resonant frequency of 8.95 GHz. The highest unloaded Q-factor measured was approximately 60 million for the first SLC and 40 million for the second. The first was used as the frequency determining element in a loop oscillator, while the second was used as a frequency discriminator to measure oscillator phase noise. The single sideband phase noise of a free running loop oscillator incorporating the first SLC was measured as -133 dBc/Hz at an offset frequency of 1 kHz, and was limited by the SLC Q-factor and the amplifier flicker phase noise. By using specially designed feedback electronics the oscillator phase noise was reduced to -156 dBc/Hz at 1 kHz. This measurement was shown to be limited by the electronic flicker noise imposed by the phase detector in the feedback electronics. Optimising the oscillator further, we measured by self discrimination a phase noise of -162 dBc/Hz at 1 kHz. Further analysis reveals that this type of liquid nitrogen cooled oscillator has the potential to reach -177 dBc/Hz at 1 kHz.

Introduction

The requirements for low noise and high stability microwave oscillators is now well recognised as an area in which sapphire dielectric resonator technology excels [1-11]. The performance of such systems is based primarily on the high-Q-factor of sapphire, which is greatly enhanced at cryogenic temperatures. At liquid helium temperatures (4.2K), Q-factors greater than 10^9 have been measured. Oscillators configured with these resonators have achieved Allen frequency deviations of 8×10^{-16} at 50 seconds integration time [1], and single sideband phase noise of -140dBc/Hz at 1 kHz offset frequency [3]. A major drawback in helium systems are their size and cost. Liquid nitrogen systems are smaller and less expensive to maintain. The Q-factor of sapphire measured at liquid nitrogen temperatures (77K) is degraded when compared to helium, but is still significantly better than the Q-factor measured at room temperature [4]. The highest Q-factor we have measured in a sapphire resonator at 77K is approximately 6×10^7.

In this paper we discuss the construction of two high Q-factor liquid nitrogen cooled SLC resonators. The first (SLC-1) was used as the frequency determining element in an ultra-low phase noise microwave loop oscillator. The second (SLC-2) was implemented as a frequency discriminator to measure the phase noise of the oscillator constructed with SLC-1. The oscillator phase noise levels detectable by our discriminator, were below the mixer limited noise floor for a typical two oscillator phase noise measurement.

The Sapphire Loaded Cavity Resonators.

The resonators were designed to operate on the $E_{12,1,\delta}$ mode at 8.95 GHz [5]. To obtain the high level of field confinement necessary to exploit the low loss tangent of sapphire, it is necessary to operate on WG modes with high azimuthal numbers. The losses then associated with the metal cavity and radiation are minimised.

The sapphire is housed inside a copper cavity within a system of evacuated cans as shown in Fig1. The resonator is then heat sunk to the nitrogen bath through a large thermal impedance. Passive temperature stabilisation was achieved due to the thermal mass of the resonator filtering out rapid temperature fluctuations in the bath. For active temperature control and tuning, the resonator has been fitted with platinum and thermistor temperature sensors and a heater. The measured temperature calibration for a typical thermistor is shown in figure 2. The sensitivity of the thermistor in percentage resistance change per K is in the order of 10 times more than a platinum thermometer, and is very effective for use at nitrogen temperatures.

Using this sensor, the frequency-temperature dependence of the resonator was measured as -115 kHz/K, as expected [5]. The passive temperature stability was measured to be better than 50μK at 10 seconds integration time. This means the resonator has a fractional frequency stability of the order of 10^{-10}. Thus with suitable temperature control using the thermistor sensors, it should be possible to control the long term fractional frequency stability to better than 10^{-10} at longer integration times.

0-7803-2500-1/95 $4.00 © 1995 IEEE

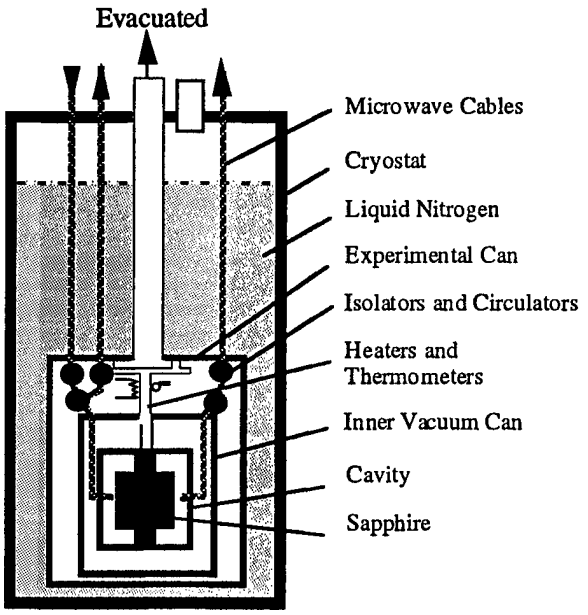


Figure 1. General schematic of a liquid nitrogen cooled sapphire loaded cavity and the cryogenic system.

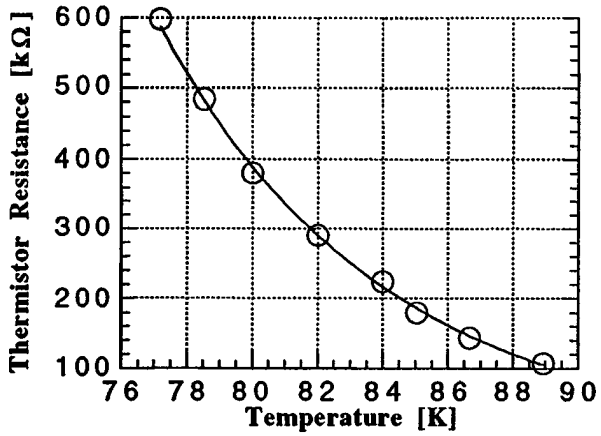


Figure 2. The measured temperature sensitivity of a typical thermistor used as thermometers for the nitrogen cooled SLC's.

In order to do any phase noise measurements, the resonant frequencies of the SLC's must be closely matched. The resonant frequency of SLC-1 was measured to be approximately 700 kHz less than SLC-2 at 77K. To tune the frequencies SLC-2 was heated to approximately 83 K.

Resonator Couplings and Q-factors:

The coupling to the resonators was accomplished with two, fixed magnetic field probes. The coupling was set low at room temperature to allow for the increase in Q-factor on cooling down. On cooling a doublet appears, as the resonance bandwidth becomes less than the doublet spacing of 5 kHz. The couplings and Q-factors of these doublets can be different. We report values associated with the highest Q of the doublet pair. Figure 3 shows the Q values measured in SLC-1 for the fundamental E mode family.

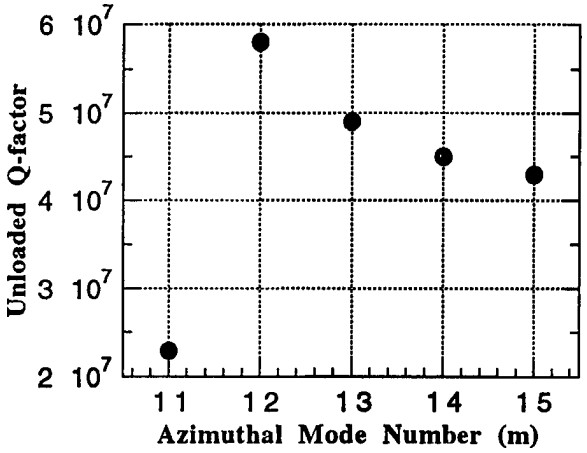


Figure 3. Unloaded Q-factors for the $E_{(m)\,1\,\delta}$ mode family in SLC-1.

The couplings measured were $\beta_1 = 0.78$ and $\beta_2 = 0.30$ for the operational $E_{12,1,\delta}$ mode at 8.95 GHz. The unloaded Q-factor of 6×10^7 is to the authors' knowledge the highest measured in sapphire at 77K. The unloaded Q-factor of the $E_{11,1,\delta}$ mode was degraded due to over coupling. As Expected the Q-factor of the higher order modes decrease roughly in proportion to the increase in frequency due to the intrinsic losses in sapphire. The unloaded Q-factor measured in SLC-2 was 3.5×10^7 for couplings, $\beta_1 = 0.31$ and $\beta_2 = 0.45$. The lower Q-factor is probably due to some contamination of the second sapphire.

Oscillator Phase Noise Measured by Self Discrimination

A loop oscillator with noise cancellation was assembled using SLC-1 as the frequency determining element, a schematic diagram is shown in fig. 4. The phase noise cancellation system was implemented using the SLC in a phase bridge configuration, similar to that described previously [5,6,12,13].

Phase fluctuations in the loop oscillator are detected by the phase bridge containing mixer 1, then the error signal is amplified, filtered and fed back to a varactor phase shifter in the loop. The phase bridge and the resonator are collectively referred to as a 'discriminator'. The phase noise is cancelled down to the limit set by the discriminator noise floor, providing there is enough servo gain. The servo consists of a variable gain low noise amplifier, with multiple filters having a similar transfer function to a system described previously [5].

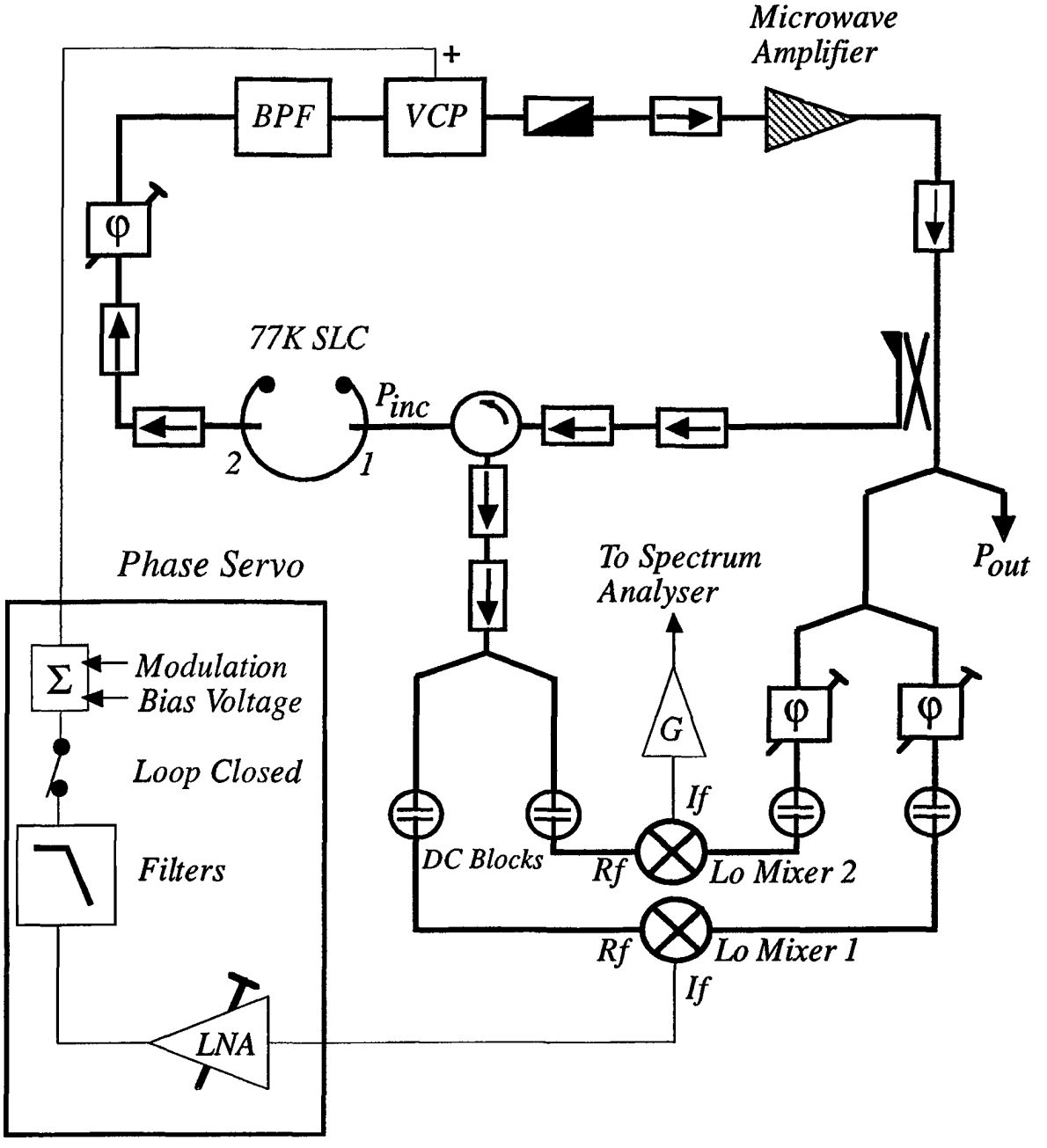


figure 4. Layout of the 8.95 GHz loop oscillator incorporating a liquid nitrogen cooled SLC with room temperature microwave and phase servo electronics. The mixers combined with the cavity are highly sensitive phase detectors. Phase noise cancellation is achieved by amplifying and filtering the error signal from mixer 1 and feeding it back to the voltage controlled phase shifter. In some cases a second mixer (2) was placed in parallel with mixer 1 for the purpose of monitoring the electronics component of the phase noise.

The sensitivity of the discriminator to frequency fluctuation is characterised as a frequency to voltage conversion ratio in V/Hz. When the frequency of the oscillator is tuned exactly to the resonance of the SLC cavity, for Fourier frequencies less than the SLC bandwidth, the discriminator conversion ratio is given by;

$$S_{Disc}^{DC} = k\sqrt{P_{inc}}\ \frac{2\beta_1^*}{1+\beta_1^*}\ \frac{2}{\Delta f_L} \qquad (1)$$

where k [V/$\sqrt{\text{Watt}}$] is the mixer power to voltage conversion ratio, P_{inc} [Watt] is the power incident on the cavity, β_1^* is the effective coupling into port 1 and Δf_L is the resonator loaded bandwidth [Hz].

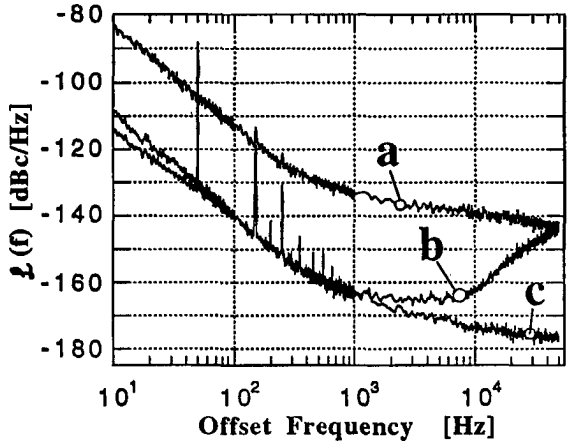


*Figure 5. Phase noise of the X-band SLC oscillator measured by the self discriminating technique using mixer 2 in figure 4. Curve **a** shows the single sideband phase noise of the free running loop oscillator. Curve **b** shows the phase noise when the servo has maximum gain. The closed loop phase noise follows the phase detector noise floor (**c**) until the noise cancellation servo looses gain.*

The typical DC conversion ratio measured for the discriminator in figure 4 is 3 V/Hz, for 30mW of incident power. The second mixer outside the loop is needed to monitor the electronic phase noise by a self discrimination technique. The drawback of this technique is that the noise measurement is insensitive to any correlated components of oscillator noise in the mixer Rf and Lo ports, such as vibration and cavity residual noise. The electronics single sideband phase noise was calculated using (2) below.

$$\mathcal{L}(f) = \frac{\left(\delta V_{rms}\right)^2 - \left(\delta V_{rms}^{NF}\right)^2}{2\left(G\, S_{Disc}^{DC}\, f\right)^2}\left[1+\left(\frac{2f}{\Delta f_L}\right)^2\right] \qquad (2)$$

where δV_{rms} is the measured spectral density of voltage noise, and δV_{rms}^{NF} is the noise floor of the measuring system. G is the low noise amplifier gain and f is the offset frequency.

The measured oscillator phase noise using self discrimination is shown in figure 5. The lowest phase noise measured was -162 dBc/Hz from 1 to 10 kHz when the servo was operating with its maximum gain. Because the self discriminating technique is insensitive to vibration and resonator residual noise, it was necessary to verify the phase noise level by another method.

Oscillator Phase Noise Measured by External Discrimination

This section describes the phase noise measurements of the oscillator constructed with SLC-1, by an external phase sensitive discriminator constructed with SLC-2. The schematic of the external discriminator is shown in figure 6. Due to lower couplings in SLC2 the sensitivity of the external discriminator to oscillator phase noise was not sensitive enough. To compensate the oscillator output power was increased by altering the power splitters in the loop, which provide a discriminator conversion ratio of 1.5 V/kHz. However increasing the oscillator output power is at the expense of reducing the incident power to SLC1 and hence the volts to frequency conversion ratio. Thus the oscillator phase noise was degraded to -156 dBc/Hz at 1 kHz offset frequency, confirmed by the self discrimination technique.

The discriminator was then tuned to the oscillator frequency by heating the resonator in SLC-2 to approximately 83K. The initial measured spectrum showed enhanced sensitivity to vibrations at low frequencies. After raising the oscillator and discriminator on suspension rubbers, there was marked improvement in the low frequency phase noise shown in figure 7. This noise was determined to be due to the vibration sensitivity of the mechanical phase shifters and the microwave circuit in general.

Some measured spectral densities of phase noise using an external discriminator are shown in figure 8. The anomalies, **e** and **f** are caused by the presence of the doublet resonance in the discriminator, and are discussed in more detail in figure 9. At 1 kHz offset the oscillator phase noise was 3 dB above the noise floor and therefore equal to the noise floor measured as -156 dBc/Hz.

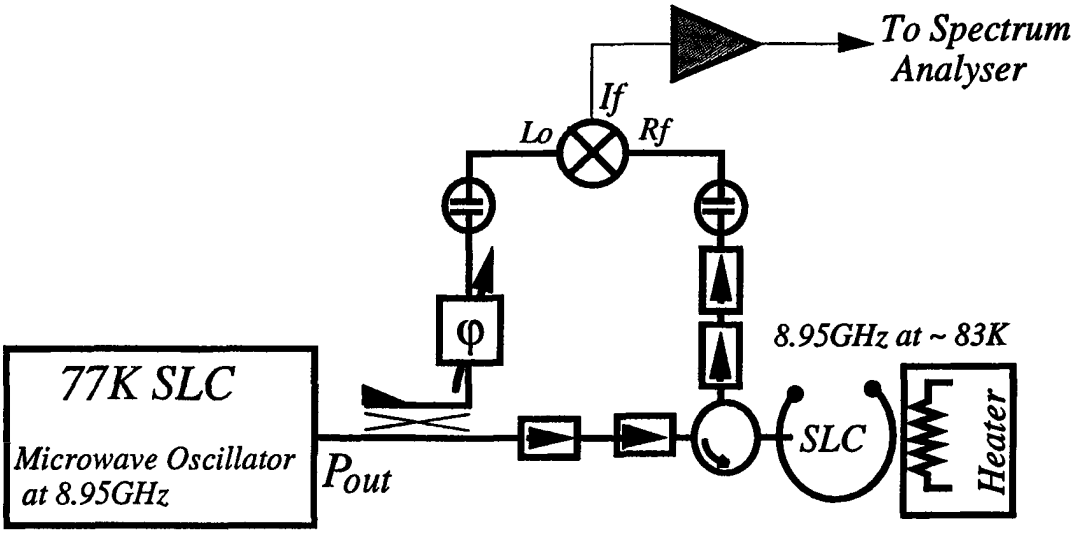


Figure 6. Schematic of the SLC discriminator phase bridge used to measure oscillator phase noise.

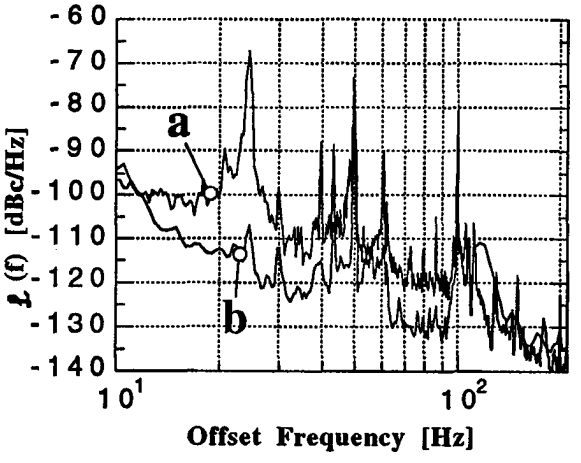


*Figure 7. Low frequency effects in the SLC oscillator: **a.** phase noise spectra without vibration isolation; **b.** phase noise spectra measured with the oscillator and discriminator mounted on rubber blocks.*

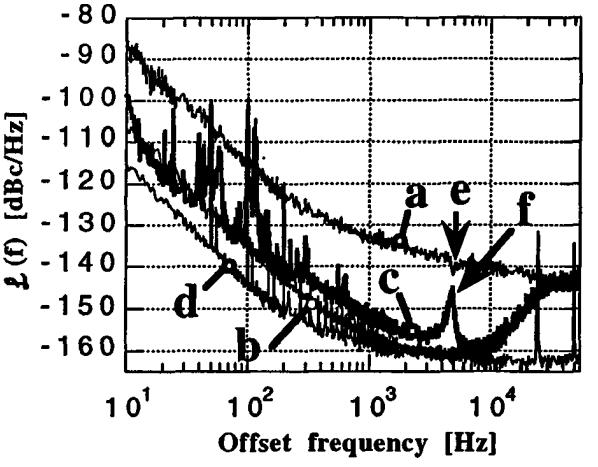


*Figure 8. The SLC oscillator single sideband phase noise measured with the external phase noise discriminator as shown in figure 6: **a.** Free running loop oscillator phase noise; **b.** Closed loop phase noise measured by the self discrimination technique; **c.** Closed loop single sideband phase noise. The low frequency peaks are due to external vibrations of the microwave electronics. The phase noise spectra follows the electronic phase noise component **(b)** closely until it runs into the SLC phase discriminator noise floor **(d)** where the phase noise is then 3 dB above the noise floor. The actual phase noise measured is therefore equal to the noise floor and is -156 dBc/Hz at 1 kHz offset. The $E_{12,1,8}$ mode in SLC-2 occurs as a doublet with about a 5 kHz spacing. The doublet causes the anomalies **e** and **f** in the spectra, which are discussed fully in figure 9.*

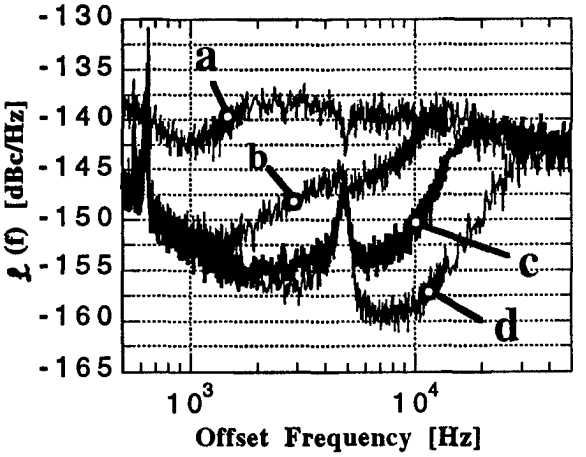


*Figure 9. The effect of increasing the servo loop gain on the phase spectra "anomaly". The dip and peaks are due to the presence of a sapphire doublet 5 kHz away from the discriminating mode. We emphasize that this effect is not due to the oscillator phase noise: **a**. The open loop plot shows a small dip because of reduced phase noise sensitivity. The servo loop is then closed and the amplifier gain increased through curves **b**,**c** and **d**. As the phase noise level drops a peak becomes visible and remains at the -147 dBc/Hz level. This peak is due to amplitude noise in the SLC oscillator. The doublet forces the external discriminator to have reduced PM noise sensitivity and enhanced AM noise sensitivity at this offset frequency.*

The phase noise spectral density measured by the self discriminator method as shown in figure 8, is in agreement with the phase noise spectral density measured by external discrimination above 500 Hz. Therefore we can conclude that the self discrimination technique is a viable method for measuring oscillator phase noise, provided the vibrational performance of the oscillator is known. For the 3 V/Hz oscillator configuration in figure 4, we can be confident that the same spectral density of phase noise would have been measured with an external discriminator, corresponding to -162 dBc/Hz from 1 kHz to 10 kHz (fig. 5). This level is close to what we predicted possible for this type of oscillator previously [5].

An upper limit to the residual noise of the 77K resonator was measured and shown in figure 10. The structure of the upper limit is defined by the driving oscillator phase noise.

There is potential for improvement in the performance of the liquid nitrogen cooled SLC oscillator. If the power incident on SLC-1 was increased to 100 mW, the phase detector sensitivity would be improved by a factor of 4 to 5. Further optimisation of the oscillator parameters and components should result in a phase noise floor of -177 dBc/Hz at 1 kHz offset. Figure 11 shows the predicted improved oscillator performance along with the results measured in this paper and the performance of other microwave oscillators.

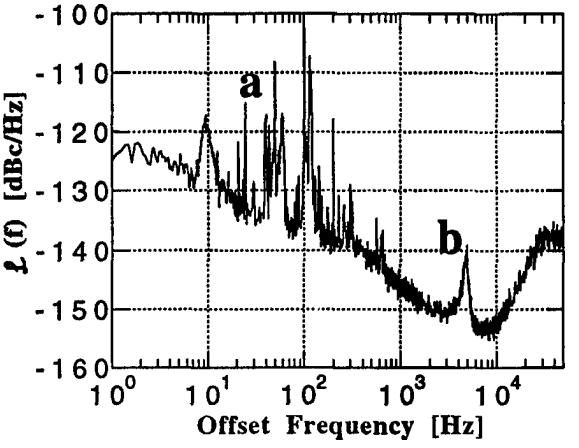


*Figure 10. The measured upper limit to the residual phase noise of a nitrogen cooled SLC resonator. The level is set by the driving oscillator phase noise: **a**. shows peaks due to vibration sensitivity of the oscillator; **b**. Anomaly due to the AM noise of the oscillator as shown in fig 9.*

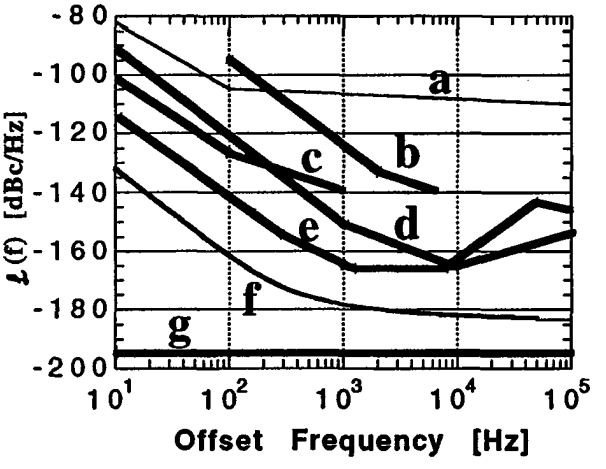


*Figure 11. Progress of phase noise measured in X-band sapphire microwave oscillators at UWA: **a**. Typical multiplied quartz; **b**. Room temperature SLC oscillator results, published in 1994 [5,6]; **c**. Liquid helium cooled SLC oscillator results, published in 1993 [14] and 1994 [3] ; **d**. Room temperature SLC oscillator results 1995, in these proceedings [2]; **e**. Liquid nitrogen cooled SLC oscillator results, 1995, in this paper; **f**. Predicted phase noise results for a further optimised liquid nitrogen cooled SLC oscillator; **g**. The thermal noise floor limiting the phase noise performance of a liquid nitrogen cooled SLC oscillator with 50mW of power incident on the resonator.*

Conclusion

We have measured a phase noise of -156 dBc/Hz at 1 kHz offset in a liquid nitrogen cooled SLC based oscillator using an external phase noise discriminator. Optimising this oscillator further, we measured using self discrimination a phase noise of -162 dBc/Hz from 1 kHz 10 kHz offsets. The noise level was limited by the phase detector noise floor in the noise cancellation servo below 1 kHz, and the microwave amplifier noise above 1 kHz. The vibration sensitivity of the oscillator at low frequencies can be improved further by improving the design and making the microwave electronics less sensitive to vibrations. With further optimisation of the liquid nitrogen cooled SLC oscillator we predict a phase noise performance of -177 dBc/Hz at 1 kHz.

Acknowledgments

The authors thank Poseidon Scientific Instruments for their general support and their making of the servo electronics. This project was supported by the Australian Research Council.

References

[1] A.N. Luiten et al, in these proceedings.

[2] E.N. Ivanov, M.E. Tobar and R.A. Woode, "Advanced phase noise suppression technique for next generation of ultra low noise microeave oscillators, in these proceedings.

[3] M.E. Tobar and D.G. Blair, "Phase noise analysis of the sapphire loaded superconducting niobium cavity oscillator," IEEE Trans. MTT., vol. 42, no. 2, Feb. 1994.

[4] M.E. Tobar, A.J. Giles, S.J. Edwards and J.H. Searls, "High-Q TE stabilized sapphire microwave resonators for low noise applications," IEEE Trans. Ultrason. Ferroelec. and Freq. Contr., vol. 41, May 1994.

[5] M.E. Tobar, E.N. Ivanov, R.A. Woode, J.H. Searls, "Low noise microwave oscillators based on high-Q temperature stabilized sapphire resonators," in Proc. 1994 IEEE Int. Freq. Contr. Symp. pp. 433 to 440.

[6] M.E. Tobar, E.N. Ivanov, R.A. Woode, J.H. Searls, A.G. Mann, "Low noise 9 GHz sapphire resonator-oscillator with thermoelectric temperature stabilization at 300 K," IEEE MWG let., May 1995.

[7] G.J. Dick and J. Saunders, "Measurement and analysis of a microwave oscillator stabilized by a sapphire dielectric ring resonator for ultra-low noise," IEEE Trans. Ultrason., Ferroelec., Freq. Contr., vol. 37, pp. 339-346, Sept. 1990.

[8] M.M. Driscoll et al., "Cooled ultra-high Q, sapphire dielectric resonators for low noise microwave signal generation," IEEE Trans. Ultrason. Ferroelec. and Freq. Contr., vol. 39, pp 405-411, May 1992.

[9] G.J. Dick and D.G. Santiago, "Microwave frequency discriminator with a cryogenic sapphire resonator for ultra-low phase noise," in Proc. 1993 IEEE Int. Freq. Contr. Symp. pp. x to x.

[10] D.P. Tsarapkin, "An uncooled microwave oscillator with 1-million effective Q-factor," in Proc. 1993 IEEE Int. Freq. Contr. Symp. pp. 779 to 783.

[11] C.A. Flory and R.C. Taber, "Microwave oscillators incorporating cryogenic sapphire dielectric resonators," in Proc. of the 1993 IEEE International Frequency Control Symposium pp 763-773.

[12] Z. Galani et al, IEEE Trans. Microwave Theory Tech., vol. 32, pp. 1556-1565, 1984.

[13] F.L. Walls, C.M. Felton, T.D. Martin, "High spectral purity X-band source," Proc. 44th Symp. Freq. Contr., pp. 542-547, 1990.

[14] M.E. Tobar, PhD thesis, University of Western Australia, 1993.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A NON-LINEAR CAD MODEL FOR THE REGENERATIVE DIELECTRIC RESONATOR OSCILLATOR

ALI A. BEHAGI, Ph.D.
Associate Professor, Penn State University - Harrisburg
777 W. Harrisburg Pike, Middletown, PA. 17057

STEPHEN D. TURNER, P.E.
MAXTECH, Inc.
2120 Old Gatesburg Road, State College, PA. 16803

ABSTRACT - Much has been written in recent years on the design and measurement of the regenerative dielectric resonator oscillator. This type of oscillator consists of an independent amplifier followed by a power divider that is connected to a feedback path. The feedback path consists of a two port dielectric resonator cavity and a phase shifter. The output of the phase shifter is then connected to the amplifier's input to complete the oscillator feedback loop. It has been demonstrated that this type of oscillator can achieve far superior phase noise performance over other designs in the microwave frequency range.

Specifically, this paper describes the process of design and analysis of a 5 GHz regenerative dielectric resonator oscillator. The oscillator is based on a multistage amplifier using the AT41435 packaged bipolar device. A two port dielectric resonator cavity has been assembled and characterized separately. A mechanical phase shifter is used in the loop to adjust the oscillator phase. LIBRA, a nonlinear CAD simulator by HP-EESOF, is used to perform both linear and nonlinear computer simulation.

1.0 INTRODUCTION

A detailed three step approach to the oscillator model is presented in the paper.

Step 1) Accurate characterization of a four stage bipolar transistor amplifier including gain, power and compression.

Step 2) Accurate characterization of a two port dielectric resonator cavity including measured S-parameters.

Step 3) Nonlinear analysis of closed loop high Q microwave oscillator.

The oscillator parameters under study include: oscillator start up, limiting and output power. The paper details the procedure of using the non-linear simulator to efficiently extract these important design parameters. As in any non-linear simulation of high Q circuits, there are many pitfalls that must be avoided for accurate results. The computer model results are then compared to the measured oscillator's performance.

It is the goal of this paper to help streamline the CAD model of the regenerative dielectric resonator oscillator. Non-linear simulation of high Q oscillators is often rather complex and time consuming. Once a reliable model of the basic regenerative oscillator is developed, the techniques can be extended to perform noise analysis of this popular low phase noise microwave oscillator.

As shown in Figure 1, the regenerative DRO consists of a multistage transistor amplifier, a two-port dielectric resonator cavity, a phase shifter, and a power splitter.

The dielectric resonator cavity used in the experiment is a two port rectangular cavity containing an adjustable height dielectric resonator. The coupling strips are arcs concentric to the dielectric resonator itself. Transition to the input and output SMA connectors is through the sides of the cavity using 50 Ohm microstrip lines.

The phase shifter consists of a mechanically adjustable length concentric coaxial lines connected to the output of the resonator to adjust the phase of the feedback signal .

2.0 LINEAR SIMULATION

The purpose of the linear simulation is to find a rough estimation of the oscillator startup conditions and the frequency of oscillation. To achieve this, the source is swept over a large frequency range and the power levels are recorded.

0-7803-2500-1/95 $4.00 © 1995 IEEE

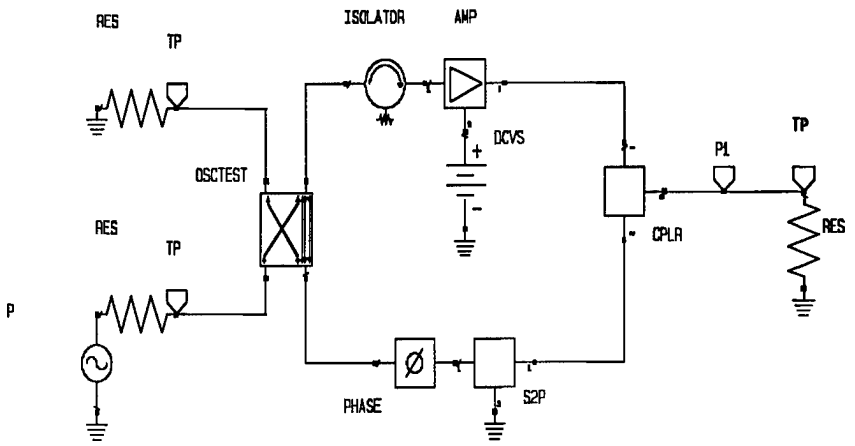


Figure 1. Oscillator Schematic prepared for Nonlinear Simulation

Figure 1 represents the transmission form of the DRO as required for nonlinear simulation. Note that the four stage bipolar amplifier is represented by one subcircuit block. The cavity resonator is represented by its S-parameter data file. The 10 dB directional coupler is also inserted in its subcircuit form in the final oscillator schematic. The oscillator simulation is described in three primary steps, beginning with the more familiar linear simulation.

The frequency of oscillation is obtained where the phase shift of the RF signal at the amplifier input crosses the zero point or the imaginary part of the input impedance becomes zero.

By defining R to be the ratio of the feedback voltage to the input voltage, the magnitude and the phase of R indicates when the conditions for oscillation are met.

From linear analysis it was found that the input power level for oscillation is about -2 dBm and the frequency of oscillation is about 5.2 GHz

When performing nonlinear simulation using the Harmonic Balance simulator, it is important to have the circuit tuned very close to its operating state. This is particular true of high Q circuits which can cause numerous convergence problems. One of the more efficient means of accomplishing this task is by using linear analysis to get the oscillator circuit close to its final state. In the regenerative oscillator this is done by manually tuning the phase shifter to peak the output reflection coefficient at the resonant frequency of the dielectric resonator cavity.

When observing the output of the oscillator of Figure 1, the designer must remember that there is an integral 10 dB loss due to the directional coupler. The coupler's loss may be normalized out in the software by using the **OUTVAR** and **OUTEQN** features in the software. This allows the oscillator's output signal to be measured directly at the output of the amplifier. The oscillator's output reflection coefficient and directional coupler's forward transmission are first defined as variables.

```
OUTVAR
      OSCR= OSC  DB[S11]
      COUP=CPLR  DB[S31]
OUTEQN
      OSCNORM = OSCR - COUP
```

It is usually best to begin with a wide frequency sweep range and large frequency step size for faster simulation speed.

```
FREQ
      5.10  5.30  .001   ! 1 MHz step size
```

The phase shifter is then manually tuned for maximum reflection coefficient in the vicinity of the cavity resonator frequency which in this case is 5.20 GHz. With the phase shifter set at 100 ° the output of Figure 2 is obtained. Achieving positive reflection coefficient at the output of the oscillator indicates the potential for oscillation. At this point the circuit is prepared for nonlinear simulation

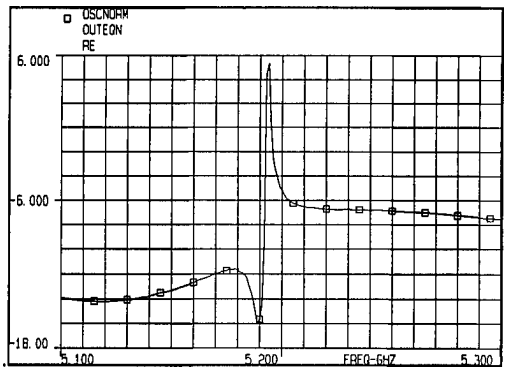


Figure 2. Oscillator Output Reflection Coefficient using Linear Sweep and Loop Phase Shift of 100 °

3.0 NONLINEAR SIMULATION

The application of the harmonic balance algorithm has some drawbacks in oscillator analysis. By nature the harmonic balance algorithm requires the presence of an input signal to the circuit. However in most oscillator designs, a fundamental input signal does not exist. In order to provide an input signal to the circuit the OSCTEST element is utilized. This element is an ideal mathematical entity and has no insertion phase affect on the circuit. The OSCTEST element passes all signals bilaterally in its through path while passing only the fundamental at its sample ports. The harmonic balance excitation frequency is applied to the input sample port. The oscillator loop reflected signal is observed at the output sample port. This allows the incident and reflected signals to be compared for the existence of the oscillation criteria. The first criteria for oscillation to exist in the loop is that the ratio of the magnitudes of the reflected to incident signal must be unity. In addition, the angle of this ratio must be equal to 0 degrees.

Because the frequency can change as oscillation develops in a large signal circuit the frequency of the excitation source must be very close to the final solution. If it is not the simulation will terminate or result in nonconvergence. Another problem of the harmonic balance simulation is that the power level of the excitation source must be very close to that which is required to find the final solution to obtain accurate results. To further complicate the tedious nature of the harmonic balance analysis is the application of this procedure to very high Q oscillator circuits such as the regenerative DRO. High Q circuits can result in numerical instabilities which also result in nonconvergence. Therefore a certain amount of manual control of the harmonic balance simulator is required to achieve accurate results. The remainder of this paper outlines the steps required to achieve quick and accurate results from the harmonic balance simulator for the regenerative DRO.

4.0 DETERMINE EXCITATION FREQUENCY

From the previous linear simulation it has been shown that the circuit has the potential to oscillate around 5.2 GHz. Therefore the same frequency sweep range will be used for the initial nonlinear sweep. A moderate number of harmonics should be chosen for the initial nonlinear sweep. In this case, set NH = 3 . The power level shall initially be set at -10 dBm.

The incident and reflected signals from the **OSCTEST** element shall be defined as output variables so that their ratio can be computed.

```
OUTVAR
      VIN = OSC VFC TP_T28 H1=1
      VFB=OSC VFC TP_T31 H1=1
OUTEQN
      RATIO = VFB/VIN
OUT
      OUTEQN MAG[RATIO] SCN
      OUTEQN ANG[RATIO] SCN
```

The circuit is then swept using the harmonic balance simulator. The output is viewed in tabular format. Table 1 shows the approximate oscillation frequency to be around 5.1935 GHz

Table 1. Magnitude and Angle of R at the OSCTEST.

(Input Power = - 10 dBm)

FREQ	RATIO	
RATIO		
GHz	MAG	ANG
5.19000	1.85755	44.9464
5.19100	1.59600	30.6008
5.19200	1.33566	16.2599
5.19300	1.07651	1.92353
5.19400	0.97982	-2.14097
5.19500	0.88238	-6.21514
5.19600	0.78422	-10.2957
5.19700	0.74690	-13.0293

5.0 DETERMINE EXCITATION POWER

The next step is to find the optimum excitation power level to perform full harmonic balance oscillator analysis. Set the frequency range for a fixed CW at the frequency found in the previous step (5.193 GHz). Then sweep the power over a range from -10 dBm to 0 dBm. Again use the harmonic balance sweep to obtain a tabular output of the **OSCTEST** element.

As Table 2 shows, the optimum excitation power lies between -5 and -6 dBm. For this circuit -5.5 dBm is chosen as the power level for the **OSCTEST** element. Now the circuit is ready for full harmonic balance analysis. The excitation frequency is set at 5.1935 GHz and the power level at -5.5 dBm.

Table 2. Magnitude and Angle of R at the OSCTEST
(Input Freq. = 5.19350 GHz)

PWR dBm	RATIO MAG	RATIO ANG
-10	1.0758	2.0523
-9	1.0760	1.7560
-8	1.0764	1.3226
-7	1.0770	0.6261
-6	1.0736	-.3110
-5	1.0145	3.9678
-4	0.9359	12.043
-3	0.8481	19.847
-2	.7758	28.212

6.0 DETERMINE INITIAL CONDITION

To further aid the convergence of the harmonic balance simulation an initial condition must be set at any node in the oscillator circuit. In this design the output node is chosen to set the initial voltage. To find the initial voltage at the output node, simply set the excitation source frequency and power as previously determined and perform a harmonic balance sweep while observing the output voltage waveform. As Figure 3 shows the initial output voltage is approximately 0.9 volts

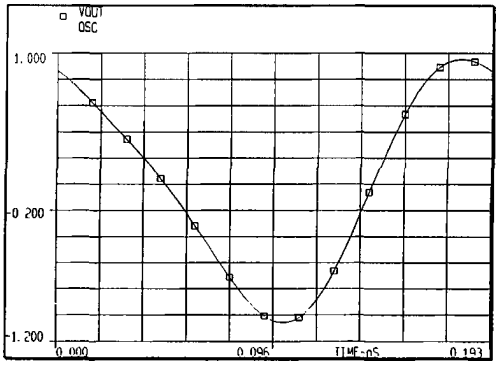


Figure 3. Oscillator Circuit Output Voltage Waveform for Initial Condition Definition.

7.0 FULL HARMONIC BALANCE SIMULATION

Finally the circuit is ready for full harmonic balance simulation. To initiate the oscillator analysis mode in Libra, the initial conditions must be entered in the harmonic balance control block.

```
HBCNTL
    OSCNODE=TP_T32
    OSCVOLT= 0.9
```

In a simple oscillator circuit such as a single transistor oscillator the harmonic balance control block listed may be all that is necessary to perform the nonlinear analysis. Due to the complex nature of the regenerative DRO namely the four stage amplifier and the high Q resonator, this simulation requires some manual control over the calculations. First the simulation speed versus accuracy should be modified to achieve solution convergence. This is best accomplished with the SAMPLE and RELTOL controls. The SAMPLE rate controls the multiple of sample points used by the simulator to convert node voltages from frequency domain to time domain. The RELTOL controls the accuracy of the harmonic balance error function. The default error function that is based on branch currents is 0.00001 . In a microwave oscillator such as this, power calculations to 0.1 dB and frequency calculations to the nearest 1 MHz is considered acceptable. Therefore the error function tolerance can be relaxed by a large amount. This will result in much faster simulation time and reasonably good accuracy.

For complex oscillator circuits the harmonic balance simulator will require further modification of the oscillator mode analysis to achieve solution convergence. The OSCDAMP and OSCNDAMP controls have direct effect on the numerical calculations The harmonic balance method uses Newton's method of finding the zeros of the current error function.[2] This technique finds the zeros by taking the first derivative of a function and extrapolating to the independent variable's axis. The process is repeated iteratively until the desired accuracy is achieved. The OSCDAMP control changes the algorithm to a modified Newton-Raphson technique. The OSCNDAMP specifies the number of modified steps to be taken before resuming the normal iterative solution.

The resulting control block is then given below.

HBCNTL
OSCNODE=TP_T32
OSCVOLT=0.9
SAMPLE= 2
RELTOL = .1
OSCDAMP = .1
OSCNDAMP = 20

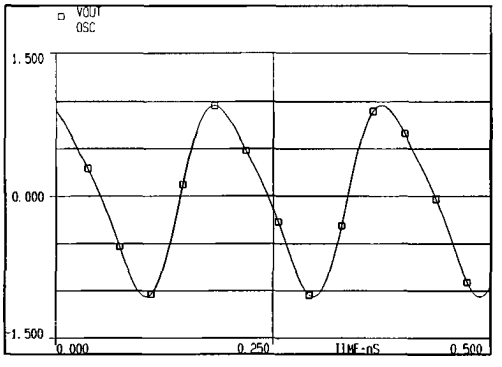


Figure 4. Output Voltage Waveform of Oscillator using full Harmonic Balance Simulation.

The oscillator circuit is then ready for full harmonic balance simulation. The output voltage of the oscillator is shown in Figure 4.

Figure 5 below shows the simulated output spectrum of the oscillator using the full harmonic balance simulation. The usable output power of the oscillator is +9.6 dBm from the directional coupler.

Therefore, taking the 10 dB loss of the directional coupler into account, the power out of the amplifier is +19.6 dBm. This is very close to the compressed output power of the amplifier as previously determined.

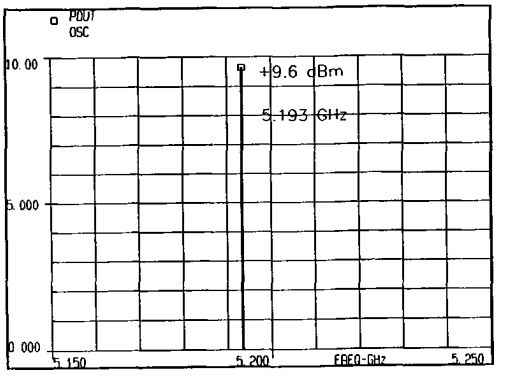


Figure 5. Oscillator Output Spectrum using full Harmonic Balance Simulation.

8.0 MEASURED RESULTS

Figure 6 shows a spectrum analyzer display of the oscillator's output spectrum. The measured output power was 8.7 dBm. This is 0.9 dB less than that predicted by the nonlinear simulation. Approximately 0.5 dB loss can be attributed to the SMA connectors, adapters, and semi-rigid cabling from the output of the amplifier to the output of the directional coupler. Therefore the final output power error is about 0.4 dB. This error may be further reduced by increasing the number of harmonics in the simulation but at the expense of extremely long simulation time. For the purposes of this initial nonlinear model the 0.4 dB error was considered acceptable.

The frequency shown in Figure 6 is approximately 5.196 GHz. This 3 MHz discrepancy is attributed to the repeatability of the puck height on its movable plunger. The movable plunger was capable of achieving about 300 MHz of mechanical tuning.

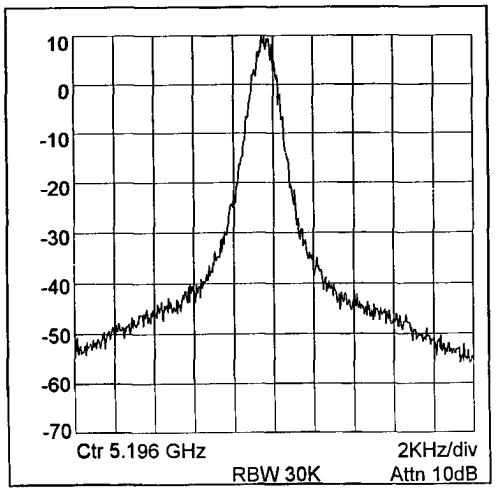


Figure. 6 Measured Output Spectrum of DRO

9.0 CONCLUSIONS

A basic nonlinear model for the regenerative dielectric resonator oscillator has been presented. Excellent correlation has been achieved between the simulated and measured output frequency and power. It has been demonstrated that the harmonic balance simulator can be efficiently applied to complex oscillator circuits. By combining good nonlinear device models along with measured S-parameters for the resonator, very good simulation accuracy can be obtained.

There are many advantages to having a good nonlinear CAD model for the regenerative oscillator. It is much more efficient to perform device load line analysis to determine the optimum load impedance for an oscillator. A graph of output power versus load impedance can be easily obtained. The designer could also analyze the effects of load impedance on oscillator harmonics. Temperature effects on loop gain margin can also be addressed.

Because the regenerative oscillator has demonstrated superior phase noise performance compared to other classes of oscillators, this CAD model can be applied to phase noise characterization. [3] Many of the newer generation of microwave simulation tools provide for the analysis of oscillator phase noise. This CAD model can be easily imported to those software platforms for phase noise simulation. Gaining insight to the oscillator phase noise at the simulation stage can save countless hours of trial and error laboratory work.

Acknowledgment

The authors would like to thank the Microwave Engineering Company of Harrisburg, PA, for the assembly and testing performed on the regenerative DRO.

References

[1] Robert G. Rogers, Low Phase Noise Microwave Oscillator Design: Boston, MA., Artech House, 1991

[2] Stephen A. Maas, Nonlinear Microwave Circuits: Boston, MA., Artech House, 1988

[3] M.Mizan, R.C. .McGowan, T.Lukaszek, A.Ballato, " Determination of the Limiting Factors in the Absolute Phase Noise of an L-Band Dielectric Resonator Oscillator", Proceedings of the 45th Annual IEEE Symposium on Frequency Control, pp 687-692

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

LATEST RESULTS OF THE U.W.A. CRYOGENIC SAPPHIRE OSCILLATOR

A.N. Luiten, A.G. Mann, N. J. McDonald and D.G. Blair,

Department of Physics, University of Western Australia, Nedlands, W. A. 6907, Australia.

Abstract

Two nominally identical separate 12 GHz oscillators based on 5cm diameter cryogenic sapphire resonators were compared using a double heterodyne method. Locking of the oscillator frequency to the resonator is provided by active Pound stabilization and a second servo which removes the deleterious effects of amplitude modulation generated by the phase modulator. The resonator temperature and incident microwave power are also kept under servo control. The diode detectors for the power and Pound frequency stabilization servos are adjacent to the resonator in the cryogenic environment, which eliminates their room temperature sensitivity. At short times, from 0.3 to 30 s, the Allan standard deviation is about $2.5 \times 10^{-15}\ \tau^{-1/2}$ and limited by the measurement system and servo system noise floors. At about 50 s the Allan deviation reaches a minimum of 8×10^{-16}. This represents frequency stabilization to better than 1 ppm of the resonator bandwidth. Up to 100 seconds the stability degrades as approximately 1 to $2 \times 10^{-16}\ \tau^{1/2}$, which is consistent with the earlier measurement of 4×10^{-15} for one oscillator against a H-maser. This medium term drift appears to be associated with a high oscillator acceleration sensitivity (~0.5 to $3 \times 10^{-9}\ g^{-1}$) and mechanical instability of the resonator.

Introduction

Over the last three years we have been constructing high stability microwave oscillators based around large (5cm diameter) cryogenic sapphire resonators [1, 2, 3]. The resonator is used as the feedback element in a loop oscillator at 12 GHz. Locking of the oscillator frequency to the center of resonance is provided by active Pound stabilization and a second servo which removes the deleterious effects of amplitude modulation generated by the Pound stabilization phase modulator. The resonator temperature and incident microwave power are both kept under servo control to minimize fluctuations in the resonator frequency. During the last year we have implemented this full set of four servos on both oscillators so that we are now approaching the intrinsic stability limit of the resonator rather than being limited, as in the past, by environmental influences acting on the oscillator components. We focus on the frequency stability results from a comparison of the two sapphire oscillators and the noise floors of their servo systems. We report an improvement to the stability which has been achieved by implementing an improved resonator temperature servo of one of the sapphire oscillators and discuss the residual resonator drift.

The Resonator

Two nominally identical resonators were cut from Crystal Systems "HEMEX" material and operated on an $H_{14\,1\,1}$ (quasi-$TE_{14\,1\,\delta}$) mode at 11.9 GHz, which displays a loaded Q of 10^9 at 6 K. Each sapphire dielectric resonator (SDR) is mounted inside a cylindrical niobium shield which is placed in a permanently evacuated vacuum pot. The pot is attached to a cryostat for cooling to liquid helium temperatures. Temperature control to a precision of about 10^{-5} K was provided by a carbon glass thermometer and a heater mounted in good thermal contact with the shield. Details of the two SDRs have been presented elsewhere [1, 2]. Resonator 1 was cut from a typical boule and has a frequency maximum at 6 K described by a curvature $1/f\ d^2f/dT^2 \sim 10^{-9}\ K^{-2}$. This frequency maximum considerably relaxes the temperature control requirement necessary for $\sim 10^{-15}$ fractional frequency stability. Resonator 2 came from atypical material, known to be slightly contaminated with titanium. This resonator has no extrema in our preferred temperature range. However, it has a minimum slope of $3 \times 10^{-10}\ K^{-1}$ at 6 K. The mode frequency temperature dependence of both resonators can be explained by a

0-7803-2500-1/95 $4.00 © 1995 IEEE

combination of the temperature dependence of the sapphire dielectric constant and the ac susceptibility of residual paramagnetic ions (Mo^{3+} and Ti^{3+}) [1, 2]. The high confinement of field energy to the sapphire in the $H_{14\,1\,1}$ mode results in the superconducting shield having a negligible influence on the mode frequency. Both resonators show an approximately linear dependence of mode frequency on the microwave power dissipated in the resonator, $\sim 10^{-10}$ per mW, consistent with the permittivity change due to electromagnetic radiation pressure [1].

Loop Oscillator

The oscillator circuit is essentially our original configuration [1, 2] with the addition of an amplitude modulation servo and is shown in Figure 1. In transmission the resonator is the primary frequency determining element of the free-running loop oscillator and in reflection it is the dispersive element in an active Pound-type frequency stabilisation scheme. A bandpass filter ($Q \sim 10^3$) is necessary to select the required SDR mode. The circuit losses are compensated by the gain of the GaAs FET amplifier (A) which runs in slight saturation.

The loop electrical length is servo controlled by the Pound frequency stabilization scheme to lock the oscillation frequency to the center of resonance. The carrier power incident on the SDR is phase modulated by a varactor diode phase shifter (PS1) driven by a crystal reference oscillator, at a rate of 80 kHz, which is much larger than the resonator bandwidth. PM-to-AM conversion upon reflection provides the discriminator signal which is recovered by synchronous detection with the reference oscillator. The SDR input coupling is set near unity for maximum carrier suppression which allows square law operation of the detector diode (D1). This maximizes the discriminator sensitivity and minimizes the sensitivity to incidental AM generated by the phase modulator PS1 (see below). After amplification and filtering the discriminator signal is fed back to a second varactor diode phase shifter (PS2) to control the frequency of the loop oscillator. The transfer function of the Pound frequency stabilization circuit is essentially that of a single-pole integrator and has a unity gain crossover point at about 1 kHz. Power from the oscillator is extracted after the VCA and before PS2.

All of the loop oscillator components are at room temperature except for the circulator (C) and isolators (I) (typically 3 per transmission line) that are placed near the resonator to minimize frequency pulling effects. The microwave power dissipated in the resonator is typically less than 0.5 mW and is regulated by the power servo comprising detector (D2) and voltage controlled attenuator (VCA). This servo fully compensates for amplifier power fluctuations, the variation of insertion loss of PS2 with the error voltage from the frequency stabilization servo and the long term variation of incident power on the resonator due to the variation of transmission line attenuation with liquid helium bath level.

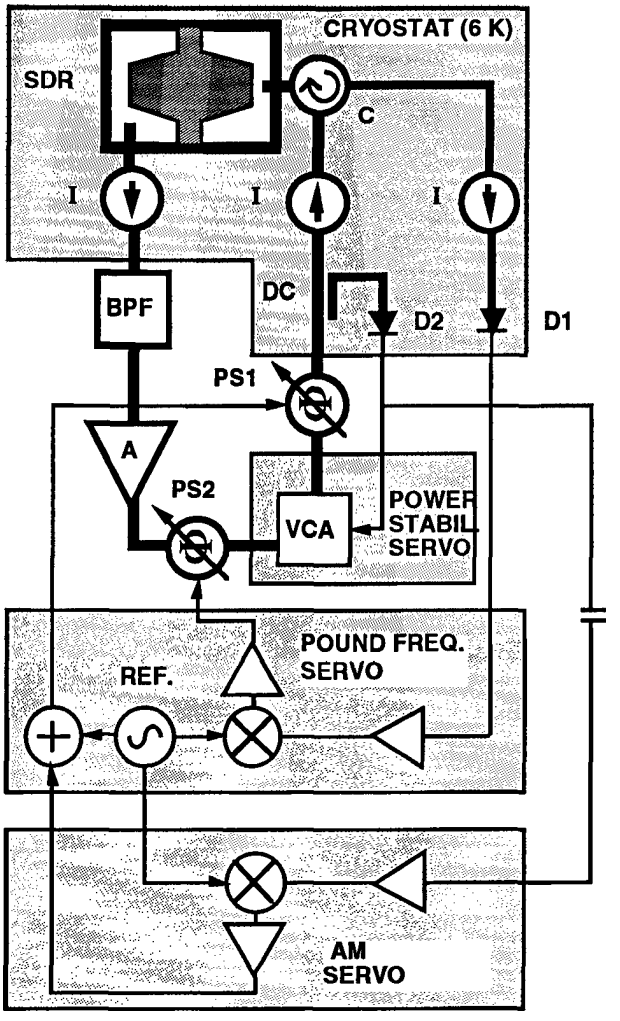


Figure 1 The loop oscillator circuit.

D1 and D2 are tunnel diode detectors placed near the resonator in the cryogenic environment. The advantage of this position is that it eliminates the room temperature dependence of detector sensitivity and improves the signal-to-noise ratio by eliminating transmission line loss (typically 5 dB) before the detectors. The power servo detector (D2) is operated at a power level corresponding to its highest sensitivity. By judicious choice of the coupling ratio of the directional coupler (DC) that provides power to D2 we can change the ratio of loop power to the fixed diode power. This allows us to trade off the noise floor of the power servo against the frequency stabilization servo noise floor in order to optimize the stability of the oscillator.

Both phase modulators have a variation of insertion loss (IL) with bias level that causes incidental amplitude modulation (AM) together with the desired phase modulation (PM). Incidental AM due to the error signal in PS2 is removed by the power servo. However the component of the AM generated by PS1 which is in phase with its PM will masquerade as a frequency offset error at the discriminator. Any variations in the level of this AM will induce frequency fluctuations. In some of our modulators we find a turning point in the variation of IL with bias. If we operate at this extremum we can eliminate the AM. This turning point has a temperature coefficient resulting in changes of the AM level (hence frequency offset) with room temperature. To reduce the variation in unwanted AM we have implemented an AM servo which keeps the phase modulator bias level at the turning point. A lock-in amplifier monitoring the power servo detector (D2) recovers the in-phase component of the AM and provides an error signal for the bias of the phase modulator (PS1). When this fourth (AM) servo system is running the temperature sensitivity of the oscillator drops to about 1×10^{-14} K^{-1} [1].

Oscillator Stability

Measurement System

To evaluate oscillator performance we have constructed two nominally identical systems, oscillator 1 and 2 which are mounted in separate cryostats, and have performed a standard two oscillator comparison using a double heterodyne configuration. The first difference frequency of 3.75 MHz (which is due to the slight dimensional mismatch between the resonators) is mixed with a similar frequency from a HP8662A synthesizer which has less absolute instability. For time domain measurements longer than 0.4 second the synthesizer is locked to a HP5065A rubidium frequency standard and is set to generate a second difference frequency near 220 Hz which is then counted in a 300 Hz bandwidth on a HP53131A frequency counter. The noise floor of this method is inadequate to characterize our oscillators at short times (see below). Previously we had used a "zero-beat" technique to circumvent the counter noise for measurement times under a few seconds [1, 4]. In the measurements reported here the complete hardware to implement this scheme was not available.

Time Domain Measurements

The measured oscillator stability, combining new (Feb 1995) frequency counter data (curves 1a and 1b) and our earlier (Aug 1993) "zero-beat" technique data (curve 3) [3] is shown in Figure 2. Also shown (point 2) is the upper bound on the stability of sapphire oscillator 1 at 1000 s obtained by beating it against a Shanghai Observatory H-maser, model H-M12A [1]. A factor of $\sqrt{2}$ has been removed from curves 1a, 1b and 3 to display the Allan deviation due to a single oscillator, assuming equal instability in each. The noise floors of these measurements and estimates of various oscillator instabilities are discussed in the next section and shown in Figure 3. At short times (<0.1 s) the Allan deviation appears to fall faster than τ^{-1} because of the bright line structure from vibrational modes in the resonator and power line frequency modulation. Amplifier flicker phase noise was present in the ealier measurements [3] as the oscillator output was taken from the loop immediately after the amplifier and before PS2. The measured stability below 0.1 s [3] was consistent with amplifier flicker phase noise of $S_\phi = 100/f$ dBc Hz^{-1} which translates to an Allan deviation of approximately $5 \times 10^{-16}/\tau$. From 0.3 to 30 s the Allan standard deviation is about $2.5 \times 10^{-15}\ \tau^{-1/2}$ and limited by the measurement system and servo system noise floors (see below). At about 50 s the Allan deviation reaches a minimum of 8×10^{-16}. Up to 100 seconds the stability degrades as approximately 1 to $2 \times 10^{-16}\ \tau^{1/2}$, which is consistent with the earlier measurement of 4×10^{-15} for oscillator 1 against a H-maser [1]. Curve 1b was obtained several days after the initial liquid helium transfer, at an time when the after-effects of the resulting mechanical disturbance had decayed away, whilst curve 1a was obtained one to two days after the transfer.(see below).

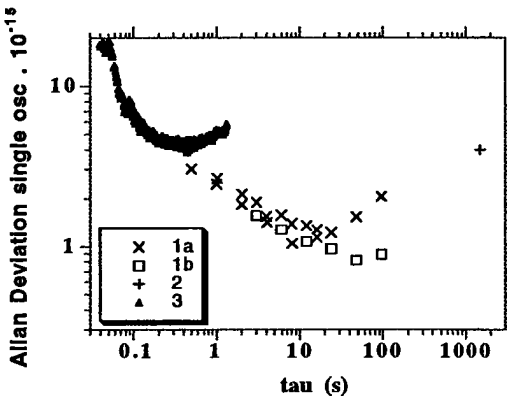


Figure 2 Frequency stability of sapphire oscillator:
1a, 1b frequency counter data (Feb 1995);
2 sapphire osc. 1 vs H-maser (May 94) [1];
3 "zero-beat" technique data (Aug 1993) [3].

System Noise Floors

Before installing cryogenic detectors in oscillator 2 we measured the noise floors of the power and frequency control servo systems (for oscillator 2) by placing an identical open loop servo system in parallel with the original, including a second detector which shares half of

the microwave power of the original detector [1]. The inferred Allan deviation due to the noise floor of the frequency control servo (F) is approximately $1 \times 10^{-15} \tau^{1/2}$, as is expected for the white frequency noise of the discriminator. This is in agreement with $1.4 \times 10^{15} \tau^{-1/2}$, calculated from the known detector sensitivity and noise floor and the -30 dB sideband to carrier power ratio of the 80 kHz modulation [5]. A similar calculation for the AM servo yields about $1 \times 10^{-15} \tau^{-1/2}$ [5]. The inferred Allan deviation due to the power servo (P) is flat, which is consistent with operation of the detector in its 1/f frequency regime. The box displayed on Figure 3 is an upper bound to the noise floor when operating with cryogenic detectors, especially in oscillator 1 which has a lower frequency vs power slope at the operating power [1].

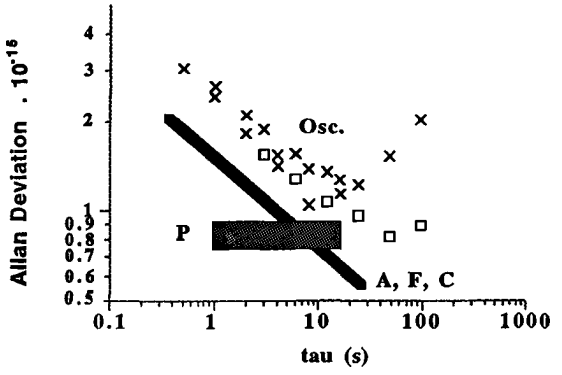


Figure 3 Frequency stability , noise floors of the measurement system (C) and oscillator servo systems (A, F, P).

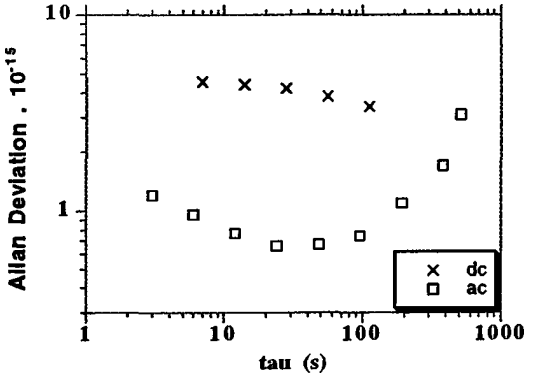


Figure 4 Inferred frequency stability of resonator 2 temperature servo systems, using a commercial dc and temperature bridge.

The requirement for temperature stability in resonator 2 is very strict as it has no turning point and only a minimum slope of $3\ \mathrm{Hz\,K^{-1}}$. We have implemented a 3 microKelvin resolution ac temperature bridge [6], initially using a room temperature reference resistor. Previously we had used a commercial dc bridge (Lake Shore DRC-91CA temperature controller) [1] which we have since found is affected by thermoelectric voltages generated in the room temperature connections to the thermometer leads. This results in a sensitivity to room temperature and medium term drift. The inferred Allan deviation resulting from the temperature fluctuations measured using a second (monitor) ac temperature bridge (thermometer near the resonator), when the temperature was servoed by either an ac temperature bridge or the commercial dc bridge is shown in Figure 4. It is clear that the high noise floor of the dc bridge was probably responsible for limiting the previously reported oscillator stability [1] to an apparent floor near 3 to 4×10^{-15}. The temperature instability due to the ac bridge is almost negligible. The drift beyond a few hundred seconds is due to the room temperature reference resistor (~5 ppm/K). We subsequently replaced it with a cryogenic one at the resonator and found no change in the oscillator drift.

Residual Drift

The frequency drift at times longer than 20 to 50 seconds is not correlated with any of our servos, room temperature or liquid helium bath pressure. It amounts to about 10^{-13} over one day. Both oscillators are somewhat microphonic, in proportion to their measured acceleration sensitivities: $\sim 3 \times 10^{-9}\ \mathrm{g^{-1}}$ for oscillator 1 and $5 \times 10^{-10}\ \mathrm{g^{-1}}$ for oscillator 2. When a severe mechanical disturbance is applied to the dewar top-plate, oscillator 1 shifts about 10^{-10} to 10^{-9} while oscillator 2 shifts about 10^{-13}. These frequency offsets decay away slowly, reaching $\sim 10^{-13}$ per day after a few days. They are not observed on any of the servos which implies that the offsets are changes in the resonator frequency. The amount of offset depends on the thermal history. In Dec 1994, for the first time in more than 2 years, we opened the vacuum can containing resonator 2 to change the input coupling (to unity) and retighten all the bolts, most of which were loose after dozens of thermal cycles from room temperature to liquid helium temperature and back. The first cooldown after retightening of the bolts showed shifts $\sim 10^{-13}$ due to mechanical disturbance, whilst on subsequent cooldowns its sensitivity increased to $\sim 10^{-10}$. This is consistent with the bolts becoming loose again. A second opening of both resonator cans in April 1995 (the first opening for resonator 1 in more than 2 years) to retighten probe and resonator mounting bolts did not greatly change the observed frequency shifts in response to large mechanical disturbances or acceleration sensitivity of resonator 1. The best data (curve 1b in Figure 2) was obtained five days after the initial liquid helium transfer

and the resulting mechanical after-effects had decayed away.

Frequency pulling from nearby cavity-like modes and/or the input coupling probe are the only viable mechanisms to account for the observed drift and acceleration sensitivity. The frequency spectrum around the $H_{14\,1\,1}$ operating mode is shown in Figure 5. In each resonator the nearest cavity-like modes are only 15 to 25 MHz away. These modes are grossly overcoupled, with loaded Qs of 3×10^4. We investigated the nearest one in resonator 1. One can estimate that the frequency pulling of the sapphire mode due to reactive coupling between sapphire and cavity-like modes will be attenuated by about 40 dB [7]. This is in rough agreement with the measured acceleration sensitivities: about 4×10^{-7} g^{-1} for the cavity-like mode compared to $\sim 3 \times 10^{-9}$ g^{-1} for the $H_{14\,1\,1}$ mode. Further work needs to be done to reduce the influence of nearby cavity-like modes and possibly the probes on the frequency of the operating mode.

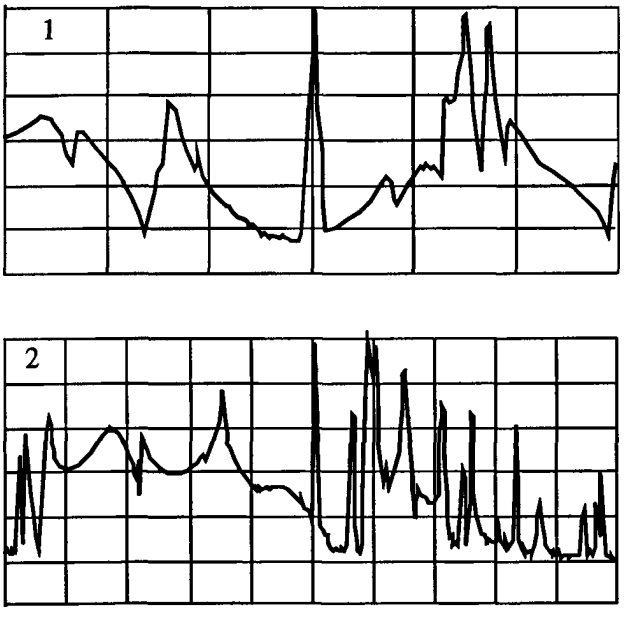


Figure 5 Frequency spectrum around $H_{14\,1\,1}$ (quasi-$TE_{14\,1\,\delta}$) mode at 11.9 GHz (center) for resonators 1 and 2. Vertical axis is 10 dB per division. Horizontal axis is 20 MHz per division. Spectrum analyzer bandwidth 200 kHz.

Conclusions

We have demonstrated that a power, phase, AM and temperature stabilized loop oscillator based on a cryogenic sapphire resonator can achieve an Allan standard deviation of about 2.5×10^{-15} $\tau^{-1/2}$ from 0.3 to 30 s, limited by the measurement system and servo system noise floors. At about 50 s the Allan deviation reaches a minimum of 8×10^{-16}. This represents frequency stabilization to better than 1 ppm of the resonator bandwidth. Up to 100 seconds the stability degrades as approximately 1 to 2×10^{-16} $\tau^{1/2}$, which is consistent with the earlier measurement of 4×10^{-15} for one oscillator against a H-maser. This medium term drift appears to be associated with a high oscillator acceleration sensitivity (~ 0.5 to 3×10^{-9} g^{-1}) and mechanical instability of the resonator.

Acknowledgments

This research was supported by the Australian Research Council. The authors would like to thank the Physics Dept. Mechanical Workshop, in particular Derek Newman, for the fabrication of sapphire clock hardware.

References

[1] A.N. Luiten, A.G. Mann, M.E. Costa and D.G. Blair, "Cryogenic Sapphire Resonator-Oscillator with Exceptional Stability: an Update", Proc. 48th Annual Frequency Control Symposium, Boston MA, 1994, pp. 441-446.

[2] A.N. Luiten, A.G. Mann, M.E. Costa and D.G. Blair, "Power Stabilized Exceptionally High Stability Cryogenic Sapphire Resonator Oscillator", IEEE Trans. Instr. & Meas., vol. IM 44, accepted for publication, April 1995.

[3] A.N. Luiten, A.G. Mann and D.G. Blair ,"Cryogenic Sapphire Microwave Resonator-Oscillator with Exceptional Stability", Electronics Letters, vol. 30, pp. 417-419, 1994.

[4] M.E. Costa, A.N. Luiten, M.E. Tobar and D.G. Blair "Oscillator Performance from the time evolution of relative phase", Electronics Letters, vol. 30, pp. 149-151, 1994.

[5] A.N. Luiten, Sapphire Oscillator Secondary Frequency Standard, Ph.D. Thesis, Physics Dept., University of Western Australia, 1995, unpublished.

[6] N.J. McDonald, F. van Kann and M.J. Buckingham, "MicroKelvin Thermometry Using Conventional Resistance Sensors", in preparation.

[7] M. E. Tobar, "Effects of spurious modes in cavity resonators", J. Phys. D: Appl. Phys., vol. 26, pp. 2022-2027, Nov. 1993.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

1 GHz STW BASED OSCILLATOR WITH CONTINUOUS TEMPERATURE COMPENSATION

Marian Angelov Taslakov,
Institute of Electronics, 72 Tzarigradsko Chaussee,
1784 Sofia, Bulgaria

Abstract

This paper presents results from a novel 1 GHz temperature compensated oscillator (TCO) featuring a continuous temperature frequency dependence. This work is a further development of the TCO using combined analogue and digital compensation which was studied in [1] and [2] in detail. The major advantage of the TCO, described here, is that there are no jumps in the temperature frequency dependence which results in a significant improvement of the spectral purity of the output signal. The overall temperature instability is ±1.25 ppm in the temperature range of (-30 ... +60) °C.

Introduction

Temperature compensation (TC) techniques using analogue and digital methods, have been widely used in crystal and surface acoustic wave oscillators to improve their temperature stability [4,5]. In 1991 we proposed a novel combined TC method that was successfully applied to a 1 GHz oscillator stabilised with a surface skimming bulk wave / surface transverse wave (SSBW/STW) combined mode resonator filter (CMRF) [1,6,7]. According to this method the temperature dependence is partly compensated for by an analogue circuit. Precision compensation is performed by a digital circuit. The overall instability achieved was ±1.4 ppm in the temperature range (-45...+75) °C. In 1993 we used a similar method for combined TC of a voltage controlled oscillator (VCO), stabilised with a CMRF. The VCO has a tuning range of a 1.2MHz which makes TC at 128 discrete frequencies with a step of 6.25kHz within this tuning range possible [2]. In this paper we describe a new greatly enhanced method for TC, combining the advantages of the digital and analogue methods and featuring a continuous temperature compensation curve.

Circuit description and compensation principle

The block diagram of the CMRF stabilised VCO is shown in Fig. 1. It consists of two amplifiers A1 and A2 with a phase shifter between them and a CMRF in the feedback loop. A3 is a buffer amplifier. The tuning range of this oscillator is about 1.2MHz at room temperature for a control

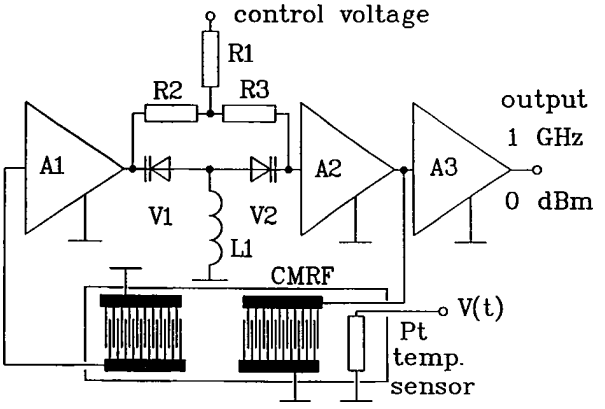


Fig. 1 Block diagram of the uncompensated CMRF based VCO with a tuning range of 1200ppm.

voltage ranging from 0.4 to 12.4 V. The CMRF is mounted on a precise Pt resistive temperature sensor which senses the crystal temperature and, by means of a resistance to voltage converter (RVC), converts the temperature into a temperature proportional voltage used by the compensation circuit. The dependence of the frequency versus temperature and control voltage is shown in Fig. 2. When cutting this surface with horizontal planes, each corresponding to a constant frequency, curves, giving the dependencies between the control voltage and the temperature, necessary for achieving a temperature independent frequency, are obtained. They show how the control voltage, as a function of the temperature, must change to keep the frequency constant. This dependencies are shown in Fig. 3 for 11 frequencies. For the sake of circuit simplicity we chose the curve corresponding to 997.3 MHz, because the control voltage, necessary to compensate the oscillator was less than 5 V, which allows the entire compensation circuit to be supplied from a 5 V d.c. source. The following figures illustrate the operation principle of the new temperature compensation method and compare it to the old method [1,2]. Fig. 4 shows part of parabola like curve (dotted line), and a two other curves, which approximate it: a stepped curve illustrates the old compensation principle and a curve consisting of 5 chords, illustrating the new method. Fig. 5 shows two curves, giving the frequency drift of the oscillator, compensated by both methods: the continuous curve corresponding to the new method and the dashed curve

0-7803-2500-1/95 $4.00 © 1995 IEEE

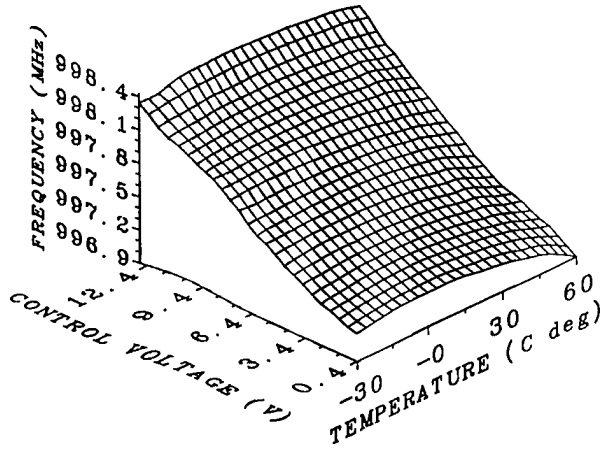


Fig. 2 Frequency as a function of two arguments - temperature and control voltage.

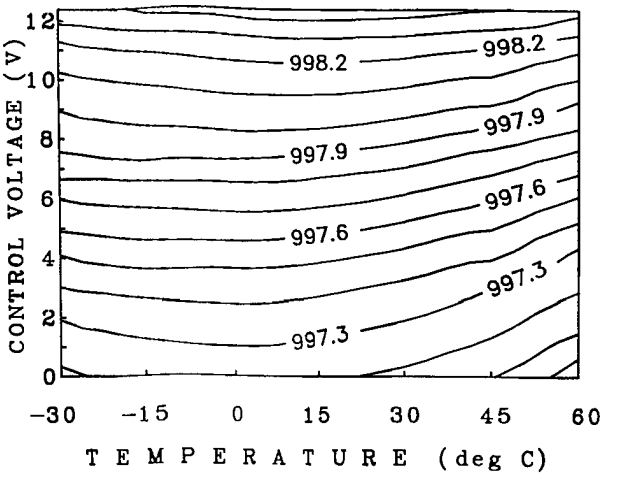


Fig. 3 These control voltage / temperature dependencies have to be synthesised by the TC circuit in order to obtain temperature independent discrete frequencies. A few of them are shown in the plot.

corresponding to the previous method. Obviously, there is a big difference between both curves. The major advantage of the new method is that the frequency temperature characteristic is continuous. However, there are also other advantages:

- the overall temperature induced frequency shift is more than 10 times less for the same number of points compared to the old method - this makes it possible to reduce the number of compensation points which facilitates the compensation procedure,
- the temperature induced drift is nearly independent of the temperature frequency characteristic slope of the uncompensated oscillator,
- the compensation principle is fully digital compared to the combined analogue and digital principle described in [1] and [2] and the analogue method proposed in [3]. It has

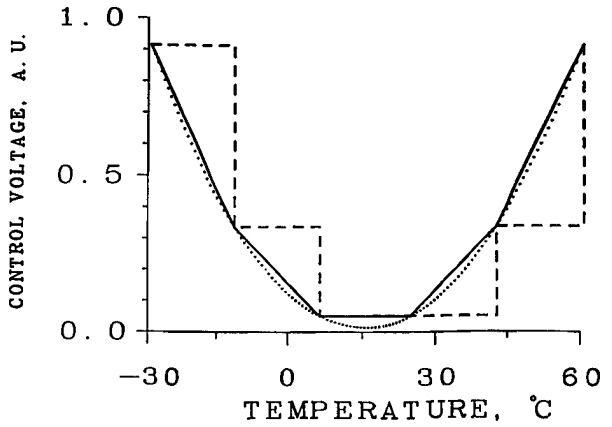


Fig. 4 Comparison of the new and the old compensation principle

(-------) old method
(_____) new method
(.....) compensation goal

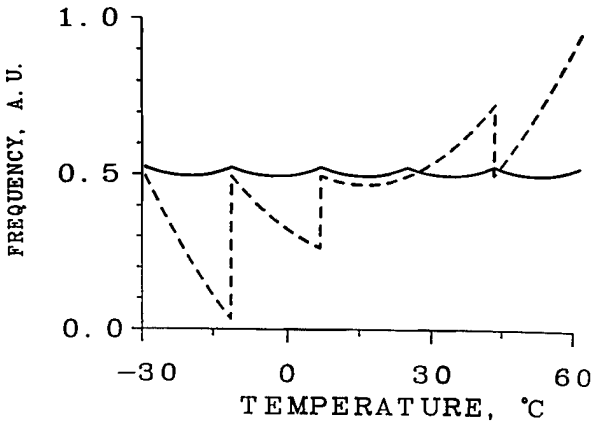


Fig. 5 An example for a compensation result using both methods

(-------) instability achieved by the old method
(_____) instability achieved by the new method

all advantages of the digital methods since it allows a fully computerised, very precise and greatly simplified compensation procedure. On the other hand it has the basic advantage of the analogue principle which is the continuous temperature frequency dependence.

The compensation principle is based on a new method for functional conversion, appropriate for slow processes as the time-temperature dependence of an oscillator is. Fig. 6 shows a block diagram of the temperature compensation circuit. The temperature proportional voltage (TPV), generated by a RVC, is summed by a summer (SUM) with a triangular signal, generated by triangular pulse generator (TG). The summed signal is applied to a 5 bit analogue to

digital converter (ADC). The ADC output generates a signal, which we will call multilevel pulse width modulated (ML PWM) signal. The temperature proportional voltage (TPV), generated by the RVC, is the modulating signal and a low frequency periodic pulse is the carrier. The pulse-to-width ratio is proportional to the momentum value of the TPV between two adjacent levels of the ADC, converting the TPV into a digital code. The ML PWM signal addresses an

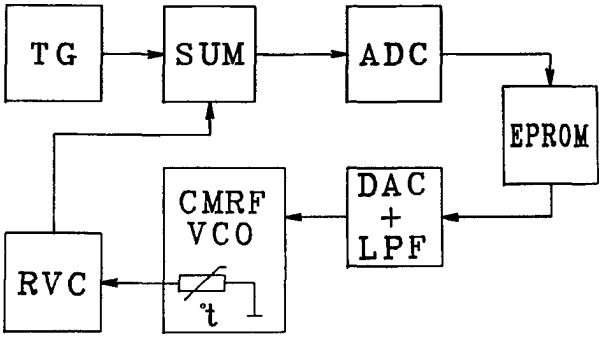


Fig. 6 Block diagram of the TC circuit.

EPROM in which the values of the compensation voltage are written for 32 equidistant temperature points. This signal is shown in Fig. 7. It is a function of temperature and time. It is evident that the ADC output signal is different for different temperatures. For comparison, with the previous method [1,2] the ADC output signal was constant between two adjacent temperature points. The EPROM output is converted to an analogue signal by a digital to analogue converter (DAC). Its output signal is also ML PWM modulated but the height of the steps depends on the local slope of the frequency - temperature characteristic. This waveform is illustrated in fig. 8. Furthermore it is integrated by a low pass filter (LPF) and applied to the control voltage port of the oscillator. The integration yields an average value of the compensation voltage at the DAC output so that the jumps between adjacent steps are completely eliminated. The operation principle is illustrated by the following equations:

$$U(T) = x.M + (1-x).(M+1), \qquad (1)$$

$$Uc = x.P + (1-x)Q, \qquad (2)$$

where M and M+1 are integers, corresponding to two adjacent levels of the conversion characteristic of the ADC, U(T) is a rational number, corresponding to the TPV. Its momentum value is between M and M+1. Uc is the oscillator control voltage and P and Q are the values of Uc stored under the addresses M and M+1 respectively. Finally, $0<x<1$ is a rational number, corresponding to the pulse-to-width ratio. By properly selecting the x value, U(T) and Uc can be determined with an arbitrarily high accuracy. The realisation of this method utilises an amplifier with a controlled gain and sign (ACGS) (inverting/noninverting). The gain corresponds to the

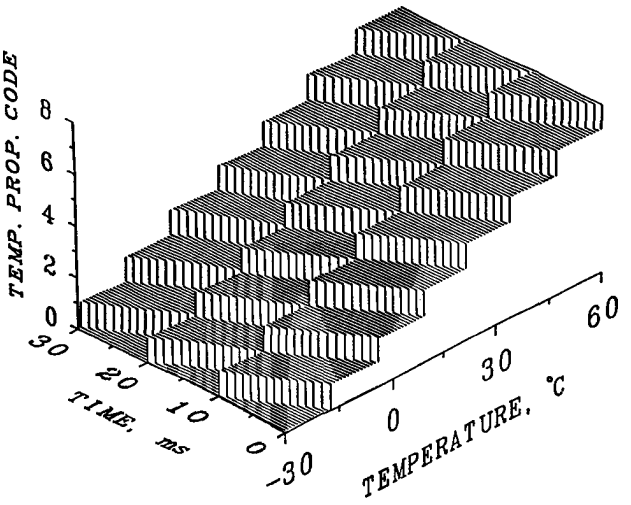


Fig. 7 Illustration of the temperature proportional code behaviour at the ADC output with time and temperature

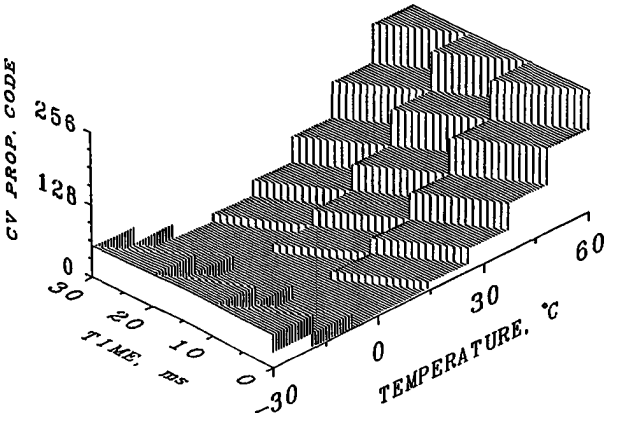


Fig. 8 Illustration of the control voltage (CV) proportional code behaviour at the DAC output with time and temperature.

slope of the function Uc = f(U(T)) that has to be synthesised, and the sign will change dependent on whether this function is increasing or decreasing. The ACGS is realised as a combination of the functional converter described in [1] and [2] and the modified pulse-width-modulation circuit the function of which was described above. For the triangular waveform we chose a frequency of about 100 Hz for the following reason. The voltage change rate of the triangular signal must be much higher than the rate of the temperature proportional sensor voltage. This condition makes the filtering of the 100 Hz signal very easy, preventing it from modulating the oscillator carrier frequency. For good approximation of the functional dependence Uc = f(U(T)), synthesised by the circuit, the amplitude of the triangular voltage is selected to be 1 to 2 times higher than the conversion step of the ADC. It is

important to note that the harmonics of the 100 Hz signal are present in the DAC output spectrum. These harmonics are filtered by a sixth order active low-pass filter with a cut-off frequency of 1 Hz. Since the amplitude of the 100 Hz is very low (tens of mV before filtering), these parasitic signals are well enough suppressed by the low pass filter and practically do not modulate the oscillator frequency.

Results

Fig. 9 is the experimentally obtained overall temperature instability at the goal frequency after the compensation. It does not exceed ±1.25 ppm. The highest instability is at negative temperatures. Most probably, the reason for that is ice covering the circuit. Another possible reason is that the used were low-cost types without specified temperature characteristics.

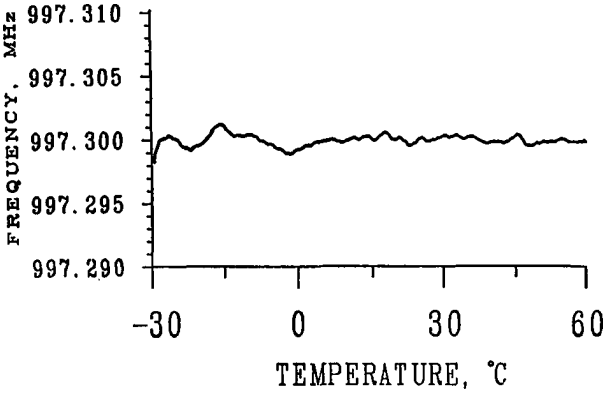


Fig. 9 Overall temperature instability after the compensation

Discussion

The maximum theoretical temperature instability, that can be achieved with the selected number of 32 chords, is ± 0.3 ppm. This value can readily be achieved by improving the oscillator packaging, using more stable analogue circuitry and evacuating the humidity from the temperature chamber.

Summary and conclusions

We have demonstrated a new principle of temperature compensation which can be applied to any kind of electronically tuneable stable oscillators. The method guarantees zero warm-up time. The calibration procedure is fully digital and can easily be computerised. Since the compensation circuitry can be realised entirely in micropower bipolar and CMOS technology, it is the author's belief that it can be successfully applied to portable communications.

Acknowledgements

The author wishes to thank Dr. Ivan D. Avramov from the Institute of Solid State Physics in Sofia for developing the SSBW/STW VCO used in this study and for the useful discussions.

References

[1] Taslakov, I.D. Avramov, "A Temperature Compensated 1GHz STW Oscillator", Proc. of the 1991 IEEE Ultrasonics Symposium, December 8-11, 1991, Lake Buena Vista, Florida, pp. 275-278.

[2] M.A. Taslakov, "A Temperature Compensated 1 GHz STW Based Multifrequency Oscillator", Proc. of 1993 IEEE Frequency Control Symposium, June 2-4, 1993, Salt Lake City, Utah, pp.275-278.

[3] M. A. Taslakov, " A Low Consumption Analogue Circuit for Temperature Compensation of SAW or BAW Devices", Proc. of 6th Conference "Acoustoelectronics'93", September 19-25, 1993, Varna, Bulgaria, pp.150-155.

[4] Kinsman, R.G., "Temperature compensation of crystals with parabolic temperature coefficients", Proc. 36-th Ann. Symp. on Frequency Control, 1982, pp.102-107.

[5] Cowan, V.D., Slobodnik, A. J., Roberts, J. A., and Silva, J. H.,"A 300 MHz digitally compensated SAW oscillator", IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control, Vol 35, No 3, May 1988, pp. 380-385.

[6] Avramov, I. D.,"1 GHz low loss coupled resonator filter using surface skimming bulk waves and Blewstein- Gulyaev waves", Electron. Lett., 1991, 5, pp.414-415.

[7] Avramov, I. D., and Kozlov, A. S., "Design of combined mode resonator filters based on surface transverse and surface skimming bulk waves", Proc. of the 10-th European Conference on Circuit Theory and Design ECCTD'91, Lyngby, Denmark, 1991, pp. 729-738.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Precise Coupling-of-Modes Simulation of Surface Transverse Wave Devices For Frequency Control Applications

Ken-ya Hashimoto[+] and Masatsune Yamaguchi[++]

Department of Electrical and Electronics Engineering
Faculty of Engineering, Chiba University
Inage-ku, Chiba-shi 263 Japan
[+]ken@sawlab.te.chiba-u.ac.jp, [++]masa@cute.te.chiba-u.ac.jp

Abstract

The paper proposes a precise simulation method for devices employing surface transverse waves (STWs) which propagate under grating structure and possess strong dispersive nature. The method is based on the coupling-of-modes (COM) theory, where four parameters required for the analysis, i.e., static capacitance, and self-coupling, mutual-coupling and transduction coefficients are determined as a function of frequency by the discrete Green function theory. The method was applied to the simulation of STW resonator filters on quartz employing metallic gratings. The result was in good agreement with experiments, and showed that the passband shape as well as the resonance frequency are most dependent on the thickness and width of grating electrodes. This indicates that STWs on quartz with weak piezoelectric coupling are substantially guided close to the surface by mass-loading effect of gratings.

1 Introduction

Because of such distinctive features as very large velocity, low propagation loss, superior temperature stability, surface transverse waves (STWs)[1] on quartz substrates have been expected to be promising for frequency control devices at GHz ranges[2].

Basic STW device structure is mostly the same with conventional surface acoustic wave (SAW) devices. However, the STW devices usually exhibit some very complicated frequency responses which are resulted from strong dispersion of the velocity and electromechanical coupling factor; this strong dispersion arises from the fact that the STW energy is trapped close to the propagation surface by mass-loading and/or topographic effects of grating structure[1]. Hence, the development of computer models and simulation methods, which can take account of the frequency dispersion, must play a key role in realising high performance STW devices.

The authors have recently proposed the discrete Green function theory[3] which rigorously relates charge and surface potential on periodic metallic gratings. The theory enables us to estimate various properties associated with piezoelectric acoustic waves including the mass-loading effect of a metallic grating[4, 5]. The paper proposes a simulation method for STW devices based on the coupling-of-modes (COM) theory[6], where four parameters required for the analysis, i.e., static capacitance, self-coupling, mutual-coupling and transduction coefficients are determined as a function of frequency by the discrete Green function theory.

The method was applied to the simulation of STW resonator filters on quartz employing metallic gratings[2, 7]. Without employing any filling and/or empirical factors, the result was in good agreement with the experiment. It was also shown that the passband shape as well as the resonance frequency are most dependent on the thickness and width of grating electrodes. This indicates that STWs on quartz with weak piezoelectric coupling are mainly guided close to the surface by mass-loading effect of gratings.

2 Theory

2.1 Original Bløtekjær's theory

Consider a metallic grating with the periodicity, width and height of p_g, w and h, respectively (See Figure 1).

When $h = 0$, the behaviour of the grating-modes having the wave number $\beta_0 = 2\pi s$ can be analysed by Bløtekjær's theory[8]. The calculation is performed according to the following procedure:

Let us represent the charge distribution $q(x_1)$ on

0-7803-2500-1/95 $4.00 © 1995 IEEE

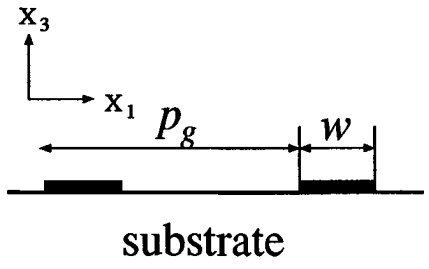


Figure 1: Coordinate System for the Analysis.

an electrode by,

$$q(x_1) = \sum_{m=M_1}^{M_2} \frac{A_m \exp(-j\beta_{m-1/2}x_1)}{\sqrt{\cos(2\pi x_1/p_g) - r}}, \qquad (1)$$

where $r = \cos(\pi w/p_g)$, $\beta_m = 2\pi m/p_g + 2\pi s$, and A_m is the unknown coefficient. Relative values of A_m are determined from the following simultaneous linear equations,

$$\sum_{m=M_1}^{M_2} A_m P_{n-m}(r) \left\{ S_n - S_{n-m} \frac{\epsilon(\beta_n/\omega)}{\epsilon(\infty)} \right\} = 0$$
$$\text{for } n = [M_1, M_2 - 1]. \qquad (2)$$

In the equation, $S_n = \mathrm{sgn}(\beta_n)$, $\epsilon(S)$ is the effective permittivity [9] of the substrate, $P_s(r)$ is the Legendre functions of the s-th order, and ω is the angular frequency.

Using the determined A_m, one can relate the potential $\Phi(s)$ and total charge $Q(s)$ on an electrode by the strip admittance $Y(s)$,

$$Q(s)/\Phi(s) = -2jY(s)/\omega = 2\sin(s\pi)\epsilon_g(s). \qquad (3)$$

Here $\epsilon_g(s)$ is the effective permittivity for the grating structure, and given by,

$$\epsilon_g(s) = \epsilon(\infty) \frac{\sum_{m=M_1}^{M_2} A_m P_{m+s-1}(r)}{\sum_{m=M_1}^{M_2} (-)^m A_m P_{m+s-1}(-r)}. \qquad (4)$$

The dispersion relation of the grating-modes under open-circuited (OC) and short-circuited (SC) gratings can be determined by $\epsilon_g(s) = 0$ and $\epsilon_g(s)^{-1} = 0$, respectively.

Distinctive feature of this method is exceptionally rapid convergence with respect to M_2 and M_1 with sufficiently small numerical errors; $M_2 = -M_1 = 1$, for example, gives the error of the order of 10^{-4} for SAW velocity calculation[3]. The error further decreases markedly with an increase in M_2 and $-M_1$.

2.2 Extended Blϕtekjær's theory

The authors proposed two methods of taking account of the effect of finite h in the above theory. One[4] employs the spectral domain analysis (SDA) for the derivation of acoustic wave fields in both the substrate and electrode regions. The other[5] employs the finite element method (FEM) in the electrode region, while the SDA is used only in the substrate region. Here the latter method is mainly described.

Let us designate the Fourier transforms of stress $T(x_1)$ normal to the surface, displacement $u(x_1)$, surface potential $\phi(x_1)$ and $q(x_1)$ at $x_3 = 0$ as $T(\beta)$, $u(\beta)$, $\phi(\beta)$ and $q(\beta)$, respectively.

In the lower half (substrate) region in Fig. 1, these field variables are related to each other by,

$$\begin{pmatrix} \dot{u}(\beta) \\ \phi(\beta) \end{pmatrix} = \frac{1}{|\beta|} \begin{pmatrix} R_{11}(\beta) & R_{12}(\beta) \\ R_{21}(\beta) & R_{22}(\beta) \end{pmatrix} \begin{pmatrix} T(\beta) \\ q(\beta) \end{pmatrix}, \qquad (5)$$

where $R_{ij}(\beta)$ is the effective acoustic admittance[10] which is determined by the material constants of the substrate. The "conventional" effective permittivity $\epsilon(S)$ in eq.(2) is given by,

$$\epsilon(\beta/\omega) = \left. \frac{q(\beta)}{|\beta|\phi(\beta)} \right|_{T(\beta)=0} = \frac{1}{R_{22}(\beta)}. \qquad (6)$$

Since the grating structure of current interest is fully periodic, the field variables should be composed of partial waves having discrete wavenumbers, i.e., β_m. Thus, $T(x_1)$ of a grating-mode with the wave number β can be expressed in the form of,

$$T(x_1) = \sum_{m=-\infty}^{+\infty} T(\beta_m) \exp(-j\beta_m x_1). \qquad (7)$$

Multiplying $\exp(+j\beta_m x_1)$ and integrating over one period length give,

$$T(\beta_n) = \frac{1}{p_g} \int_{-w/2}^{+w/2} T(x_1) \exp(j\beta_n x_1) dx_1, \qquad (8)$$

because of $T(x_1) = 0$ on the unelectroded region.

For the upper half (electroded) region, define the vectors $\hat{u}$ and $\hat{T}$ composed of $u(x_1)$ and $T(x_1)$ at the nodal points of $x_3 = 0^+$. If the driving frequency ω is specified, the application of FEM to the grating electrode region relates this two vectors in the form of,

$$\hat{T} = -[F]\,\hat{u}, \qquad (9)$$

where $[F]$ is the matrix derived from the FEM analysis.

From the Poynting theorem, acoustic complex power $P^{\pm}$ supplied from the boundary at $x_3 = 0^{\pm}$ is given by,

$$P^{\pm} = \mp j\omega \int_{-w/2}^{+w/2} u(x_1)T(x_1)^*|_{x_3=0} dx_1. \quad (10)$$

Application of the Parseval theorem gives,

$$P^{\pm} = \mp j\omega p_g \sum_{n=-\infty}^{+\infty} u(\beta_n)T(\beta_n)^*|_{x_3=0}. \quad (11)$$

The total power P supplied from the boundary is $P^+ + P^-$. Since $u(x_1)$ is continuous at the boundary, P should be zero if the solution is rigorous. Although P generally takes non-zero value because of numerical evaluation, it must satisfy the following condition,

$$\frac{\partial P}{\partial u(x_\ell)^*} = 0, \quad (12)$$

for each component $u(x_\ell)$ of $\hat{u}$.

Note P^- can be expressed in terms of $T(\beta_m)$ and $q(\beta_m)$ by substituting eq.(5) into eq.(10), while P^+ can be expressed in terms of $\hat{T}$ by substituting eq.(9) into eq.(11). Thus the condition represented by eq.(12) gives,

$$\hat{u} = \sum_{n=-\infty}^{+\infty} [L(\beta_n)]q(\beta_n), \quad (13)$$

where

$$[L(\beta_n)] = \frac{R_{12}(\beta_n)}{|\beta_n|}[A]^{-1}\,[G(\beta_n)]^{*t}.$$

Here $[G(\beta_n)]$ is a matrix giving the integration in eq.(8), and

$$[A] = \frac{1}{p_g}[I] + \sum_{n=-\infty}^{+\infty} \frac{R_{11}(\beta_n)}{|\beta_n|}[G(\beta_n)]^{*t}[G(\beta_n)][F].$$

Substituting this into eq.(5), one can finally obtain the relation between $\phi(\beta_m)$ and $q(\beta_m)$ in the form of,

$$\phi(\beta_m) = \frac{1}{|\beta_m|} \sum_{\ell=-\infty}^{+\infty} H_{m\ell} q(\beta_\ell), \quad (14)$$

where

$$H_{m\ell} = -R_{21}(\beta_m)[G(\beta_m)][F][L(\beta_\ell)] + R_{22}(\beta_m)\delta_{m\ell}. \quad (15)$$

By applying this expression to the original Blϕtekjær's theory, the simultaneous linear equations which govern A_m are obtained by,

$$\sum_{m=M_1}^{M_2} A_m \left\{ S_n \sum_{\ell=-\infty}^{+\infty} H_{n\ell} P_{\ell-m}(r) - S_{n-m} P_{n-m}(r) \right\} = 0 \quad \text{for } n = [M_1, M_2 - 1]. \quad (16)$$

Hence, substituting thus determined A_m into Eqs.(3) and (4), one can specify the relation between $\Phi(s)$ and $Q(s)$ on a grating structure of finite thickness. This result enables us to characterise the STW propagation taking account of the mass-loading effect.

Note that although SDA can relate $T(x_1)$ and $u(x_1)$ shown in eq.(9) in a closed form in this case[4], the closed form solution does not always result in reduced computational time. This is because the number of nodal points is not very large and the FEM computation is required only once at each frequency point.

Figure 2 shows the calculated STW velocity dispersion under an SC grating on 36°Y-cut 90° off X-

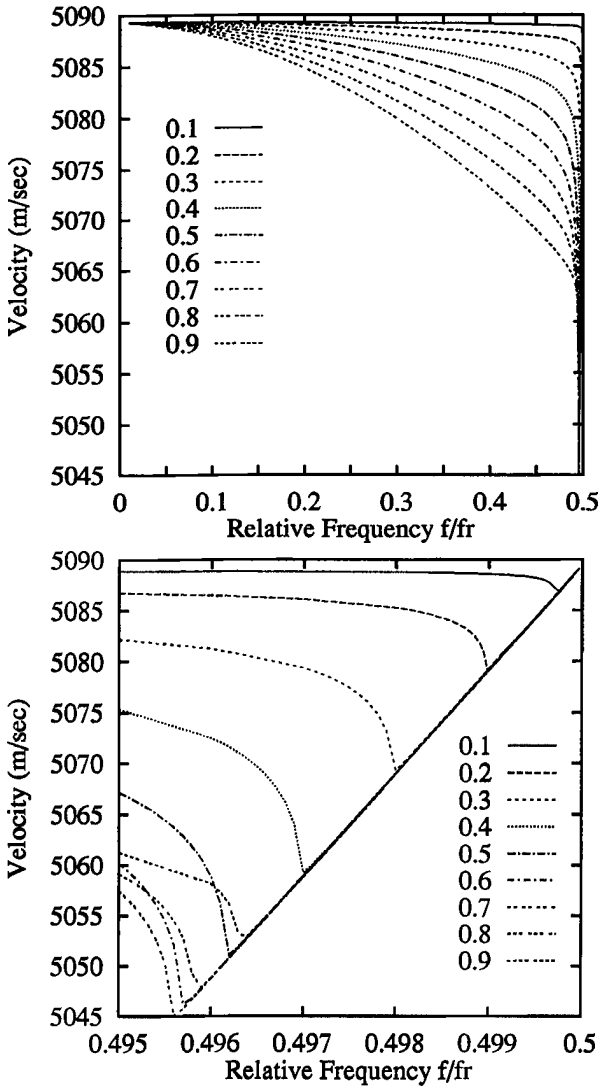


Figure 2: Change in the dispersion characteristics of STW velocity with w/p_g.

prop. quartz with w/p_g as a parameter. In the figure, $f_r = V_B/p_g$ and $V_B = 5,089$m/sec is the SSBW velocity[11]. Thickness h of grating electrodes (Aℓ) was assumed to be $0.032p_g$. For all w/p_g, the STW velocity decreases monotonically with f, and there appears a stopband at $f \cong f_r/2$, which is resulted from the coherent reflection by the grating. The velocity dispersion and stopband width become large with an increase in w/p_g. Within our calculation, no STW solution was found at $f > f_r/2$. The reason is that STWs become "leaky" and their electromechanical coupling becomes small because the wave-guide mechanism is weakened.

Figure 3 shows calculated STW attenuation due to the reflection with w/p_g as a parameter. It is clearly seen that not only the reflection amplitude but also the frequency giving maximum attenuation markedly changes with w/p_g.

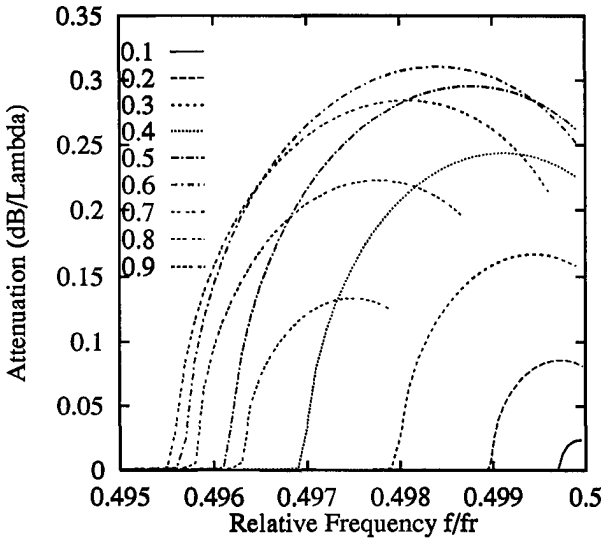


Figure 3: Change in STW attenuation with w/p_g.

Figure 4 shows the w/p_g dependence of the resonance frequency giving the maximum reflection. Except when $w/p_g \cong 0$, the frequency decreases almost linearly with an increase in w/p_g in the form of $0.50036 - 0.0032(w/p_g)$. It may be of current interest to note that the frequency should show the quadratic dependence if the topographic effect is dominant in the wave-guide mechanism. So the result in Fig. 4 suggests that STWs could substantially be guided close to the surface by the mass-loading effect.

Figure 4 also shows the maximum STW attenuation due to the grating reflection. The attenuation is almost proportional to $\sin(\pi w/p_g)$, which is similar to that for the conventional Rayleigh SAWs.

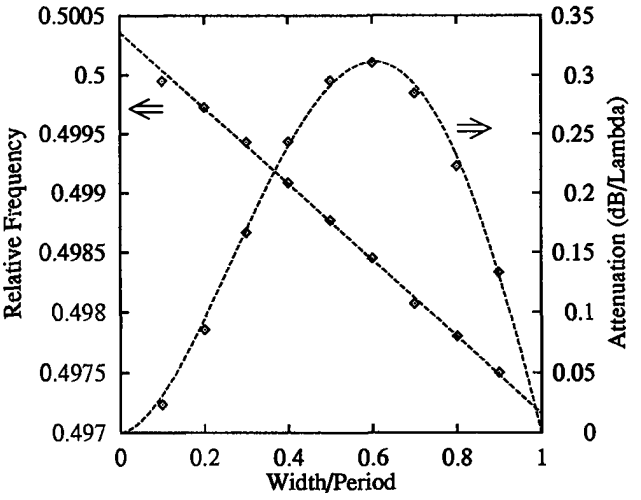


Figure 4: w/p_g dependence of STW resonance frequency and maximum attenuation. Dots:calculated, Dashed-line:fitted.

However, the w/p_g dependence of the attenuation is a little canted toward the larger values of w/p_g. So, the maximum reflection occurs at $w/p_g \cong 0.6$, whereas $w/p_g \cong 0.5$ gives the maximum reflection for Rayleigh SAWs. This difference may be attributed to the fact that although STW reflection at electrode edges phase-matches at $w/p_g \cong 0.5$, the STW penetration depth, with an increase in w/p_g, decreases and STWs become more sensitive to the surface perturbation. By numerical fitting, the w/p_g dependence of the attenuation is represented by $0.455x^{0.645}\sin(\pi x)$.

For the device design, w/p_g should be chosen to be a proper value giving the maximum reflection, so that the effects of the deviation of the electrode sizes occurring during the device fabrication is to be minimised. Note also, though, that the optimal value of w/p_g may be dependent on h/p_g.

2.3 Discrete Green function

Define the following variables $Q(s)$ and $\Phi(s)$ as,

$$Q(s) = \sum_{n=-\infty}^{+\infty} q_n \exp(+2\pi jns), \qquad (17)$$

and

$$\Phi(s) = \sum_{n=-\infty}^{+\infty} \phi_n \exp(+2\pi jns), \qquad (18)$$

where q_n and ϕ_n are the total charge and potential on the n-th electrode, respectively.

Since q_n and ϕ_n can be regarded as the Fourier expansion coefficients for $Q(s)$ and $\Phi(s)$, respectively, eqs.(17) and (18) can be rewritten in the form of,

$$q_n = \int_0^1 Q(s)\exp(-2\pi jns)ds, \qquad (19)$$

and

$$\phi_n = \int_0^1 \Phi(s)\exp(-2\pi jns)ds. \qquad (20)$$

Eqs.(19) and (20) suggest that q_n and ϕ_n could be expressed as a sum of contributions of various grating modes having wave numbers of $2\pi s/p_g$, where $0 \leq s \leq 1$. In 2.1 and 2.2, it was shown that $Q(s)$ and $\Phi(s)$ for each grating-mode are related to each other by $\epsilon_g(s)$. Thus, substitution of eq.(3) into eq.(19) gives,

$$q_n = 2\int_0^1 \sin(s\pi)\epsilon_g(s)\Phi(s)\exp(-2\pi jns)ds. \qquad (21)$$

Furthermore, substituting eq.(18) into eq.(20), one may obtain the following expression,

$$q_k = \sum_{\ell=-\infty}^{+\infty} \phi_\ell G_{k-\ell}, \qquad (22)$$

where G_k is the discrete Green function[3] given by,

$$G_k = 2\int_0^1 \sin(s\pi)\epsilon_g(s)\exp(-2\pi jks)ds. \qquad (23)$$

It is seen from eq.(22) that G_k gives the induced charge q_k when unit potential is applied on the 0-th electrode, while the potential on other electrodes is zero (see Fig. 5). Because of the reciprocity, $G_k = G_{-k}$ holds, which results in $\epsilon_g(1-s) = \epsilon_g(s)$.

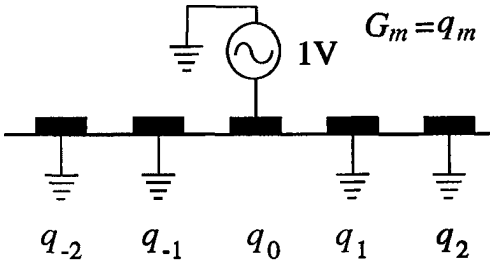


Figure 5: Discrete Green function.

Complex power P supplied through the electrodes can also be calculated by $\epsilon_g(s)$ as follows,

$$P = \frac{j\omega}{2}\sum_{k=-\infty}^{+\infty} \phi_k^* q_k = j\omega\int_0^1 |\Phi(s)|^2\epsilon_g(s)\sin(s\pi)ds. \qquad (24)$$

In Ref. [3], the authors applied the complex integral theory for evaluating G_n and P in eqs.(23) and (24). The result showed that the contribution G_{gn} of the grating-mode radiation is given by,

$$G_{gn} = -j\frac{4}{\pi}\epsilon(\infty)K_g^2\exp(-2\pi js_g|n|), \qquad (25)$$

and the power $P_g^{\pm}$ of the grating-mode radiated toward $\pm x$ is given by,

$$P_g^{\pm} = \frac{\omega K_g^2\epsilon(\infty)}{\pi}\left|\sum_{k=-\infty}^{+\infty}\phi_n\exp(\mp 2\pi jns)\right|^2, \qquad (26)$$

where K_g^2 is the effective electromechanical coupling factor of the grating-mode given by,

$$K_g^2 = \pi^2\sin(s_g\pi)\left\{\epsilon(\infty)\left.\frac{\partial\epsilon_g(s)^{-1}}{\partial s}\right|_{s=s_g}\right\}^{-1}. \qquad (27)$$

2.4 Determination of COM parameters

The COM equations considered are given by,

$$\frac{\partial A^+(x_1)}{\partial x_1} = -j\theta A^+(x_1) - j\kappa_{12}A^-(x_1) + j\zeta v, \qquad (28)$$

$$\frac{\partial A^-(x_1)}{\partial x_1} = +j\kappa_{12}A^+(x_1) + j\theta A^-(x_1) - j\zeta v, \qquad (29)$$

$$\frac{\partial I(x_1)}{\partial x_1} = -4j\zeta A^+(x_1) - 4j\zeta A^-(x_1) + j\omega Cv, \qquad (30)$$

where $\theta = \kappa_{11} + \omega/V_B - 2\pi/p_i$, $p_i = 2p_g$ is the IDT periodicity, v is the IDT voltage, C is the static capacitance per unit electrode-pair, κ_{11} and κ_{12} are the self- and mutual-coupling coefficients, and ζ is the transduction coefficient.

From the definition, C is determined by the effective permittivity $\epsilon(\infty)$ as[12],

$$Cp_i = \epsilon(\infty)\frac{P_{-1/2}\{-\cos(2\pi w/p_i)\}}{P_{-1/2}\{\cos(2\pi w/p_i)\}}, \qquad (31)$$

where P is the Legendre function.

As a solution to eqs.(28) and (29), one may obtain the wave number β_p of the "perturbed" mode under an SC grating by,

$$\beta_p = 2\pi/p_i \pm \sqrt{(\beta_u - 2\pi/p_i)^2 - \kappa_{12}^2}, \qquad (32)$$

where β_u is the wave number of an "unperturbed" mode given by,

$$\beta_u = \kappa_{11} + \omega/V_B. \qquad (33)$$

The grating-mode estimated in 2.2 includes all the effects caused by grating structure. This implies that the estimated mode corresponds to the "perturbed" one having β_p.

On the other hand, the "unperturbed" mode is an ideal one decoupled from the backscattered field components by the reflection. Since the field components are expressed as a sum of the Fourier components in 2.1 to 2.2, it is possible to remove the terms representing the backscattered field components, and therefore, the dispersion characteristics of the "unperturbed" grating-mode having β_u can be calculated. By substituting β_g and β_u into eqs.(32) and (33), κ_{11} and κ_{12} are determined as a function of ω.

Figure 6 compares the calculated dispersion characteristics of the "unperturbed" mode with the "perturbed" mode for $h/p_g = 0.032$ and $w/p_g = 0.5$. It is seen that the backscattering little affects the dispersion characteristics of the "unperturbed" mode when $f \neq f_r/2$, and that the stopband at $f \cong f_r/2$ disappears by removing the terms representing the backscattered field components.

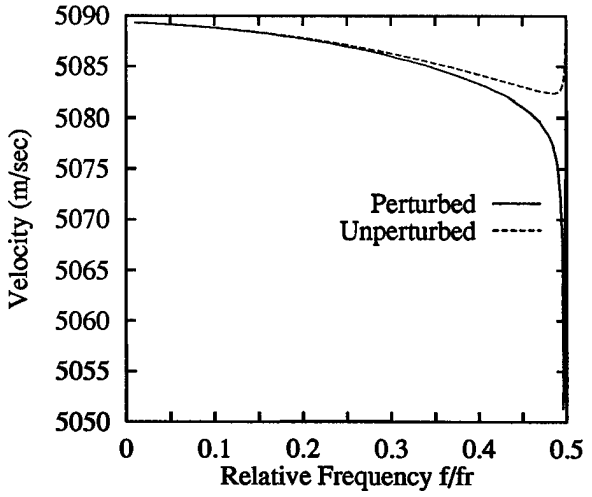


Figure 6: Velocity dispersion of STWs on 36°Y-cut 90° off X prop. quartz.

Next, in order to determine ζ, let us relate the mode amplitude $A^{\pm}(x_1)$ with the charge $q_{gk}^{\pm}$ carried by the grating-mode propagating toward $\pm x_1$. That is,

$$q_{gn}^{\pm} = \eta^{-1} A^{\pm}(np_i/2)\exp(\mp\pi jn). \qquad (34)$$

The factor $\eta = 4\sqrt{\omega\epsilon(\infty)K_g^2/\pi}$ is determined so that the power carried by the grating-mode having $|A^{\pm}(x_1)| = 1$ becomes unity[3]. By the definition, the current i_k flowing into the k-th hot electrode is given by,

$$i_k = j\omega(q_{gk}^{+} + q_{gk}^{-} + Cp_i v). \qquad (35)$$

Since i_k is equivalent to an increase in $I(x)$ within a unit IDT period, one may rewrite eq.(35) in the form of eq.(30) by the substitution of $p_i \partial I(x)/\partial x$ into i_k in eq.(35). This gives,

$$\zeta p_i = -4^{-1}\omega/\eta = \sqrt{\omega\epsilon(\infty)K_g^2/\pi}. \qquad (36)$$

Note that $\epsilon_g(s)$ for the "unperturbed" mode should be used for the estimation of K_g^2 by eq.(27). This is because eqs.(17) and (18) suggest that ζ could correspond to the excitation efficiency when the effect of the backscattering is ignored.

Figure 7 shows thus estimated COM parameters for STWs on 36°Y-cut 90° off X-prop. quartz, where $w/p_g = 0.5$ and $h = 0.016p_i = 0.032p_g$. First, κ_{11}, ζ, and κ_{12} increase monotonically with f. They take the maximum values at $f \cong 0.5f_r$, and then rapidly decrease. This is because the STW penetration depth increases due to the increased STW velocity within the stopband, and STWs become less sensitive to the surface perturbation.

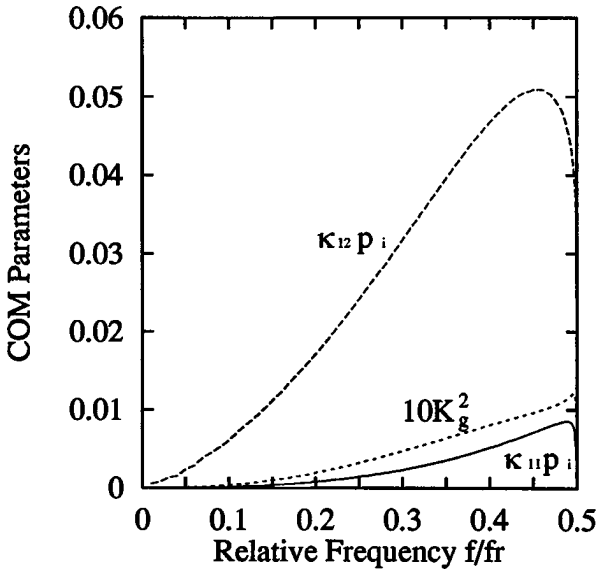


Figure 7: Estimated COM parameters for STWs under metallic-grating on quartz.

3 Device Simulation

By using the COM parameters shown in Fig. 7, the two-port STW resonator described in Ref. [2]

was analysed.

Figure 8 compares the calculated result with the experiments. Both calculated and experimental results are in very good agreement with each other. In particular, the cut-off nature at frequencies higher than the resonance is well simulated. In the experiment, however, there remains appreciable response above 780MHz (see Fig. 8). This is caused by SSBW, contribution of which was not included in the analysis.

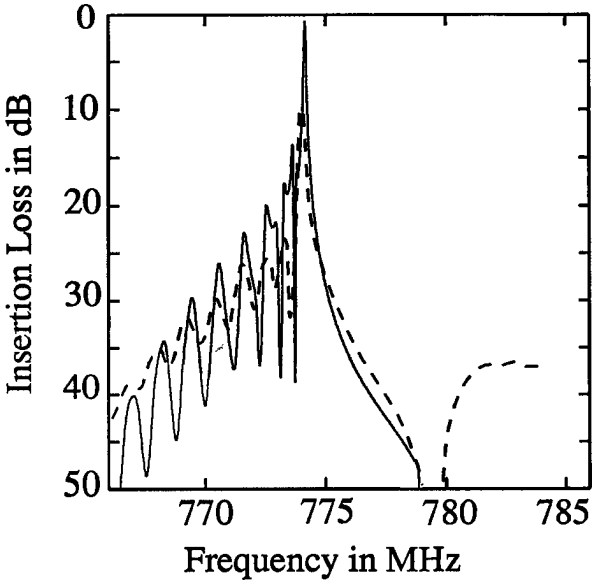


Figure 8: Frequency response of STW resonator. Solid-Line:Calculation, Broken-Line:Experiment[2]

By the numerical fitting, the propagation loss was estimated to be 0.0035dB/λ.

It should be pointed out that for comparison, the experimental result shown in the figure was translated by about 2.2MHz toward lower frequencies. This frequency shift may arise from a slight change in h/p_i occurring during the fabrication process.

Another two-port STW resonator described in Ref. [7] was then analysed. Figure 9(a) shows the calculated result; the traces from right to left represent the results for the devices with h/p_i =1.22%, 1.39%, 1.62%, 1.76% and 1.98%, respectively. Compared with the experimental results shown in Fig. 9(b), it is clear that the present theory well simulates the device characteristics, in particular, h dependence.

The simulation also predicts effects of deviation in w/p_i from 0.5. Note the simulation in Fig. 9(a) takes a slight deviation in w/p_i (less than 4%) reported by the authors[7] into consideration. If the effect is ignored, the shapes of the resonant peaks significant change as shown in Fig. 10.

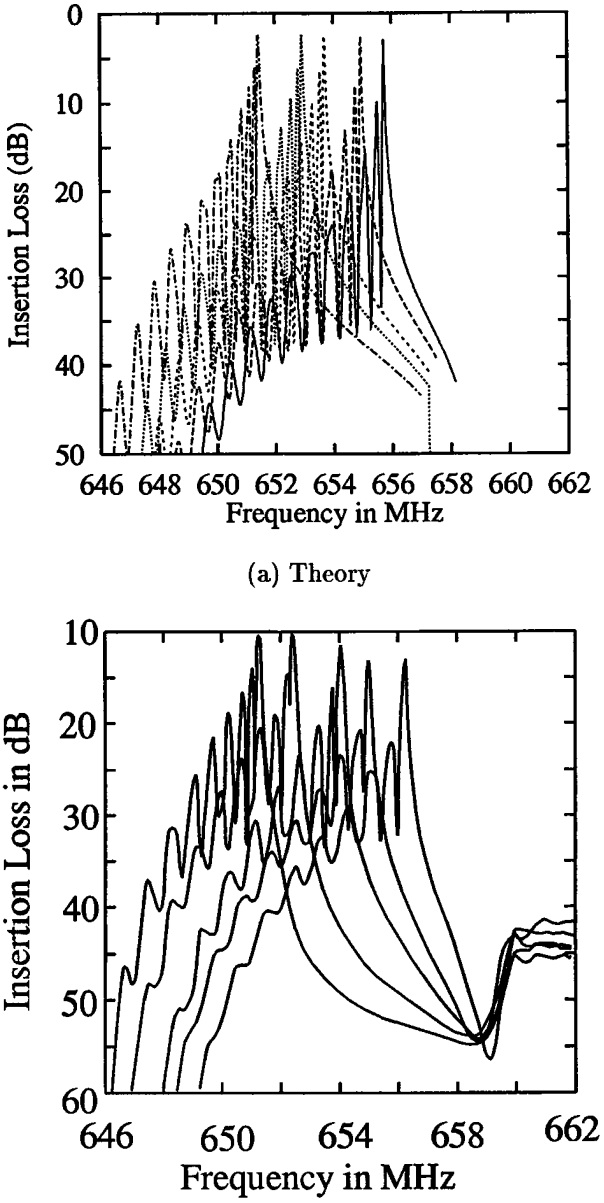


Figure 9: Frequency response of STW resonators. From right to left: h/p_i =1.22%, 1.39%, 1.62%, 1.76% and 1.98%.

4 Conclusion

The paper proposed a precise simulation method for STW devices using frequency dependent COM

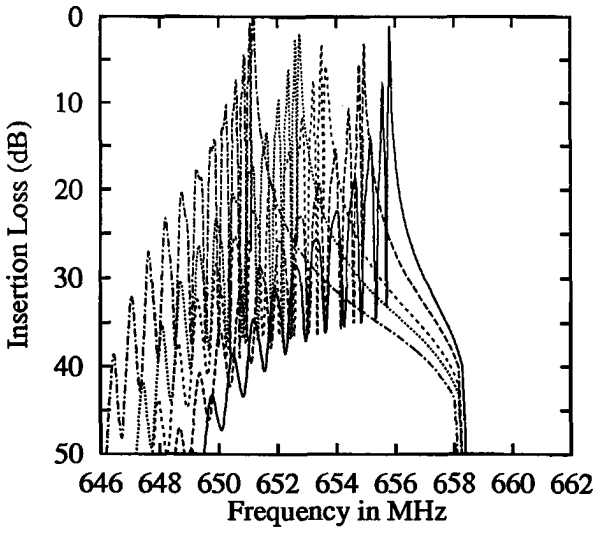


Figure 10: Theoretical frequency response of STW resonators when the deviation in w/p_i is ignored. From right to left: h/p_i =1.22%, 1.39%, 1.62%, 1.76% and 1.98%.

parameters. The transduction coefficient was determined by the discrete Green function theory, and coupling coefficients were determined by comparison of the dispersion characteristics of the "perturbed" and "unperturbed" grating-modes. This method enables us to take account of various electrical and mechanical effects in the COM simulation.

The method was applied to the simulation of two-port STW resonators on quartz, and the result was compared with experiments.

Note that since the present method is able to analyse the change in STW characteristics with film thickness and width as a function of frequency, it is also promising for the development of computer models for STW devices[13].

The method proposed here is also applicable to Rayleigh and leaky SAW devices, and the discrete Green function theory can be extended for metallic gratings with multiple electrode geometries[14]. This suggests that precise and general-purpose simulation tools could be developed by extending the present method.

Acknowledgements

It is a pleasure to thank Mr.C.S.Hartmann, Hartmann Research, Inc., and Dr.I.D.Avramov, the Institute of Solid State Science, Bulgaria, for their helpful discussion. Part of computer programming and calculation was carried out by Mr.G.Endoh of Graduate School, Chiba University.

References

[1] D.F.Thompson and B.A.Auld:'Surface Transverse Wave Propagation under Metal Strip Gratings', Proc. IEEE Ultrason. Symp. (1986) 261-266.

[2] I.D.Avramov:'Microwave Oscillators Stabilized with Surface Transverse Wave Resonant Devices', Proc. IEEE Frequency Control Symp. (1992) 391-408.

[3] K.Hashimoto and M.Yamaguchi:'Analysis of Excitation and Propagation of Acoustic Waves under Periodic Metallic-Grating Structure for SAW Device Modeling', Proc. IEEE Ultrason. Symp. (1993) 143-148.

[4] K.Hashimoto and M.Yamaguchi:'Bleustein-Gulyaev-Shimizu Wave under Metallic-Gratings of Finite Thickness', Jpn. J. Appl. Phys., **33**, 5B (1994) 3026-3030.

[5] G.Endoh, K.Hashimoto and M.Yamaguchi: 'SAW Propagation Characterisation by Finite Element Method and Spectral Domain Analysis', Jpn. J. Appl. Phys., **34**, 5B (1995) 2638-2641.

[6] D.P.Chen and H.A.Haus:'Analysis of Metal-Strip SAW Grating and Transducers', IEEE Trans. Sonics and Ultrason., **SU-32** (1985) 395-408

[7] I.D.Avramov and D.F.Thompson:'Influence of the Metallisation on the Electrical Performance of STW Two-Port Resonators', Proc. International Conf. on AcoustoElectronics (1993) 193-198.

[8] K.Blϕtekjær, K.A.Ingebrigtsen and H.Skeie:'A Method for analysing waves in structures consisting of metallic strips on dispersive media', IEEE Trans. Electron. Devices, **ED-20** (1973) 1133-1138,1139-1146.

[9] R.F.Milsom, N.H.C.Reilly and M.Redwood: 'Analysis of generation and detection of surface and bulk acoustic waves by interdigital transducers', IEEE Trans. Sonics and Ultrason., **SU-24** (1977) 147-166

[10] K.Hashimoto, Y.Watanabe, M.Akahane and M.Yamaguchi:'Analysis of Acoustic Properties

of Multi-Layered Structures by Means of Effective ACoustic Impedance Matrix', Proc. IEEE Ultrason. Symp. (1990) 937-942

[11] M.Lewis:'Surface skimming bulk waves, SSBW', Proc. IEEE Ultrasonics Symp. (1977) 744-752

[12] H.Engan:'Surface acoustic wave multielectrode transducers', IEEE Trans. Sonics & Ultrason., **SU-22** (1975) 395-401

[13] B.P.Abbott and K.Hashimoto:'A Phenomological Coupling-of-Modes Formalism for Surface Transverse Wave Devices', to be appeared in this Proceedings.

[14] K.Hashimoto and M.Yamaguchi:'Discrete Green Function Theory for Multi-Electrode Interdigital Transducers', Jpn. J. Appl. Phys., **34**, 5B (1995) 2632-2637.

Appendix

Success of the present study owes a lot to the discrete Green function theory, and the authors believe this theory can contribute to the further development of SAW device technologies. However, a large amount of experiences as well as efforts are needed to implant the sophisticated idea of the theory into practical computer softwares. The authors are much afraid that this might prevent many SAW researchers from widely applying the discrete Green function theory to device simulation and design.

From this point of view, the authors' research group has recently started the free distribution of the computer softwares "STW" and "FEMSDA" to researchers involved in SAW work.

The software "STW" is a complete set of tools used in this work. The other software "FEMSDA", which is the extended version of "STW", enables to derive the discrete Green function for any kinds of substrates[5].

The person(s) who are interested in them should sign on the "Public License and Distribution Policy" in the next page and send it to,

Address

K.Hashimoto
Dept. of Elec. & Electron. Engrg., Chiba University
1-33 Yayoi-cho, Inage-ku, Chiba-shi 263 Japan
tel: +81-43-290-3318
fax: +81-43-290-3039
E-mail:ken@sawlab.te.chiba-u.ac.jp

Public License and Distribution Policy

Copyright (C) 1994 Ken-ya Hashimoto and Masatsune Yamaguchi
SAW Research Laboratory, Department of Electrical and Electronics Engineering,
Faculty of Engineering, Chiba University
1-33 Yayoi-Cho, Inage-Ku, Chiba-shi 263 Japan
TEL:++81-43-290-3318 (ken), ++81-43-290-3317 (masa), FAX:++81-43-290-3039
e-mail: ken@sawlab.te.chiba-u.ac.jp, masa@cute.te.chiba-u.ac.jp

This computer software was developed by the SAW Research Laboratory in Chiba University (hereafter abbreviate SAW-LAB-CU) organized by Professors Yamaguchi and Hashimoto. For further development of SAW technologies, SAW-LAB-CU decided to distribute this computer software to persons who are active in the SAW technologies. Although SAW-LAB-CU does not request any charge for the use of the software, financial support for encouraging SAW-LAB-CU is very much appreciated. Especially, if the software contributes to the development of actual products, SAW-LAB-CU is very keen to receive it. Such contribution stimulates SAW-LAB-CU and enables to continue to develop new technologies, which will further benifit the contributors.

A software license with the source code will be distributed directly from SAW-LAB-CU via e-mail for persons who send completed this form. If the person moves to another organization after receiving the software, he must notify SAW-LAB-CU of his new address. If another person in the former organization wishes to continue to use the software, he must send a request for a software license to SAW-LAB-CU.

Each user can use, analyze and modify the software. Redistribution of this software must be done under the permission of SAW-LAB-CU.

As benefits of free distribution, SAW-LAB-CU requests that

1. The user must report to SAW-LAB-CU if he finds any bugs and/or trouble with the software. Note SAW-LAB-CU does not compensate any troubles and/or damages which may arise from the use of this software.
2. The user must note acknowledgment for the use of the software when the software contributes to his published work. In addition, the user must send a printed copy or photo copy of the publication to SAW-LAB-CU immediately.
3. In the event that the user makes changes to the original source code, the user must send the modified source code when SAW-LAB-CU requests it. The source code will be analyzed in SAW-LAB-CU for the further development.

I, (name)_____ of (Organization)_____, agree and guarantee to comply with the above mentioned policy.

I request that SAW-LAB-CU send (software name)_____ to me.

Signature_____ Date_____

Compete mailing address

telephone number:_____ fax number:_____

e-mail address:_____

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A PHENOMENOLOGICAL COUPLING-OF-MODES FORMALISM FOR SURFACE TRANSVERSE WAVE DEVICES

Benjamin P. Abbott[+] and Ken-ya Hashimoto[++]
[+]RF Monolithics, Inc.
4441 Sigma Road, Dallas, Texas 75240
[++]Department of Electrical and Electronics Engineering
Faculty of Engineering, Chiba University
Inage-ku, Chiba-shi 263 Japan

Abstract

Surface transverse wave (STW) devices exhibit two substantial advantages over conventional surface acoustic wave (SAW) devices. These are a substantially higher velocity and power handling capability. Each of these charactertics are often great assets in frequency control applications. However, STW devices have found little commercial application. This may be attributed to a lack of a practical and accurate analysis method.

By applying the recent advances in the analysis of STW devices by Plessky [1] and Hashimoto [2], a new COM analysis of STW devices has been developed. This new COM formalism is applied to the analysis of STW coupled resonators to demonstrate the accuracy of the method.

1 Introduction

Two recent advances have been made toward the analysis of STW devices [1, 2]. First, Plessky [1] introduced a two parameter analytic form for the STW dispersion relation. Second, Hashimoto and Yamaguchi [2] successfully applied the method of discrete Green's function to numerically determine the dispersion relation for STW's. Solutions for Plessky's COM parameter pair may be obtained by applying a non-linear least squares fitting procedure to the numerical results of Hashimoto's and Yamaguchi's discrete Green's function method.

Using these contributions as a foundation, this paper proposes a COM formalism for the analysis of STW devices. The approach assumes that the COM formalism for SAWs [3] can be adapted to satisfy the nature of STWs. Direct application of Plessky's STW dispersion is prohibitive. To alleviate this problem a new phenomenological dispersion relation is proposed.

This phenomenological dispersion relation requires five parameters rather than the two required by Plessky's STW dispersion relation. The problem of independently determining each of the five new parameters is avoided by determining an analytic equivalence between Plessky's dispersion relation and the new phenomenological dispersion relation.

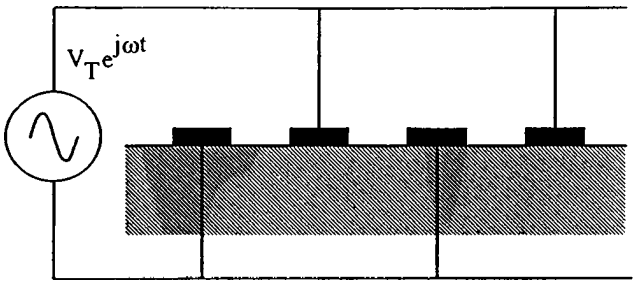


Figure 1: Interdigital transducer.

As a demonstration of the COM formalism, the method is applied to Y-rotated 36^0, X+90^0 propagating quartz. Several STW resonators have been designed. Comparison of the analysis results to experimental measurement demonstrates the accuracy of the model.

2 Dispersion Relation for STWs

Plessky successfully derived a simple analytic representation for the dispersion characteristics of STWs propagating in periodic gratings [1]. Plessky's analytic solution to the dispersion relation for such waves is

$$q(\Delta) = \sqrt{\Delta^2 - \frac{1}{4}\left(|\epsilon|^2 \pm \eta\sqrt{2|\epsilon|^2 - \eta^2 - 4\Delta}\right)^2} \quad . (1)$$

0-7803-2500-1/95 $4.00 © 1995 IEEE

where physically meaningful solutions normally correspond with the "+" sign under the radical. Δ is the wave number detuned with respect to the freely propagating SSBW, $\Delta = \omega/V_B - \pi/p$. V_B is the SSBW's velocity and p is the electrode period. The wave number of the STW as it resonates within the grating, i.e. the *perturbed wave number*, is expressed as $\beta_P = \pi/p \pm q(\Delta)$.

Plessky's analytic dispersion requires two fitting parameters η and ϵ. Each of these must be determined either empirically or by some appropriate theoretical means. The parameter η is responsible for describing the localization depth of the STW, propagating on a uniformly loaded surface. The parameter ϵ is responsible for describing the interaction between the counter propagating waves. This interaction encompasses both the STW↔STW interaction as well as the STW↔SSBW interaction.

From eqn (1), the cut off frequency for the STW↔SSBW interaction, Δ_B, as well as the upper and lower stop band edges, Δ_- and Δ_+, and the stop band center, Δ_C, can each be determined. These frequencies will each be required in the development of the COM model and therefore are given below in eqns (2) - (5). These are the same results presented by Plessky [1].

$$\Delta_B = -\frac{\eta^2 - 2|\epsilon|^2}{4} \tag{2}$$

$$\Delta_+ = \begin{cases} -\frac{(\eta - |\epsilon|)^2}{2}, & \text{if } \eta > 2|\epsilon|; \\ \Delta_B, & \text{otherwise} \end{cases} \tag{3}$$

$$\Delta_- = -\frac{(\eta + |\epsilon|)^2}{2} \tag{4}$$

$$\Delta_C = -\frac{\eta^2 + |\epsilon|^2}{2} \tag{5}$$

Since the COM formalism for conventional SAW devices is well established [3], there are great advantages if Plessky's dispersion relation could be incorporated directly into the conventional COM formalism. In the conventional COM formalism, the dispersion relation and the perturbed wave number are,

$$D(\Delta) = \sqrt{\left(\beta_U - \frac{\pi}{p}\right)^2 - \kappa_R \kappa_S} \tag{6}$$

$$\beta_P = \frac{\pi}{p} \pm D(\Delta). \tag{7}$$

The dispersion relation, $D(\Delta)$, is analogous to Plessky's $q(\Delta)$. In eqn (6), β_U is the *unperturbed wave number* . The unperturbed wave number is the that which would result if the forward and reverse surface modes were mathematically decoupled from each other. The coupling coefficients between the forward and reverse surface modes are κ_R and κ_S. In the absence of losses, these are complex conjugates.

Application of Plessky's dispersion relation into the conventional COM formalism would require that the perturbed wave number, β_P, the unperturbed wave number, β_U, and the reflection coefficients, κ_R and κ_S, be determined explicitly in terms of η, ϵ, and Δ. The functional disparity between Plessky's dispersion relation, eqn (1), and the conventional COM dispersion, eqn(6), makes the determination of the unperturbed wave number and reflection coefficients difficult. Therefore, the application of Plessky's dispersion relation within the conventional COM formalism is prohibitive.

To alleviate the difficulty in applying Plessky's STW dispersion relation to a COM analysis, a phenomenological dispersion relation is proposed. The phenomenological dispersion relation for STWs is derived by deducing appropriate forms for the unperturbed wave number and reflection coefficients. A great deal of thought and experimentation was required to determine suitable analytic forms for the unperturbed wave number and reflection coefficients. An overview of the process would not be productive. Therefore, the modified forms will be introduced directly.

$$\beta_U = \frac{\pi}{p} + (\Delta - \Delta_V) + |\kappa_B| \nu(\Delta) \tag{8}$$

$$\kappa_R = \kappa + \kappa_B \nu(\Delta) \tag{9}$$

$$\kappa_S = \kappa^* + \kappa_B^* \nu(\Delta) \tag{10}$$

The function $\nu(\Delta)$ encompasses the attenuation and strong frequency dispersion resulting from the back scattering of the surface modes (i.e. the STWs) into the counter propagating shear bulk modes (i.e. the SSBWs).

$$\nu(\Delta) = \frac{\eta_B}{\sqrt{\Delta_B - \Delta} + \eta_B} \tag{11}$$

The presence of $|\kappa_B|\nu(\Delta)$ and $\kappa_B \nu(\Delta)$ in the STW's unperturbed wave number and in the STW's effective reflectivity, respectively, is required to satisfy energy conservation. This result can be explicitly derived by considering a system of four modes. The four modes consisting of a pair of counter propagating STWs and a pair of counter propagating SSBWs. By assuming there are no SSBW energy sources and that no energy is dissipated into the bulk (this is only valid for frequencies below the SSBW cut off frequency, $\Delta < \Delta_B$) it is possible to conclude that the STW's unperturbed wave number and the STW's reflection coefficients must satisfy the general form of eqns (8) - (9). By extending the work to determine the dispersion relation for the complete system of four modes, it is found that the resulting dispersion relation has the same general form as eqn (6).

The phenomenological dispersion relation for STWs requires five parameters. They are Δ_V, Δ_B, κ, κ_B, and η_B. Each of these parameters represent physically meaningful quantities and are defined in the following paragraph.

Δ_V represents the vanishing point of the stop band. This is the point at which the stop band would vanish if the reflective coupling between the counter-propagating waves were removed. This includes the STW↔STW as well as the STW↔SSBW interaction. κ is the coupling coefficient describing the interaction between the counter propagating STWs. $\kappa_B \nu(\Delta)$ is the coupling coefficient describing the influence the SSBW presence has on the STW's unperturbed wave number and reflectivity. Δ_B found in eqn (11) is identical to the cut off frequency for scattering of the STW into the SSBW defined earlier in eqn (2).

The characterization of these five parameters represents a substantial problem for the practical application of the phenomenological dispersion relation. However, the forms of eqns (8) and (11) were not chosen arbitrarily. Using this phenomenological dispersion relation it is possible to determine a direct equivalence with Plessky's two parameter COM model. There are, in fact, a variety of equivalences. However, only two of these equivalences have been found to be physically meaningful. One of the five required COM parameters have already been presented in terms of Plessky's COM parameter pair. Δ_B is given by eqn (2). The vanishing point, Δ_V, is obtained directly from Plessky's dispersion relation. As the interaction between the counter-propagating waves is diminished toward zero, $\epsilon \rightarrow 0$, the stop band center, Δ_C, in eqn (5) converges on the vanishing point. Therefore, Δ_V can be obtained directly from eqn(5).

$$\Delta_V = -\frac{\eta^2}{2} \tag{12}$$

The remaining three phenomenological parameters are given below in eqns (13) through (15).

$$\kappa = \pm|\epsilon|\frac{2\eta + |\epsilon|}{2}\exp(j\theta_\kappa) \tag{13}$$

$$\kappa_B = \mp|\epsilon|^2\frac{\eta}{\eta + 2|\epsilon|}\exp(j\theta_\kappa) \tag{14}$$

$$\eta_B = \frac{\eta + 2|\epsilon|}{2} \tag{15}$$

These solutions were arrived at by requiring that the phenomenological dispersion relation, eqn (6), be equal to Plessky's dispersion relation, eqn (1), at three frequencies. The three frequencies are the stop band edges, Δ_- and Δ_+, as well as the cutoff frequency for back scattering of the STW into the counter propagating SSBW, Δ_B.

Mathematically, there exist two possibilities in solving for the magnitude of κ. The two possibilities represent the magnitude of κ as the difference between the vanishing point and either the lower or the upper stop band edge. That is, either $|\kappa| = \Delta_V - \Delta_-$ or $|\kappa| = \Delta_+ - \Delta_V$, respectively. However, examination of eqns (5) and (12) indicates that the vanishing point, Δ_V, must lie at (when $\epsilon = 0$) or above (when $|\epsilon| > 0$) the stop band center, Δ_C. This results in the stop band width above the vanishing point being smaller than the width below the vanishing point. Due to the proximity of the SSBW and the upper stop band edge, it can be argued that the STW's interaction with the neighboring SSBW causes a premature collapse of the upper stop band edge. From which it can be concluded, that the difference between the upper stop band edge and the vanishing point must be less than the magnitude of κ. Accepting this argument as true, the only acceptable definition of $|\kappa|$ is as the difference between the vanishing point and the lower stop band edge. Which directly leads to the form of eqn (13).

There still remains an ambiguity in the sign, or phase, of κ which is indicated in eqn (13) by the "$\pm$" and by θ_κ, respectively. If κ is real valued, the correct choice for the sign of κ may be resolved by experimental measurement. The conductance of an IDT, whose κ is negative, will be skewed toward the lower stop band edge. While the conductance of an IDT, whose κ is positive, will be skewed toward the upper stop band edge. If κ is complex, a more elaborate method must be employed to determine the exact phase of κ.

In eqn (13), if κ is positive, then κ_B in eqn (14) must be chosen as negative. Likewise, if κ is negative then κ_B must be positive. The phase, θ_κ, allows for the possibility of complex valued reflectivity. Complex values of reflectivity may occur in any anisotropic piezoelectric material. Substrates having complex reflectivity will exhibit longitudinal asymmetry between their mechanical and electrical components. Although non-symmetric substrates are usually avoided, the complex valued reflectivity has been included in the COM formalism for completeness.

The remaining two parameters, κ_B and η_B, are determined by requiring $D(\Delta_B) = q(\Delta_B)$ and $D(\Delta_+) = q(\Delta_+)$, respectively.

3 COM Equations

The COM differential equations, appropriate for the application of the phenomenological dispersion relation, are derived from Wright's work [3]. The equations presented here do not include the effects of finger resistance.

The forward propagating and counter propagating STW modes are represented as $R(x)$ and $S(x)$, respectively. The COM equations are,

$$\frac{\partial R(x)}{\partial x} = -j\beta_U R(x) + j\kappa_R S(x)e^{-j2\pi x/p} + j\alpha_R V_T e^{-j\pi x/p} \quad (16)$$

$$\frac{\partial S(x)}{\partial x} = +j\beta_U S(x) - j\kappa_S R(x)e^{+j2\pi x/p} - j\alpha_S V_T e^{+j\pi x/p} \quad (17)$$

$$\frac{\partial I(x)}{\partial x} = +2j\alpha_S R(x)e^{+j\pi x/p} + 2j\alpha_R S(x)e^{-j\pi x/p} - j\omega C_T V_T \quad (18)$$

where

$$\alpha_R = \alpha \quad (19)$$

$$\alpha_S = \alpha^* \quad (20)$$

Eqns (16) through (20) include the IDT's transduction coefficient, α, and the IDT's capacitance per unit transducer length, C_T.

In the absence of losses, κ_R and κ_S are complex conjugates. However, at frequencies where the STW scatters into the counter propagating SSBW, $\Delta > \Delta_B$, κ_R and κ_S are no longer complex conjugates. This is consistent with the effect that finite finger resistance has on these same COM parameters [3] and indicates that energy is no longer being conserved by the system. In the case of the STW, energy is lost as a result of the STW scattering into the counter-propagating SSBW.

4 STWs on quartz

To apply these COM equations to a device analysis, the four parameters, α, C_T, η, and ϵ, must be determined. The transduction coefficient, α, can be assumed to be independent of frequency, i.e. a constant, without introducing significant errors. The value of α is most reliably determined from experimental measurement. The distributed capacitance, C_T, may be determined either by experimental measurement or by electrostatic analysis methods. Plessky's COM parameter pair, η and ϵ, can be obtained by either an elaborate fit to experimental measurements or through the use of rigorous theory. Using the work of Hashimoto and Yamaguchi [2] it is possible to numerically determine the dispersion diagram for various metallization ratios and thicknesses. As an example, the response for an Aluminum grating of length 100p with a period, p, of 5 μm and thickness 0.04p deposited on Y-rotated 36^0, X+90^0 propagating quartz is shown in Figure 2. Above the SSBW cutoff frequency, about 509 MHz, a substantial

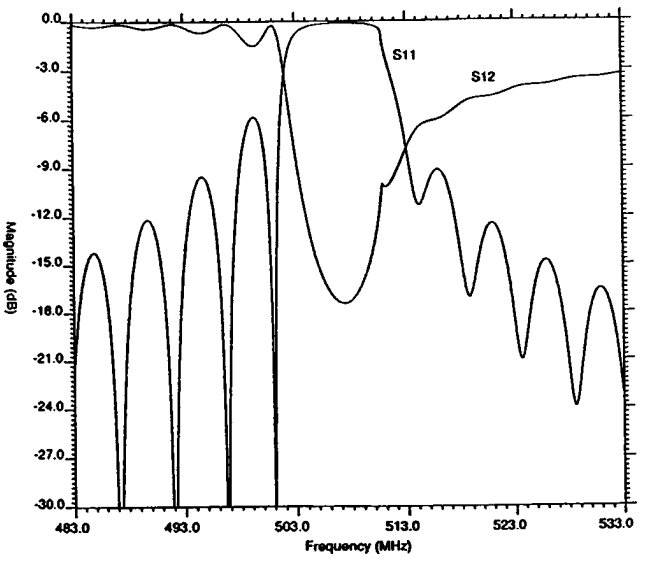


Figure 2: S11 and S12 for a grating.

amount of scattering loss is evident in Figure 2. Figure 3 illustrates the dispersion diagram for the same grating. In Figure 3, the numerical data calculated using the method of discrete Green's function is represented by the square symbols. The solid lines represent either Plessky's dispersion relation or the new phenomenological dispersion relation, which are indistinguishable from each other. The method of discrete Green's function is unable to locate solutions if the electromechanical coupling coefficient is too small. Since the electromechanical coupling coefficient diminishes substantially as frequency approaches the top of the stop band, the method of discrete Green's function was unable to converge on solutions for frequencies above about 509 MHz.

The numerical results were found to be well behaved and were easily fit to simple analytic formulae. For 36^0 Y-rotated, X+90^0 propagating quartz the analytic fits to η and ϵ are

$$\eta \approx 0.35 \tanh\left(20\frac{h}{\lambda}\right)\frac{a}{p}\sqrt{\pi/p} \quad (21)$$

$$|\epsilon| \approx 4.4 \sin\left(\pi\frac{a}{p}\right)\frac{h}{\lambda}\sqrt{\pi/p}. \quad (22)$$

From eqns (21) and (22), both κ and the velocity shift from the SSBW velocity, $\Delta v/v$ can be determined by substitution into eqns (13) and (12), respectively. Their approximate values for a 50/50 line to space are,

$$\kappa\lambda \approx -50\pi\left(\frac{h}{\lambda}\right)^2 \quad (23)$$

$$\frac{\Delta v}{v} = \frac{V_{STW} - V_{SBBW}}{V_{SSBW}}$$

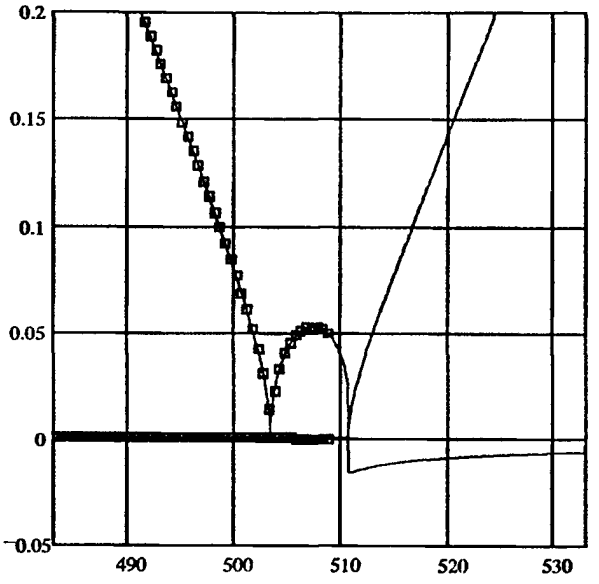


Figure 3: Sample dispersion relation for STWs on quartz.

$$\approx -6.13\left(\frac{h}{\lambda}\right)^2 \tag{24}$$

$\Delta v/v$ is not a measure of the electromechanical coupling. It is the shift in the STW velocity, V_{STW}, from that of the SSBW, V_{SSBW}, which results from the loading of the surface.

5 Experiments

To demonstrate the viability of the proposed COM formalism for STW devices, several devices have been designed and fabricated. A pair of coupled resonator filters, one with transducer lengths of 50λ and one with transducer lengths of 80λ, will be used to demonstrate the reliability of the analysis method. The measured frequency responses of these filters are shown in figures 4 and 5, respectively. The modelled responses are given in figures 6 and 7, respectively. The experimental measurements were made using high frequency probes to measure die on wafer. To eliminate RF feed through the measurements were time gated. Comparison of the measurements and calculated results reveals two consistent differences between the analysis results and measurement. These "errors" occur on the low side of the pass band and about 10 MHz above the pass band. We have identified several possible causes for each of these discrepancies. These possible causes are, piezoelectric

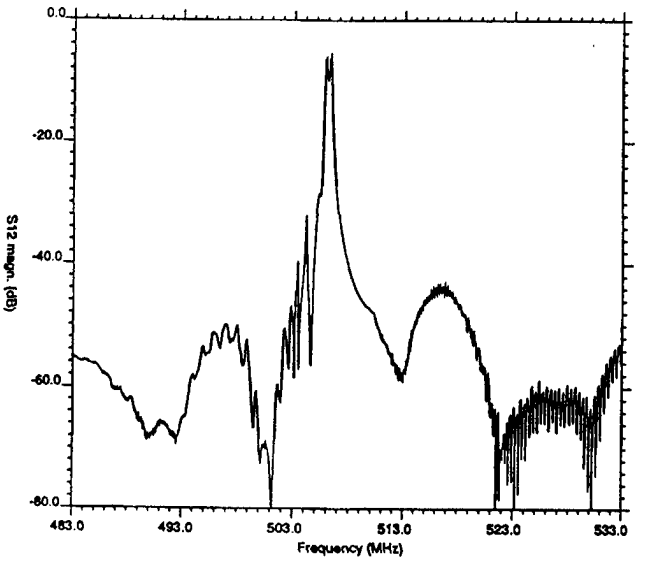


Figure 4: Experimental measurement of the CR filter with transducer lengths of 50λ.

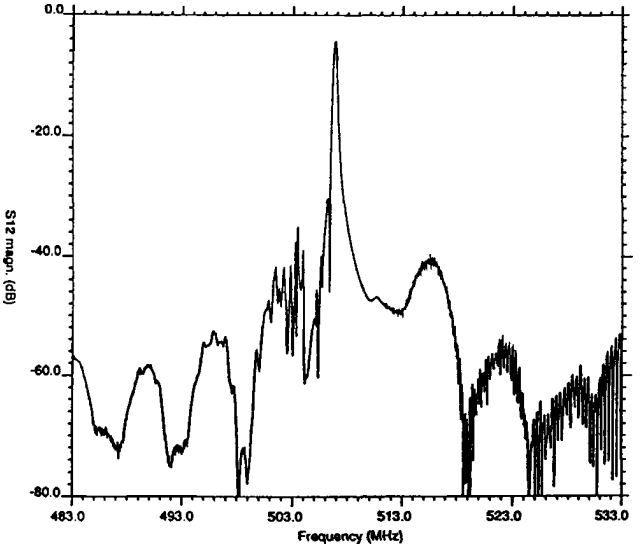


Figure 5: Experimental measurement of the CR filter with transducer lengths of 80λ.

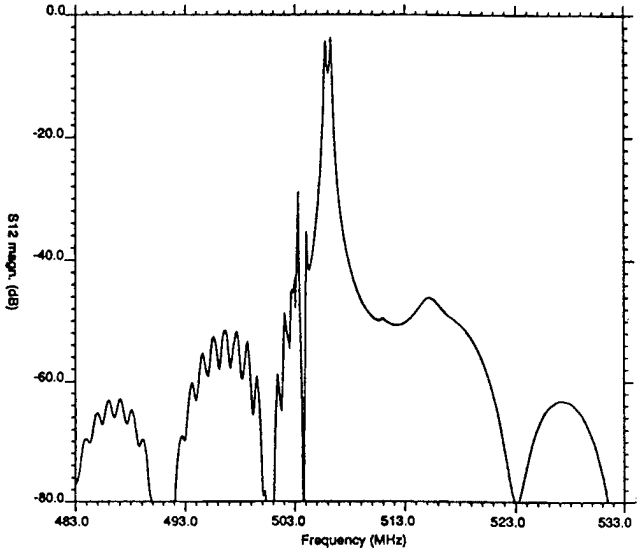


Figure 6: Theoretical response of the CR filter with transducer lengths of 50λ.

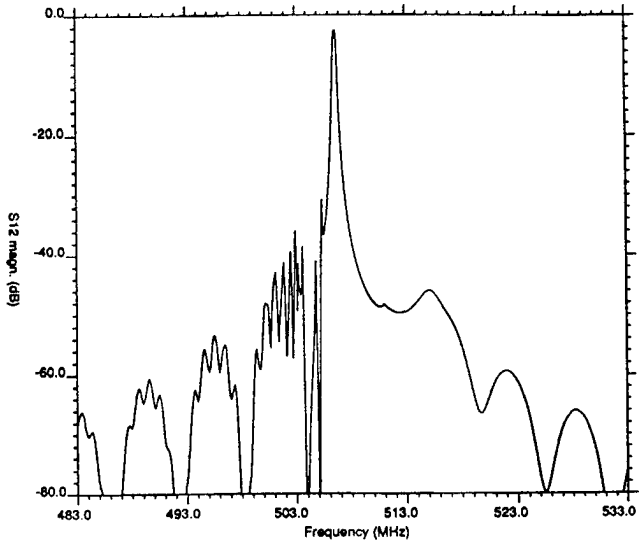


Figure 7: Theoretical response of the CR filter with transducer lengths of 80λ.

coupling to the SSBW, transverse modes, a complex reflection coefficient, and/or the non-rectangular electrode profile. Each of these possible error sources are discussed briefly below.

The response above the passband, seen in the experiment but absent in the theory, may be the result of piezoelectric coupling to the SSBW. Piezoelectric coupling between the IDT and SSBW is not included in the COM analysis. The energy storage and scattering losses resulting from reflective coupling between the STW and SSBW are included. However, the SSBW is treated as a loss mechanism, once a portion of the STW is converted into the SSBW it is no longer considered.

The errors just below the pass may be the result of transverse modes. Over a narrow range of angles, the parabolic approximation can be used to model the slowness surface of a substrate.

$$v(\theta) \approx v(0)(1-\gamma\theta^2) \tag{25}$$

For a convention quartz Raleigh wave substrate such as 36^0 Y-rotated quartz, γ is approximately +0.18. For YZ-$LiNbO_3$, γ is approximately -0.5. In the case of the conventional quartz substrate, the effect of the parabolic anisotropy is to increase the dispersion between the modes, while in the case of YZ-$LiNbO_3$ the effect is to greatly reduce this dispersion toward zero (i.e. minimum diffraction cut). In the case of 36^0 Y-rotated, $X+90^0$ propagating quartz, γ is approximately -1.85. Such a parabolic variation in velocity will force the transverse modes below the passband and could explain the discrepancy between the COM analysis and experiment.

There is one possible explanation for the errors both above and below the passband. A phase shift between the centers of transduction and reflection would modify the CR filters' responses both above and below the pass band. However, there is no experimental evidence of such a characteristic.

A final source of error is the shape of the electrode profile. The dispersion relation obtained using the discrete Green's function was calculated assuming that the electrode had an ideal rectangular shape. A non-rectangular electrode may have significantly different effect on STW propagation.

6 Conclusions

A new COM formalism which accurately represents the nature of STW propagation has been introduced. This COM formalism is based on the pioneering work of Plessky [1]. The application of the COM formalism to

the practical analysis of STW device has been demonstrated to produce good agreement with experimental measurement. Future work will concentrate on explicitly determining the source of the errors between the calculations and experiments.

7 Acknowledgments

The authors are grateful to D. Thompson, P. Wright, and M. Yamaguchi for their many helpful discussions during the development of this work.

References

[1] V. Plessky, "Two Parameter Coupling-of-Modes Model for Shear Horizontal Type SAW Propagation in Periodic Gratings", *IEEE Ultrason. Symp. Proc.*, pp. 195-200, 1993.

[2] K. Hashimoto and M. Yamaguchi, "Precise Simulation of Surface Transverse Wave Devices by Discrete Green Function Theory", *IEEE Ultrason. Symp. Proc.*, 1994.

[3] P. V. Wright, "A New Generalized Modeling of SAW Transducers and Gratings" *43rd Symp. Freq. Control Proc.* pp. 596-605, 1989.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

2 GHz Surface Transverse Wave Oscillators. Design, Performance and Limitations

Ivan D. Avramov[#], Vladimir S. Aliev[&], Sergey Denissenko[&], and Anatoliy S. Kozlov[&]

Institute of Solid State Physics, 72 Tzarigradsko Chaussee Blvd. 1784 Sofia, Bulgaria
& Institute of Semiconductor Physics, 13 Lavrentyev Avenue 630090 Novosibirsk, Russia

Abstract

The current sate-of-the-art for 2 GHz resonant devices and stable microwave oscillators using surface transverse waves (STW) is presented. Single mode low-Q resonators with an unmatched insertion loss of 4.85 dB, high-Q extended cavity resonators with an unloaded Q of 3800 and an insertion loss of 10 dB, as well as 2-pole coupled resonator filters with a 3 dB bandwidth of 4 MHz and an insertion loss within 10 dB are described and characterized in detail. It is shown that the major loss mechanism in STW resonators, operating in the 2 GHz range, is the ohmic loss limiting the maximum achievable unloaded Q to about 75% of the material Q for SAW at this frequency. High-performance feedback oscillators are designed in a straight forward procedure using transistor and STW device S-parameters. Both stripline and lumped component designs provide elegant solutions to miniature 2 GHz oscillators with a temperature stability within 100 ppm over a (-20...+70) deg. C range and an output power in excess of +21 dBm.

Introduction

Since Bagwell and Bray demonstrated the first successful high-Q resonator using the surface transverse wave (STW) mode in 1987 [1] a substantial amount of work has been dedicated to designing a variety of novel resonant devices. Combined mode resonator filters (CMRF) [2], low-Q resonators, extended cavity single mode and multimode high-Q resonators [3], impedance element narrow-band filters [4] and biosensors [5] take advantage of the well known unique properties of STW which allow significant improvement in system performance. Not long ago most SAW design engineers considered the 1.6 times higher propagation velocity as the only real advantage of STW compared to SAW. They were reluctant to use STW since the device performance depends not only on the device geometry but also on how well the wave energy is trapped to the surface of the crystal. Meanwhile the energy trapping mechanisms have been studied fairly well [1], [6], [7]. Moreover, the metallisation parameters, responsible for efficient energy trapping, are being used as an important tool for modifying the device loaded Q and modeling filter characteristics [2], [3], [7], [8]. This has allowed the design of 1 GHz resonant devices, microwave oscillators and systems benefiting from the real advantages of STW compared to their SAW counterparts. The extremely high power handling capability of STW resonators has been successfully implemented in 1 GHz power oscillators featuring an output power of 1 to 2 W, a thermal noise floor of -194 dBc/Hz and an efficiency of up to 36% [9]. Just a year ago 1 GHz high-Q STW resonators demonstrated residual 1/f phase noise levels in the -142 to -144 dBc/Hz range at 1 Hz intercept [10] which is a substantial improvement compared to SAW resonators in this frequency range. This significant advantage has been implemented in the series production of high performance 1 GHz STW oscillators featuring a close-to-carrier phase noise suppression of -35 to -38 dBc/Hz at 1 Hz intercept [11]. Finally, a very important advantage of STW resonant devices is the extremely low vibration sensitivity which makes them very attractive for state-of-the-art military applications [12]. Vibration sensitivities of $1*10^{-10}$/g were demonstrated recently with 1 GHz STW oscillators [11].

The fact that STW are doing better and better and demonstrate their real advantages at high frequencies, has inspired the authors of this paper to evaluate the performance of STW resonant devices at 2 GHz and higher. Another reason for this effort was that a new frequency range of 2.4 to 2.5 GHz was recently released worldwide for a variety of novel personal communication systems (PCS). Since the resolution of 0.5-0.6 μm, necessary for the reproducible low-cost fabrication of STW devices on piezoelectric quartz in this frequency range, is provided by most of the photolithographic equipment, available to date, we feel that STW resonator-, narowband filter- and oscillator technology may have a strong impact on low-cost PCS for large scale applications.

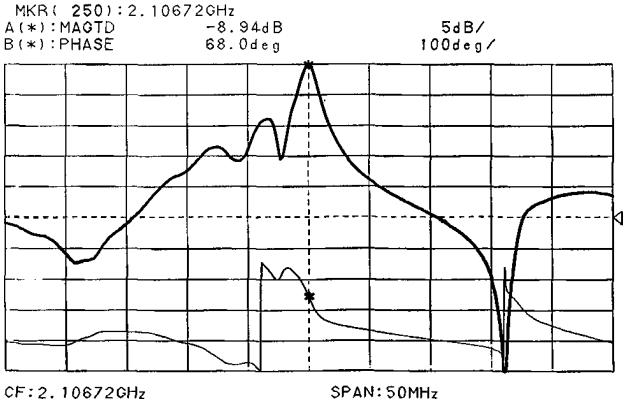
The paper will characterize different kinds of STW resonant devices which represent the current state-of-the-art with respect to device insertion loss, loaded and unloaded Q. The influence of the metallisation parameters on the device performance will be discussed by comparing 2 GHz devices with different metallisation. The loss mechanisms in 2 GHz STW resonators will be analyzed and it will be shown that the major mechanism limiting the device unloaded Q at 2 GHz and higher is the ohmic resistance loss of bus bars and interdigital transducer (IDT) fingers. The second part of the paper will present and characterize simple, high-performance 2.1 GHz STW feedback oscillators using low-cost bipolar transistors in the

0-7803-2500-1/95 $4.00 © 1995 IEEE

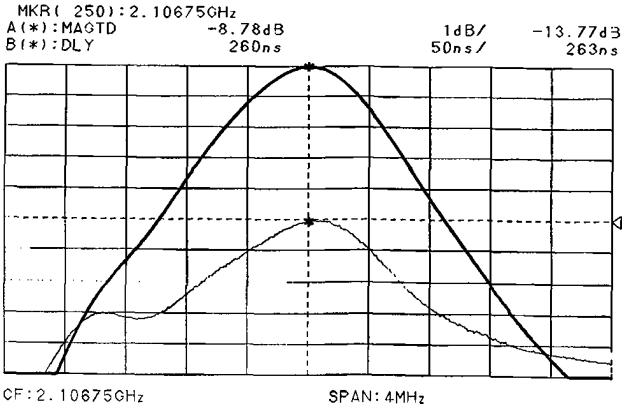
sustaining amplifier. Both stripline and lumped component designs are synthesized in a straight forward procedure using the S-parameters of the transistors and STW devices. Finally, the limits of expected phase noise performance of 2 GHz STW oscillators will be calculated using measured data available to date.

State-of-the-art 2 GHz STW resonant devices for oscillator applications

Our first efforts in designing and realizing 2 GHz STW resonators on AT-cut quartz were made in 1992. The results obtained at that time indicated that it was possible to control the Q in relatively simple two-port resonator configurations while retaining low device insertion loss and single mode operation [13]. Since then we have been able to achieve a substantial improvement of the device performance. Figure 1 a) and b) characterizes a low-Q STW resonator which was designed to provide a 1 MHz tuning range in a VCO. The loaded device Q is 1720 and the insertion loss is less than 9 dB.



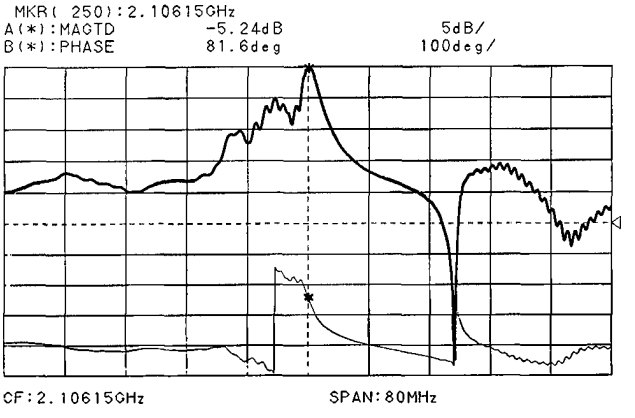
a)



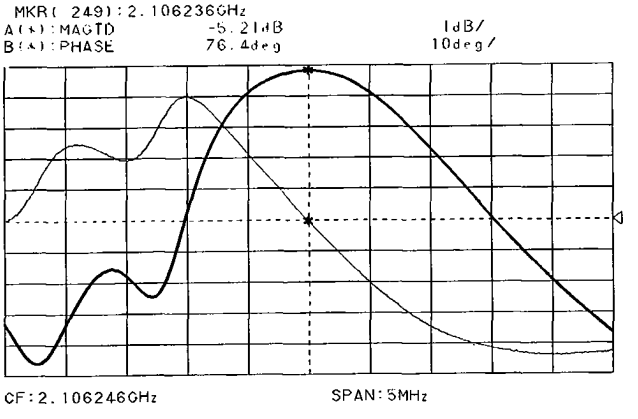
b)

Fig. 1 Characteristics of a 2.1 GHz low-Q STW resonator for VCO applications.
a) frequency and phase responses
b) detailed frequency and group delay responses.

Minimum insertion loss in STW resonators in a 50 Ω environment can be achieved by optimizing the device aperture and taking into account the bond lead inductance which can tune out the static IDT capacitance completely if the aperture is properly selected. In this case the device insertion loss can be made very low, however the device loaded Q will also decrease. The device characterized in Fig. 2 a) and b) uses the same geometry and metallisation parameters as the one from Fig. 1, however the aperture was optimized for minimum insertion loss. The value of 5.2 dB for the insertion loss was achieved at the expense of a lower device loaded Q of 1100.



a)

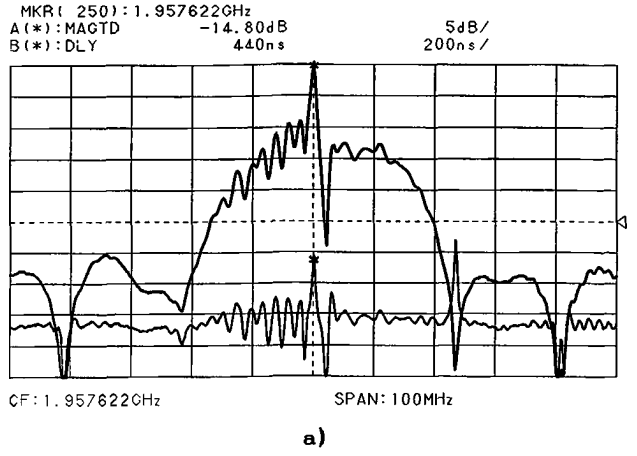


b)

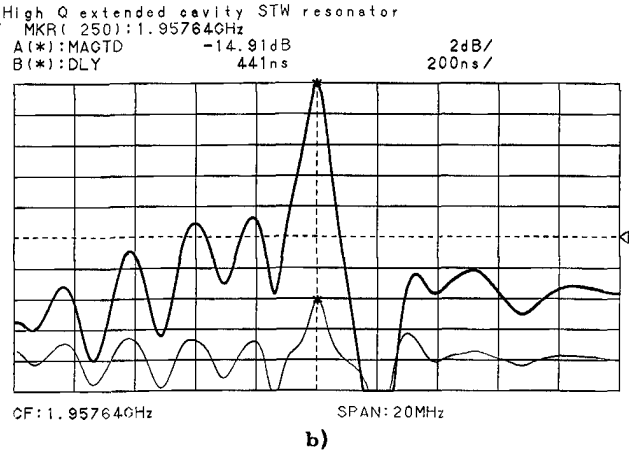
Fig. 2 Characteristics of a 2.1 GHz low-Q STW resonator optimized for minimum loss:
a) broadband and b) narrowband frequency and phase responses.

In a variety of applications in which the oscillator phase noise performance is of major significance it is necessary to maximize the device loaded Q. This can be achieved using the extended cavity single mode design described in [3]. This design uses a center waveguide grating with a length chosen in such manner that maximum group delay is achieved while single mode operation is retained. The frequency-, group delay- and phase responses of a typical extended cavity high-Q resonator design are shown in Fig. 3 a), b) and c).

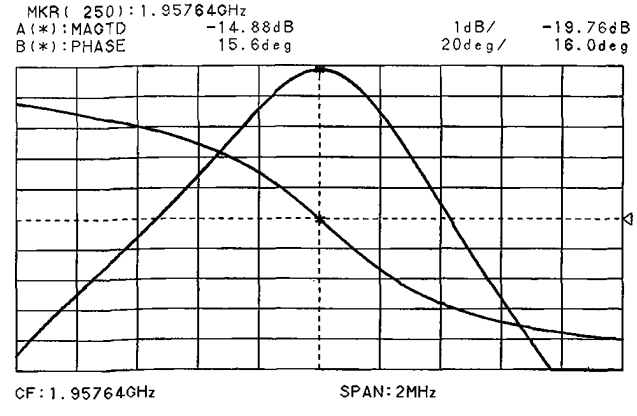
The upper curve in Fig. 3 a) shows a single sharp resonant peak which appears on top of the IDT frequency response. The loaded and unloaded Q values are 2930 and 3580 respectively and correspond to 68% of the material Q for SAW.



a)



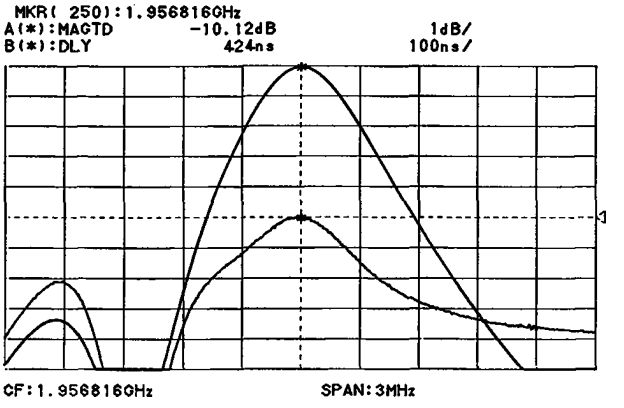
b)



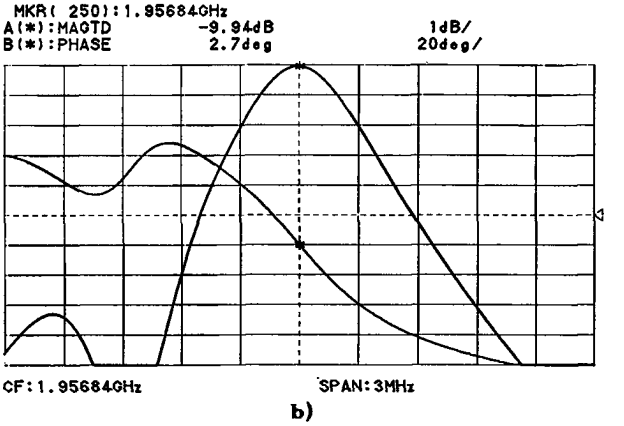
c)

Fig. 3 Characteristics of a 1.96 GHz extended cavity single-mode STW resonator: a) broadband and b) narrowband frequency and group delay responses; c) detailed frequency and phase responses of the resonant peak.

One of the best extended cavity single mode devices we fabricated is characterized in Fig. 4 a) and b). With a careful selection of the spacer in the cavity which is responsible for the optimum detection of the standing wave maxima by the IDT, and by optimizing the aperture, the device insertion loss was reduced to 10 dB and the loaded and unloaded Q values were measured as 2600 and 3800 respectively. This is as high as 70% of the material Q for SAW and to best of our knowledge is the highest Q value achieved with a 2 GHz acoustic resonator so far.



a)

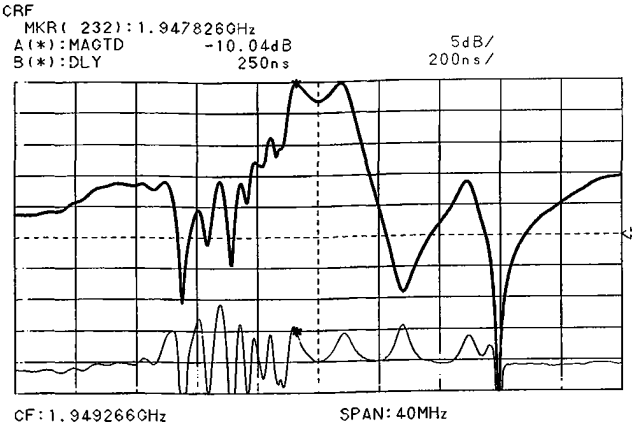


b)

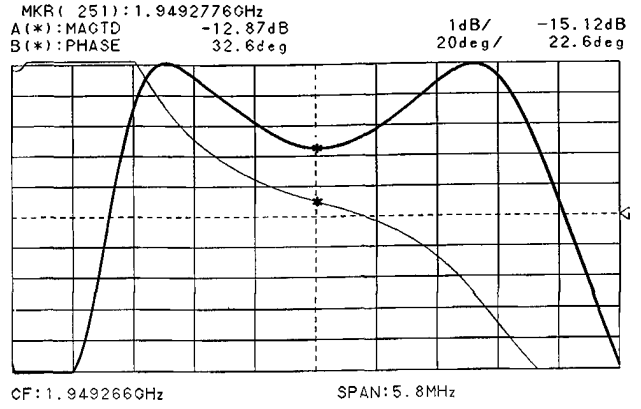
Fig. 4 Performance of the best device fabricated: a) frequency and group delay responses b) frequency and phase responses.

In some applications as microwave synthesizers and multichannel communication systems the tuning range that a low-Q STW resonator stabilized VCO provides may not be sufficient. Two-pole coupled resonator filter (CRF) designs can increase the VCO tuning range by a factor of 2 or more using fairly simple inline coupled cavities. Figure 5 a) and b) shows frequency-, group delay- and phase response data form a 1.95 GHz inline coupled CRF. It features an insertion loss of 10-13 dB, a 1 dB bandwidth of 3.5 MHz and a 3 dB bandwidth of 4.1 MHz. As evident from Fig. 5 b), this device features a very smooth S-shaped phase response providing 200 deg. of phase shift over the

entire 3 dB bandwidth. It was found to provide smooth tuning over a 4 MHz bandwidth in a VCO using a cascaded CLC varactor tuned phase shifter.



a)

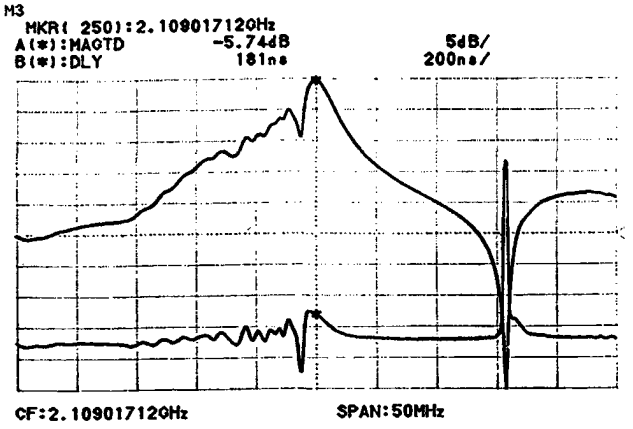


b)

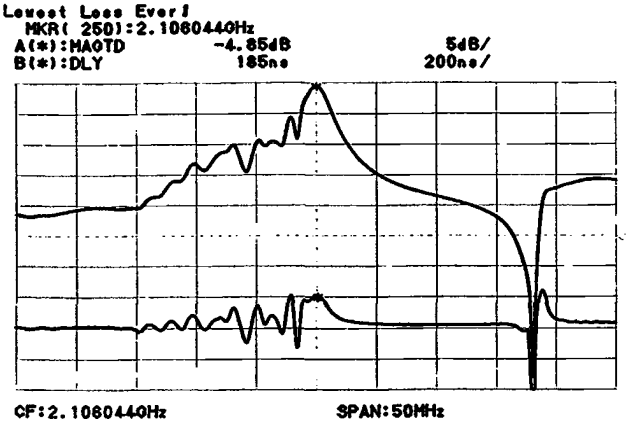
Fig. 5 Characteristics of a two-pole inline coupled 1.95 GHz STW based CRF:
a) frequency and group delay responses,
b) detailed frequency and phase responses.

Influence of the metallisation on the electrical performance of 2 GHz two-port STW resonators

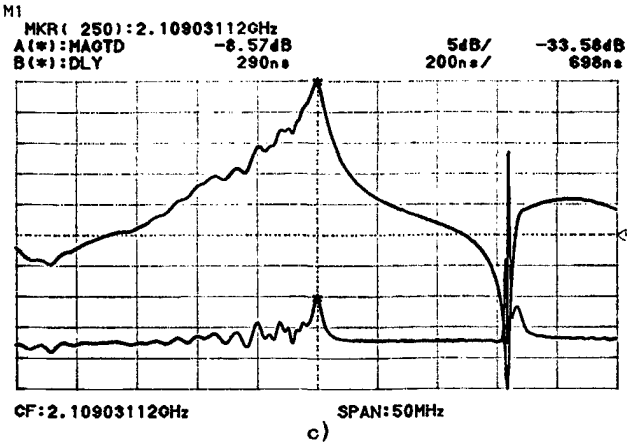
One of the most important parameters in the design of STW resonant devices are the film thickness and the finger-to-gap (f/g) ratio. Since these parameters are responsible for the energy trapping to the substrate surface they affect not only the device center frequency but also its insertion loss, sidelobe suppression, fractional bandwidth or loaded and unloaded Q in single mode resonators, as well as the behavior of the longitudinal modes in devices with longer cavities. That is why it is very important to know the effect of the metallisation parameters on the device performance before starting to design any practical devices. Fig. 6 a), b) and c) compares the frequency and group delay responses of three 2.1 GHz STW resonators with exactly the same geometry but different metallisation parameters. The devices in Fig. 6 a) and b) were fabricated with the same f/g ratio but the Al film of the device in Fig. 6 b) was by 50 Å thicker than in Fig. 6 a).



a)



b)



c)

Fig. 6 Frequency and group delay responses of 2.1 GHz STW resonators with different metallisation parameters:
a) Fo=2109 MHz, IL=5.74 dB, Ql=1200, Qu=2480
b) Al film thickness by 50 Å higher than a) Fo=2106 MHz, IL=4.85 dB, Ql=1200, Qu=2860
c) thicker film but lower f/g ratio than a) Fo=2109 MHz, IL=8.57 dB, Ql=1920, Qu=3060

The comparison of the devices from Fig. 6 a) and b) shows that, if the f/g ratio is kept constant, an increase of the film thickness by 50 Å causes a frequency downshift by 1400 ppm without any significant change of the overall frequency response. The loaded Q remains the same but the insertion loss decreases by 0.9 dB due to better energy trapping to the surface of the substrate. This results in an increase of the device unloaded Q by about 15%. The device from Fig. 6 b) demonstrates excellent energy trapping resulting in an unmatched insertion loss of only 4.85 dB. To the best of our knowledge this is the lowest loss obtained with a 2 GHz acoustic resonator so far.

The strong potential for using the metallisation parameters as a tool for controlling the device electrical performance is illustrated with the comparison of the device data from Fig. 6 a) and c). Both resonators have the same resonant frequency of 2109 MHz, however the device from Fig. 6 c) was fabricated with a thicker metal film and a lower f/g ratio than the device from Fig. 6 a). This results in a substantial increase of the device loaded Q by 60% while the insertion loss increases by only 2.8 dB. The control of the device Q by properly selecting the metallisation parameters has also the advantage that the device input and output impedances remain nearly the same and close to 50 Ω (see data in Fig. 1 c) and d) in Reference [10]), providing very good matching to the active circuitry in the oscillator loop. Of course, the device loaded Q can be increased electrically by connecting series capacitors to the resonator input and output. However, on one hand this approach seriously deteriorates matching. On the other hand, the device insertion loss increases dramatically. Figure 7 illustrates this with the frequency and group delay responses of the same device from Fig. 6 a) with two 0.47 pF capacitors in series. With a value of 2250 the device loaded Q increased just by 15% compared to the device from Fig. 6 c), but the insertion loss became 21.5 dB which is unacceptable for practical applications. Therefore, one should try to maximize the device loaded Q acoustically, rather than electrically.

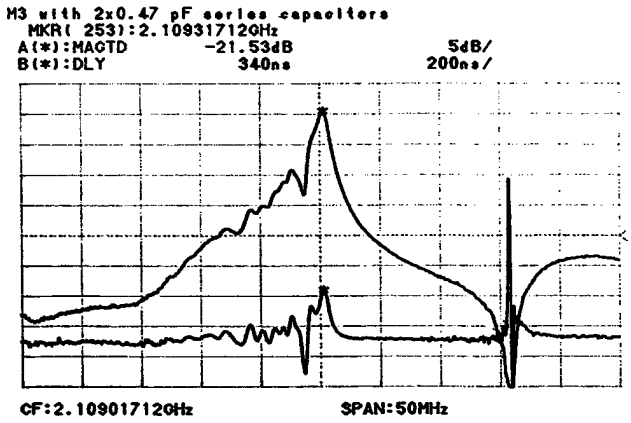


Fig. 7 Frequency and group delay responses of the device from Fig. 6 a) with two 0.47 pF capacitors in series: Q1=2250, IL=21.5 dB.

Loss mechanisms in 2 GHz STW resonators

In order to evaluate the limits of 2 GHz STW resonator technology in terms of maximum device unloaded Q and minimum insertion loss, it is necessary to investigate the loss mechanisms in devices operating at 2 GHz and higher. A fairly comprehensive model, analyzing the loss mechanisms in 2-port SAW resonators, is described in [14]. According to this model the resonator unloaded Q factor Q_u can be described as a function of the effective cavity length m in acoustic wavelengths as follows:

$$Q_u = \frac{2\pi m}{\mu_{mat} + \mu_{ohm} + \mu_{other}}, \qquad (1)$$

where μ_{mat} and μ_{ohm} are the material loss and ohmic loss coefficients respectively, and μ_{other} includes all other loss mechanisms as diffraction, losses due to leakage from the reflectors, air loading, bulk wave generation etc. Since so far no data has been published on the material Q value Q_{mat} for STW on quartz we take the material Q for SAW as a reference and assume that the material Q for STW is the same as for SAW. According to [14] the SAW material Q value is:

$$Q_{mat} = (1.07\times10^7) \,/\, F_o(\mathrm{MHz}) = 2\pi m \left\{\frac{1}{\mu_{mat}}\right\}, \qquad (2)$$

where F_o is the device frequency. Then from (1) and (2) it follows that:

$$\mu_{ohm} + \mu_{other} = \mu_{mat}\left(\frac{Q_{mat}}{Q_u} - 1\right) \qquad (3)$$

Again, using the equations in [14], it is easy to show that μ_{ohm} can be calculated as:

$$\mu_{ohm} = 32\,\hat{G}\,R_{ohm}, \qquad (4)$$

where $\hat{G}$ is the maximum value of the IDT conductance in the resonator and R_{ohm} is the ohmic IDT and bus bar resistance calculated from the relative metal film resistance ρ, the finger length W, the finger width a, the film thickness h and the number of finger pairs N as:

$$R_{ohm} = 2\rho W \,/\, ahN \qquad (5)$$

For thin films the ρ value is different from ρ_o

which is used for a massive piece of metal, and depends on the film thickness as follows [15]:

$$\rho = \rho_o(1 + 3l_o / 8h), \text{ for } l_o > h \text{ and} \quad (6)$$

$$\rho = (4/3)\,\rho_o \left[0.42 + \ln (h/l_o)\right], \text{ for } l_o < h \quad (7)$$

Assuming aluminum metallisation $\rho_o = 2.7\times10^{-6}$ Ωcm and the electron free path $l_o = 3.29\times10^{-6}$ cm.

As an example we will calculate the ohmic loss coefficient of our best 1.96 GHz STW resonator characterized in Fig. 4 a) and b). With an insertion loss of 10 dB, a loaded and unloaded Q of 2600 and 3800 respectively and an effective cavity length of 750 [3], Q_{mat} and μ_{mat} were calculated as:

$$Q_{mat} = 5470 \quad \text{and} \quad \mu_{mat} = 0.86$$

With $\hat{G}$ = 2 mS (extracted from measurements of the input and output resonator impedances) and R_{ohm} = 4.8 Ω (calculated with (5) and (6)), the ohmic and other loss coefficients were calculated as:

$$\mu_{ohm} = 0.31 \quad \text{and} \quad \mu_{other} = 0.07$$

In this example the comparison of μ_{ohm} and μ_{other} clearly indicates that in this very good resonator the ohmic resistance of IDT fingers and bus bars causes a loss which is by a factor of 4.4 higher than the other losses. This makes one conclude that, if STW resonators, operating at 2 GHz and higher, are designed for maximum device Q, the ohmic loss will be the major factor limiting the device Q and insertion loss. Our calculations show that improved designs will be able to achieve an unloaded Q value of about 4100 which is as high as 75% of the material Q for SAW. To minimize the ohmic loss one should try to:

- minimize the acoustic aperture,
- maximize the number of IDT finger pairs,
- maximize the finger width,
- maximize the film thickness and
- minimize the bus bar length or deposit bus bars with increased film thickness in a second photolithographic step.

2 GHz STW oscillators using S-parameter designs

Since STW resonant devices have very low insertion loss and input and output impedances that are relatively close to 50 Ω it is fairly easy to design and build simple feedback oscillators with them. All one needs to know, in order to design a STW based feedback oscillator, are the S-parameters of the transistors, STW devices, phase shifters, power splitters and other circuitry one intends to include in the loop. Then the design procedure is performed just by selecting proper circuits for matching the complex impedances of the various components in the loop to each other. This can be done by using either strip line transformers or lumped component T, L or Π type LC circuits. The only adjustment that one needs after completing the design is to set the phase around the loop so that the oscillator operates, or is tuned over the 1-3 dB bandwidth of the STW device. This adjustment can easily be performed by tapping a phase adjust transmission line included in the loop. Also, it is necessary to make sure that the gain compression does not exceed 2-3 dB otherwise the oscillator phase noise performance may be deteriorated [10].

In this study we investigated two different designs based on the low-cost small signal microwave transistors BFG 67 made by Philips. According to the data sheets the S-parameters of these transistors at two different collector currents I_c = 5 and 20 mA and a collector-emitter voltage of 8 V are as follows:

I_c = 5 mA (lowest noise figure):

$S_{11} = 0.49 \exp(149.8^\circ)$; $S_{22} = 0.38 \exp(-89.5^\circ)$

I_c = 20 mA (high output power):

$S_{11} = 0.45 \exp(137.8^\circ)$; $S_{22} = 0.22 \exp(-100.2^\circ)$

The first oscillator design, shown schematically in Fig. 8, was realized entirely in stripline technology. A short 29.5 Ω transmission line is used to match the impedances of the STW device and the first transistor stage and another 40.7 Ω piece of line matches the second to the first transistor stages. The matching of the second transistor

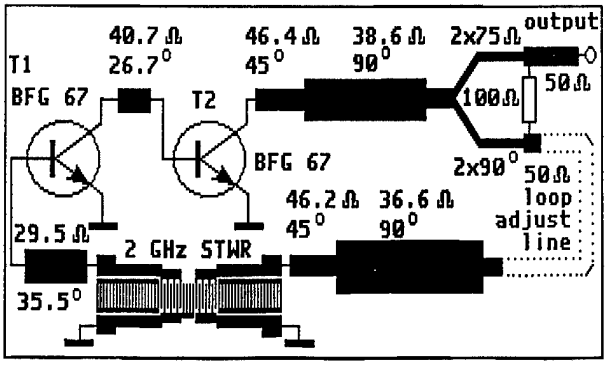


Fig. 8 Schematic of a 2 GHz stripline STW oscillator.

stage output impedance and the STW device to the 50 Ω impedances of the stripline Wilkinson power splitter for load coupling is performed by means of 3/8 λ transformers. The broadband frequency response of the two-stage sustaining amplifier from Fig. 8 is shown in Fig. 9. This frequency response was measured without STW device and Wilkinson power splitter. The 14 dB of gain conveniently compensates for a 9 dB STW device insertion loss (see for example the device in Fig. 1), and the 3.5 dB loss of the power splitter providing a gain compression of about 1.5 dB. The output power for this gain compression value was measured as 20 dBm.

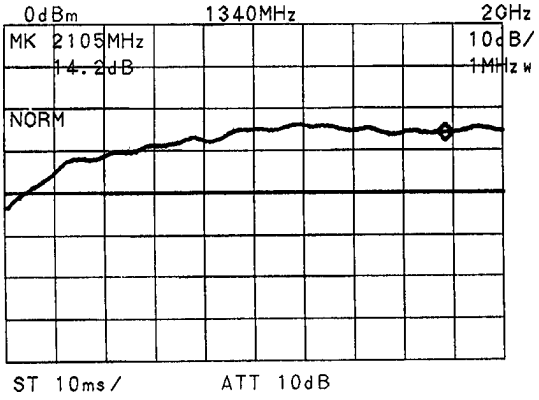


Fig. 9 Broadband frequency response of the 2-stage stripline amplifier from Fig. 8.

The stripline design was found to provide excellent oscillator performance but the printed circuit board (PCB) was fairly big. We were able to reduce the oscillator size by a factor of 5 with the lumped component design in Fig. 10. Here the stripline transformers were replaced by T and Γ type lumped LC circuits. The inductors were printed on the PCB and the capacitors were surface mounted devices (SMD) chips. Replacing the matching capacitors C1 or C2 by varactors makes the oscillator tunable. With this approach we were able to achieve fairly linear tuning over the 3 dB

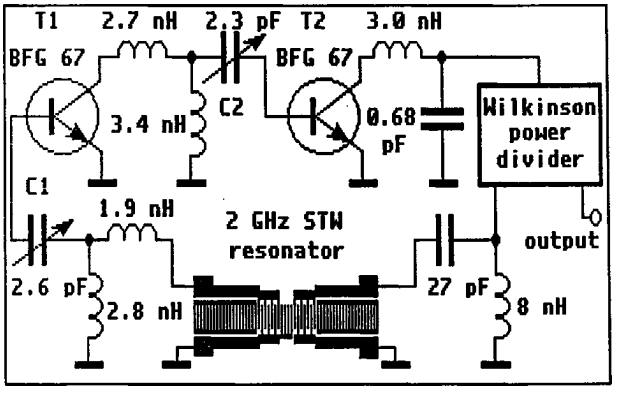


Fig. 10 Lumped component 2 GHz STW oscillator circuit.

STW device bandwidth. The performance of the lumped component circuit was very similar to the stripline one however, with this design we were able to realize the whole oscillator circuit on a miniature 4 cm^2 glass epoxy PCB. We believe that the circuit area could be reduced to 1 cm^2 if realized in thin film hybrid technology. Figure 11 shows the extremely clean output spectrum of one of the laboratory prototypes. The oscillator output power is 21.6 dBm while the loop power was measured as 25 dBm. These values were achieved with a collector current of the second transistor increased to 35 mA. With a loop loss of 11 dB and a noise figure of 4 dB this oscillator should have a noise floor of -184 dBc/Hz. A further noise floor reduction can be achieved if STW devices with an insertion loss of 5 dB (see Fig. 2 and 6 b)), and transistors with an output power of 1 W are used since STW devices can conveniently stand such high power levels [9], [10].

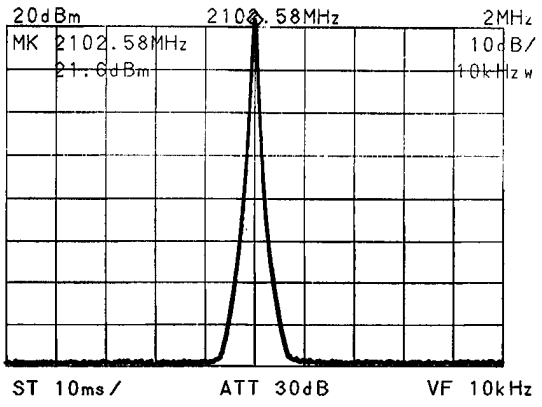


Fig. 11 Output spectrum of one of the 2.1 GHz laboratory prototypes.

Figure 12 presents typical temperature stability data obtained from the very first and the second temperature cycles on one of the lumped

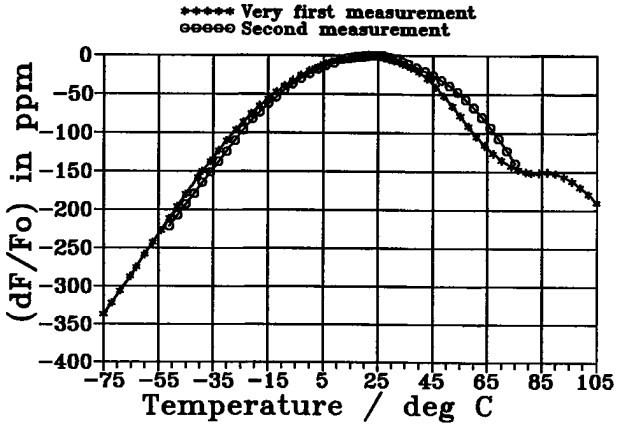


Fig. 12 Temperature stability data of a 2.1 GHz STW oscillator obtained from the first 2 temperature cycles.

component oscillator designs. It is evident that the curve of the first cycle is fairly distorted. The reason for that are static strain effects at the interface quartz - metal film [16]. We found that generally 2 temperature cycles are sufficient to release static stress and this is evident from Fig. 13 which is a detailed view of the temperature dependence of the second cycle. This curve is a

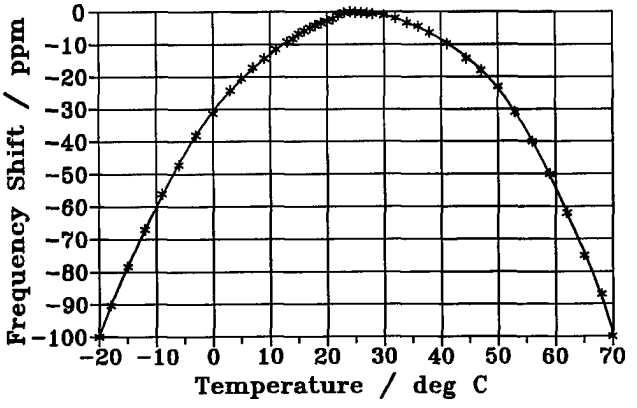


Fig. 13 Detailed temperature dependence of the 2.1 GHz STW oscillator measured during the second temperature cycle.

slightly unsymmetric parabola with a turn-over temperature of 25° C and an overall frequency shift of 100 ppm over the temperature range (-20...+70)° C. The slope above the turn-over temperature is slightly higher than below it which indicates the existence of an inversion point at very low temperatures and therefore a cubic temperature behavior of STW on AT-cut quartz [1].

Phase noise performance limits of 2 GHz STW oscillators

To evaluate the phase noise limits that we have to expect from 2 GHz STW oscillators we use the model described in [14]. According to this model the single side band phase noise $\mathcal{L}(f)$ of SAW two-port resonator based feedback oscillators which can be measured with a phase noise measurement setup, is:

$$\mathcal{L}(f) = 10 \log \left[\frac{S_\Phi(f)}{2} \right] \qquad (8)$$

Here $S_\Phi(f)$ is the power spectral density of the phase fluctuations which is expressed as:

$$S_\Phi(f) = [\alpha_R F_o^4 + \alpha_E (F_o/(2Q_L))^2]/f^3 + [(2GFKT/P_o)(F_o/(2Q_L))^2]/f^2 + (2\alpha_R Q_L F_o^3)/f^2 + \alpha_E/f + 2GFKT/P_o, \qquad (9)$$

where:

G - compressed loop amplifier power gain
K - Boltzmann's constant
T - temperature in K
P_o - carrier power level (in Watts)
F_o - carrier frequency (in Hz)
f - carrier offset frequency (in Hz)
Q_L - resonator loaded Q in the loop ($Q_L = \pi F_o \tau_g$)
α_R - resonator flicker noise constant
α_E - flicker noise constant of the amplifier
τ_g - resonator loaded group delay at F_o.

The resonator flicker noise constant can be calculated from the measured values of the residual resonator 1/f phase noise $\mathcal{L}(f)$ as:

$$\alpha_R = \left\{ \frac{1}{(2Q_L F_o)^2} \right\} f \; 2\times 10^{\left\{ \frac{\mathcal{L}(f)}{10} \right\}} \qquad (10)$$

The residual phase noise value $\mathcal{L}(1\ \text{Hz})$ of a state-of-the-art 1 GHz STW resonator at f = 1 Hz offset was measured as -144 dBc/Hz [10]. Then with the measured values of:

Q_L = 2740, F_o = 1002 MHz and $\mathcal{L}(1\ \text{Hz})$ = -144 dBc/Hz

the flicker noise constant of this STW resonator was calculated as:

$$\boxed{\alpha_R = 2.64 \times 10^{-40}} \qquad (11)$$

Since 2 GHz STW resonators are fabricated using the same technology as 1 GHz ones and since α_R is independent of the device frequency [14] then we assume that this α_R value will be true at 2 GHz too. Furthermore we assume that very soon 2 GHz silicon bipolar amplifiers with very low 1/f residual phase noise performance will be available on the market so that the STW device will be the dominant source of 1/f noise. Under these assumptions we can consider the amplifier as noiseless and calculate the expected phase noise performance limits of 2 GHz STW oscillators by substituting in equations (2) and (1) the following performance limits that we expect from STW devices

and amplifiers at this frequency:

F_o = 2000 MHz
$\alpha_R = 2.64 \times 10^{-40}$
α_E = 0
Q_L = 3000
G = 10 dB
F = 4 dB
P_o = 30 dBm (1 Watt)

The result is plotted in Fig. 14 and indicates a close-to-carrier phase noise suppression of -26 dBc/Hz at 1 Hz intercept and a thermal phase noise floor of -190 dBc/Hz. Again, this plot assumes a 1/f noiseless amplifier and is, therefore, fairly idealized. It just indicates what one should aim

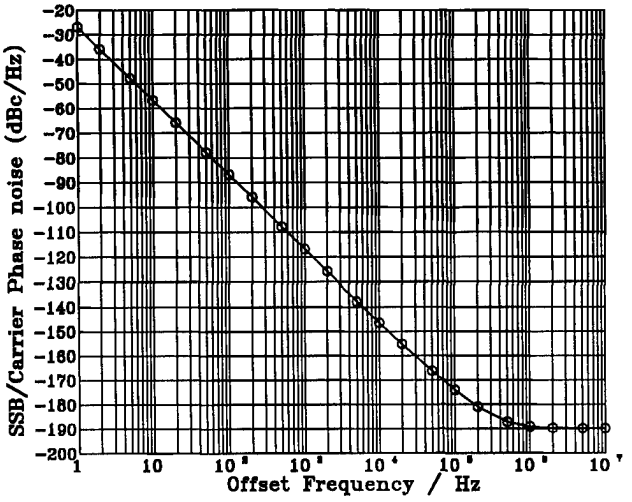


Fig. 14 Phase noise performance limits of 2 GHz STW oscillators assuming a 1/f noiseless amplifier.

at when designing 2 GHz high-performance STW oscillators. Practically, a good STW oscillator at this frequency will have a 3 to 6 dB higher phase noise than the plot in Fig. 14. Still, keeping in mind that this phase noise performance is very difficult to achieve with dielectric resonator oscillators (DRO), but can be obtained at a low-cost with a miniature, highly efficient and temperature stable STW oscillator with extremely low vibration sensitivity [11], one can see the great potential of 2 GHz STW oscillator technology for future applications.

Summary and conclusions

1. We have demonstrated state-of-the-art 2 GHz STW resonant devices featuring:

- an insertion loss of 4.85 dB at 2.1 GHz (low-Q resonator),

- a loaded Q of 2600, an unloaded Q of 3800 and an insertion loss of 10 dB (extended cavity high-Q 1.96 GHz device),

- a 1 dB bandwidth of 3.5 MHz and an insertion loss of 10 dB (1.96 GHz coupled resonator filter).

2. The main factor, limiting the device unloaded Q to about 4100 (75% of the material limit for SAW) at 2 GHz, is the ohmic resistance of the IDT and bus bars. For frequencies higher than 2 GHz the limiting effect of the ohmic loss will increase with device frequency due to smaller finger dimensions and lower film thickness.

3. The variation of the metallisation parameters (film thickness and finger-to-gap-ratio) in devices fabricated with the same geometry allow a variation of the device loaded and unloaded Q in wide limits while keeping the same device center frequency. This feature is unique for STW.

4. With careful control of the metallisation parameters a loaded Q value of 3000 can readily be achieved with an insertion loss within 10 dB.

5. S-parameters allow greatly simplified and straight forward designs of highly efficient, high-output-power 2 GHz fixed frequency and voltage controlled STW oscillators which can be realized at a low cost in both strip line and miniaturized discrete component hybrid technology. We demonstrated an output power of 21.6 dBm with a 2.1 GHz low-cost STW oscillator.

6. The phase noise performance of 2 GHz STW resonator oscillators is limited to -26 dBc/Hz for the 1 Hz intercept and -190 dBc/Hz for the noise floor, assuming that the STW device is the dominant source of 1/f noise. Achieving these values will require further work towards designing high-power 2 GHz amplifiers with greatly improved 1/f residual phase noise.

7. The temperature stability of 2 GHz STW oscillators is within 100 ppm over the temperature range (-20 ... +70) deg. C. At least two temperature runs are necessary to eliminate static strain effects.

8. 2.4 to 2.5 GHz STW oscillator technology may be extremely competitive in newly emerging personal communication systems in this frequency range.

Acknowledgements

The authors wish to gratefully acknowledge E. Yossifov from the Institute of Solid State Physics

in Sofia, Bulgaria for the expert preparation of the quartz substrates, as well as M. Dolgov, A. Meier, S. Sudarikov, N. Matveenkova, G. Strelnikova and M. Kachanova from the Institute of Semiconductor Physics in Novosibirsk, Russia for their help in realizing the STW devices.

References

[1]. T. L. Bagwell and R. C. Bray, "Novel Surface Transverse Wave Resonators with Low Loss and High Q", IEEE 1987 Proc. Ultrasonics Symposium, pp. 319-324.

[2]. I. D. Avramov, "1 GHz Voltage Controlled Oscillator Using a SSBW/BGW Combined Mode Resonator Filter", Proceedings of the 45-th Annual Frequency Control Symposium AFCS-91, pp. 230-238.

[3]. I. D. Avramov: "GHz Range Resonant Devices for Oscillator Applications Using Shear Horizontal Acoustic Waves", IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, Vol. 40, No. 5, September 1993, pp. 459-468.

[4]. J. Heighway, S. N. Kondratyev and V. P. Plessky, "Impedance Element SAW Filters", 1994 IEEE Proc. International Frequency Control Symposium.

[5]. R. McGowan, J. Foerster, R. Lindenmuth, D. Huynh and T. Lukaszek, "Chemical/Biological Contaminant Detector for Aqueous Environments: Preliminary Report", 1994 IEEE Proc. International Frequency Control Symposium.

[6]. D. F. Thompson and B. A. Auld, "Surface Transverse Wave Propagation Under Metal Strip Gratings", IEEE Proc. 1986 Ultrasonics Symposium, pp. 261-266.

[7]. F. Bigler, B. A. Auld, E. Ritz and E. Sang, "Design and Test of Surface Transverse Wave (STW) Resonators on Quartz", Proc. 5th Conference "Acoustoelectronics'91", 10-13 Sept. 1991, Varna, Bulgaria, pp. 8-22.

[8]. R. Almar, B. Horine and J. Andersen, "High Frequency STW Resonator Filters", Proc. IEEE 1992 Ultrasonics Symposium, pp 51-56.

[9]. I. D. Avramov, "Extremely High Loop Power GHz Range Surface Transverse Wave Oscillators Using AB-Class Amplifiers", Proc. 1993 IEEE International Frequency Control Symposium, pp. 728-732.

[10]. I. D. Avramov, F. L. Walls, T. E. Parker, G. K. Montress, "Extremely Low Thermal Noise Floor Power Oscillators Using Surface Transverse Waves, Proc. IEEE 1994 Frequency Control Symposium, pp. 379-394.

[11]. M. S. Cavin and R. C. Almar, "An Oscillator Design Using Low G-Sensitivity, Low Phase Noise STW Devices", Proc. IEEE 1995 International Frequency Control Symposium, (this Proceedings, in press).

[12]. D. Huynh, R. McGowan, J. Kossinski, J. Stewart, R. Piekarz and C. D. Mulford, "Experimental Studies of SAW and STW Acceleration Sensitivitiy", Proc. IEEE 1995 International Frequency Control Symposium, (this Proceedings, in press)

[13]. S. Denisenko, I. D. Avramov, "Design, Fabrication and Performance of 2 GHz Surface Transverse Wave Resonators", Proc. 1993 IEEE International Frequency Control Symposium, pp. 645-649.

[14]. T. E. Parker and G. K. Montress, "Introduction to High Stability SAW Oscillators: Design and Performance", Tutorial presented at the 1995 IEEE International Frequency Control Symposium.

[15]. "Design and Fabrication of SAW Components in the 500-800 MHz Frequency Range", Research report of the Acoustoelectronics Laboratory, Institute of Solid State Physics, Sofia, Bulgaria, 1983.

[16]. D. F. Thompson, "Theory and Application of Surface Transverse Waves", Proc. 6-th Conference Acoustoelectronics'93, Sept. 19-25 1993, Varna, Bulgaria, pp. 51-60.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DESIGN AND TEST OF 3 GHz, FUNDAMENTAL MODE STW RESONATORS ON QUARTZ

S. Denissenko*, E. Gavignet, S. Ballandras, E. Bigler, E. Cambril**

Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon - 32 avenue de l'Observatoire - 25044Besançon Cedex - France

* Institute of Semiconductor Physics, Russian Academy of Sciences,
Siberian branch, Pr. Acad. Lavrentjeva 13 - 630090 Novosibirsk - Russia

** Laboratoire de Microstructures et Microélectronique du CNRS
196 avenue H. Ravera - 92225 Bagneux Cedex

Abstract

Surface transverse wave (STW) resonators, based on the propagation of high velocity shear horizontal waves on Y-rotated quartz are designed, fabricated and tested. A model is presented to predict the resonant frequency of a 3-grating structure as a function of design parameters such as : periodicities, metal thickness, finger-to-gap ratio. Experimental devices have been fabricated by direct e-beam lithography with linewidth geometries in the range of 0.3 - 0.5 µm, operating frequency : 3 GHz in fundamental mode. Two different designs using either a quasi synchronous structure (type # 1) or a change of periodicity inside the cavity (type # 2) have been tested. Best results obtained at 3 GHz are : unloaded Q-factor Qu = 2 420, insertion loss $I.L.$ = −10.3 dB (type # 1) and Qu = 1 920, $I.L.$ = −9.5 dB (type # 2). The best experimental factor of merit $Qu \times f = 7.5 \times 10^{12}$ is close to the best results already published for quartz STW resonators.

Introduction

Developments of high frequency oscillators and filters above 1 GHz have pushed the research in the field of high velocity surface modes like surface transverse (STW) modes on quartz.

The advantages of such modes are now well-known [1-4]
- existence of temperature-compensated cuts
- 60 % higher velocity than Rayleigh modes yielding higher operating frequencies
- higher power handling [2] yielding reduced thermal noise floor of oscillators.

However, the advantages of STW modes over Rayleigh modes have been up to now balanced by a lack of experimental data on the sensitivity of the devices to design fabrication parameters like : geometrical dimensions (periodicities, finger to gap ratio, number of cavity strips, acoustic aperture, ...) and technological parameters (patterning method, metal deposition technique, ...).

In the present paper an analysis of STW propagation under metal strips is presented for Y-rotated quartz [5]. From the propagation parameters, the resonant frequency of a 3-grating structure (mirror-cavity mirror) is derived taking into account the basic parameters of the design : quartz cut angle, metal thickness, finger-to-gap ratio, linewidths and periodicities. Devices are designed for an operating frequency of 3 GHz, on (Y+36°) quartz, corresponding to a grating periodicity of 0.8 µm.

Two kinds of resonator design have been used in parallel
- type # 1 is similar to a design presented by Denissenko and Avramov [1] where the cavity has the same periodicity as the reflectors, a small gap between the cavity and one of the reflectors being allowed to adjust the resonator behavior,
- type # 2 is based on a design used by Bigler *et al.* [3] originally proposed by Bagwell and Bray [6] where the cavity has always an integer number of strips, but with a slightly different periodicity.

Devices have been fabricated at LMM-CNRS (Bagneux, France) by direct e-beam lithography and patterning of aluminum strips was done by the lift-off technique.

A good agreement has been found between the theoretical predictions of the resonant frequency and already published data [7, 8].

Piezoelectric model of STW resonators

Propagation of a STW mode

The model of a STW resonator is based on the analysis of STW propagation on a piezoelectric substrate [5]. The analysis focuses on a single mode approach, and does not take into account parasitic modes that could be generated by the transducer. A solution for STW propagation under a periodic array of metal strips or grooves can be taken as

0-7803-2500-1/95 $4.00 © 1995 IEEE

$$u_1 = \sum_{n=-\infty}^{+\infty} \left(\sum_{r=1}^{2} A_{nr}\, e^{-\alpha_{nr} a_2} \right) e^{-j\beta_n a_3}\, e^{j\omega t}$$

$$\Phi = \sum_{n=-\infty}^{+\infty} \left(\sum_{r=1}^{2} C_{nr} A_{nr}\, e^{-\alpha_{nr} a_2} \right) e^{-j\beta_n a_3}\, e^{j\omega t} \qquad (1)$$

where the propagation constant β_n obeys the Bloch-Floquet condition

$$\beta_n = \beta_o + \frac{2\pi n}{\Lambda} \qquad (2)$$

In eq. (1), the mechanical displacement and electric potential fields respectively noted u_1 and Φ will be expressed in the plates axes (a_1, a_2, a_3) (Fig. 1) under the form of an infinite sum of modes guided by the periodic structure, each composed a two partial waves as in the case of Bleustein-Gulayev's waves propagating on cubic media [9]. The propagation direction a_1 corresponds to a θ-rotated (Z') axis with respect to the crystallographic axes (XYZ) of quartz (Fig. 1).

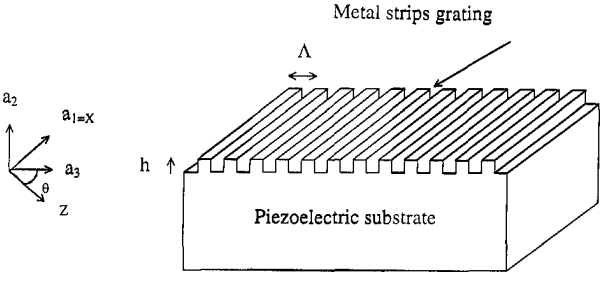


Fig. 1 : *Geometries for STW propagation under metal strips or grooves on a piezoelectric substrate*

The general boundary conditions can be expressed as a combination of mechanical stress condition (stress free interface to vacuum, continuity of stresses between the crystal and metal strips) and electrical conditions (continuity of electrical potential between the crystal and vacuum, zero potential condition on grounded electrodes).

To simplify the boundary conditions, the following approximations were made [5]

- the metal thickness is limited to a small fraction of the acoustic wavelength, thus allowing a perturbation approach
- to decouple mechanical from electrical condition, a condition of zero potential at the surface was added, corresponding to the academic case of a grounded crystal surface, periodically mechanically loaded by elastic metal strips.

In this approach, valid for weakly coupled piezoelectric crystals like quartz, an arbitrary shape of mass-loading electrodes can be taken into account, where as the coefficients C_{nr} defining the electrical potential can be directly related to the coefficients of harmonic expansion of the mechanical displacements as

$$C_{nr} = \frac{e_{26}\, a_{nr}^2 - e_{35}\, \beta_n^2 + j(e_{25}+e_{36}) a_{nn}\, \beta_n}{\varepsilon_{22}\, a_{nr}^2 - \varepsilon_{33}\, \beta_n^2 + 2j\, \varepsilon_{23}\, a_{nr}\, \beta_n}\, a_{nr} \qquad (3)$$

where e_{ij}, ε_{ij} represent the piezoelectric and dielectric constants of the crystal.

The problem is then solved for a given angular frequency ω by finding numerically the roots of two coupled nonlinear, complex equations of the parameters α_{nr}, β_n and ω [5].

Results are shown on Figs. 2 and 3 under the form of dispersion curves (ω, $\beta \equiv \beta_0$) plotted in the case of rectangular aluminum strips. All results are directly computed from material constants of quartz and aluminum, and from the geometries of the periodic array of metal strips.

Finding the resonant condition for a 3-grating structure

To predict the resonant condition for a 3-grating structure (mirror-cavity-mirror) a COM approach is followed, similar to the same model already proposed for nonpiezoelectric STW propagation [3, 10]. From the expressions of the mechanical displacement u_1 (Eq. 1), the forward and backward propagated amplitudes of the equivalent COM Model u_{11}^+ and u_{11}^- (Fig. 4) can be expressed as

$$\begin{pmatrix} u_{11}^{+} \\ u_{11}^{-} \end{pmatrix} = \begin{pmatrix} \sum_{n=0}^{+\infty} x_n\, e^{-j(\frac{\pi}{\Lambda}(2n+1)+\delta\beta)\, d} & \sum_{n=0}^{+\infty} x''_n\, e^{-j(\frac{\pi}{\Lambda}(2n+1)-\delta\beta)d} \\ \sum_{n=-\infty}^{-1} x_n\, e^{-j(\frac{\pi}{\Lambda}(2n+1)+\delta\beta)\, d} & \sum_{n=-\infty}^{-1} x''_n\, e^{-j(\frac{\pi}{\Lambda}(2n+1)-\delta\beta)d} \end{pmatrix} \begin{pmatrix} A_0 \\ A''_0 \end{pmatrix} \qquad (4)$$

where

$$x_n = \frac{1}{A_o} \sum_{r=1}^{2} A_{nr} \quad \textit{for the first Floquet's mode} \qquad x''_n = \frac{1}{A''_o} \sum_{r=1}^{2} A''_{nr} \quad \textit{for the second Floquet's mode} \qquad (5)$$

and, since the device will operate close to the stopband,

$$\delta\beta = \beta_o - \pi/\Lambda \tag{6}$$

where $\delta\beta$ will be small with respect to π/Λ

From Eq. (4), a matrix COM formalism is used as presented in [3, 10].

If Γ_m denotes the complex reflection coefficient of the reflectors (Fig. 4) and if the COM matrix of the cavity is expressed as

$$\begin{pmatrix} u_{13}^{+} \\ u_{13}^{-} \end{pmatrix} = \begin{pmatrix} a_r & c_r \\ b_r & d_r \end{pmatrix} \begin{pmatrix} u_{14}^{+} \\ u_{14}^{-} \end{pmatrix} \tag{7}$$

The resonant condition takes finally the following form

$$1 = \Gamma_m \frac{b_r + d_r \Gamma_m}{a_r + c_r \Gamma_m} \tag{8}$$

Comparison of the model with already published data

The previous model has been compared to experimental data already published by Avramov *et al.* [7, 8] for the resonant frequency of STW resonators as a function of metal thickness (Fig. 5). In these results, no fitting was made, theoretical predictions directly come out from material constants, metal thickness and geometrical dimensions. It can be seen that the agreement is very satisfactory.

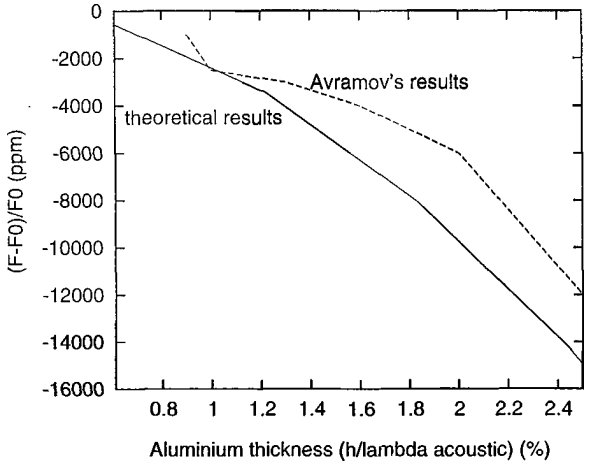


Fig. 5 *Comparison of theoretical predictions (this work) with experimental data [7, 8] for the influence of metal thickness (a) and finger-to-gap ration (b) on the resonant frequency of STW resonators*

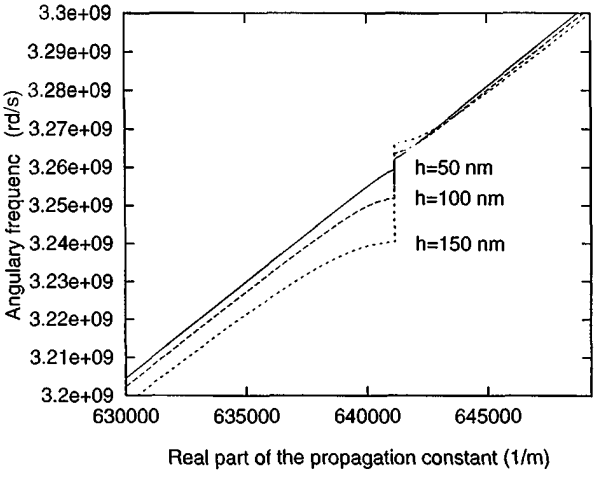


Fig. 2 : *Dispersion curves (passband) for STW propagation on (Y+36°) quartz as a function of metal thickness*

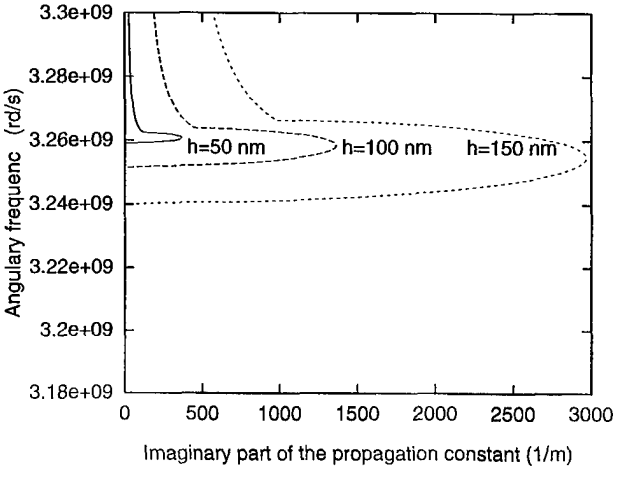


Fig. 3 : *Dispersion curves (stopband), same configurations as Fig. 2*

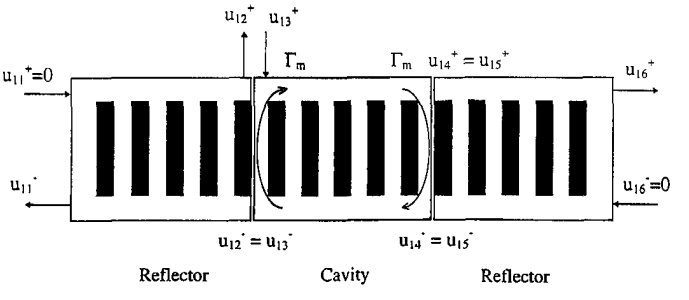


Fig. 4 : *COM approach for solving the resonant frequency of a 3-grating STW structure*

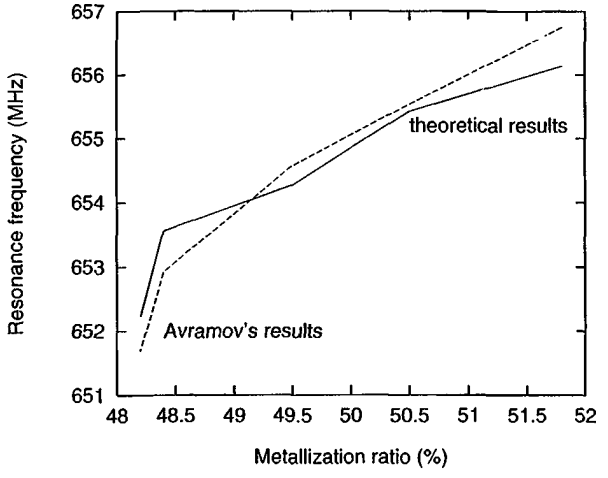


Fig. 5 b
Comparison of theoretical predictions (this work) with experimental data [7, 8] for the influence of metal thickness (a) and finger-to-gap ration (b) on the resonant frequency of STW resonators

Design, fabrication and test of 3 GHz STW resonators

Design

Two different designs were used (Fig. 6 and Tables I-II). Type # 1 is based on a design presented by Denissenko and Avramov [1] and was chosen here to work in a quasi synchronous mode with a small gap at the end of the cavity. Here this gap was adjusted to keep the periodicity of the whole device to 0.8 µm.
Type # 2 corresponds to a design where the cavity operates in the passband [6] at the resonant frequency, thus the periodicity inside the cavity is slightly higher (here : 0.85 µm) than the periodicity in the reflectors (here : 0.82 µm). An integer number of strips is kept within the cavity and a fine tuning is done by changing the number of strips to ensure that the resonance takes place at the lower edge of the mirror stopband [10, 3].

Fabrication

Devices were fabricated at LMM-CNRS (Bagneux, France) by direct e-beam photolithography. Four resonators were fabricated on a 10×20 mm quartz plate (AT-cut, Y+36°) (Fig. 7). Table III summarizes the fabrication procedure.

Most of the devices have been built using type # 1 design with various parameters
- metal thickness range was from 250 Å to 600 Å
- finger to gap data were taken as 0.6/0.2 ; 0.5/0.3 and 0.4/0.4 (in micrometers).

Two kinds of cavity lengths were taken : a long cavity with 130 strips length (plus gap), and a short cavity with 12 strips length (plus gap).

Table I : Design parameters and results for type #1

Metal thickness	500 Å
Number of strips in the cavity (Nc)	12
Number of strips in the reflector (Nr)	800
Number of pairs of electrodes by I.D.T (Nf)	70
Aperture (A)	65 µm
Gap1 (standard gap)	0.3 µm
Gap2 (adjusted gap)	1.1 µm
Periodicity in the cavity	0.8 µm
Periodicity in the reflector	0.8 µm
Width of a strip in the cavity	0.5 µm
Width of a strip in the reflector	0.5 µm
Resonant frequency	3.102 GHz
Quality factor (loaded)	1680
Quality factor (unloaded)	2420
Insertion loss	-10.3 dB

Table II : Design parameters and results for type #2

Metal thickness	400 Å
Number of strips in the cavity (Nc)	26
Number of strips in the reflector (Nr)	1000
Number of pairs of electrodes by I.D.T (Nf)	70
Aperture (A)	150 µm
Gap1 (standard gap)	0.3 µm
Gap2 (standard gap)	0.3 µm
Periodicity in the cavity	0.82 µm
Periodicity in the reflector	0.85 µm
Width of a strip in the cavity	0.33 µm
Width of a strip in the reflector	0.34 µm
Resonant frequency	2.966 GHz
Quality factor (loaded)	1280
Quality factor (unloaded)	1920
Insertion loss	-9.5 dB

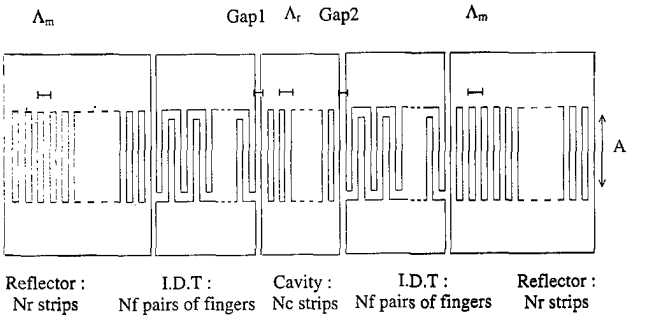


Fig. 6 : *Design parameters for 3 GHz STW resonators on (Y+36°) quartz*

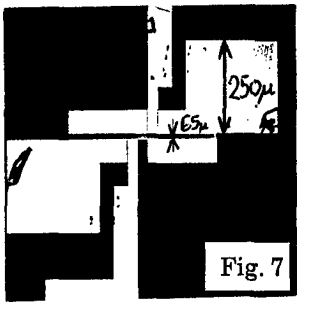


Fig. 7

Table III
Fabrication parameters by e-beam lithography

Machine type	JEOL JB × SD2U
Quartz sample size	10 × 20 mm²
Photoresist	PMMA, 150 nm thick
Beam current	100 nA / spotsize 0.3 µm (connection pads) 1 nA / spotsize 0.1 µm (electrode fingers)
Fabrication time	1 H 30 per plate (4 devices)
Line widths	0.3 to 0.5 µm
Metal layer	25 to 50 nm / lift-off

Experimental results

Devices were measured with high frequency (50 Ω) probes (picoprobe # 40A-SG-250P) on a network analyzer (HP8720A).

Best results for type # 1 were obtained with a short cavity length (Fig. 6, Table I).

Figure 8 presents network analyzer response curve of the device and Table I summarizes all the characteristics of the resonator.

Best results for type # 2 design are shown on Fig. 8b and Table II summarizes the characteristics of this second kind of resonator.

The measured resonant frequencies were compared to theoretical values predicted either from a nonpiezoelectric model of STW propagation [3] or to the prediction of the present improved model taking piezoelectricity into account [5]. Figure 9 summarizes the comparison of both models with experimental data.

Experimental values for the Q-factor and insertion losses of type # 1 resonators are reported on Fig. 10.

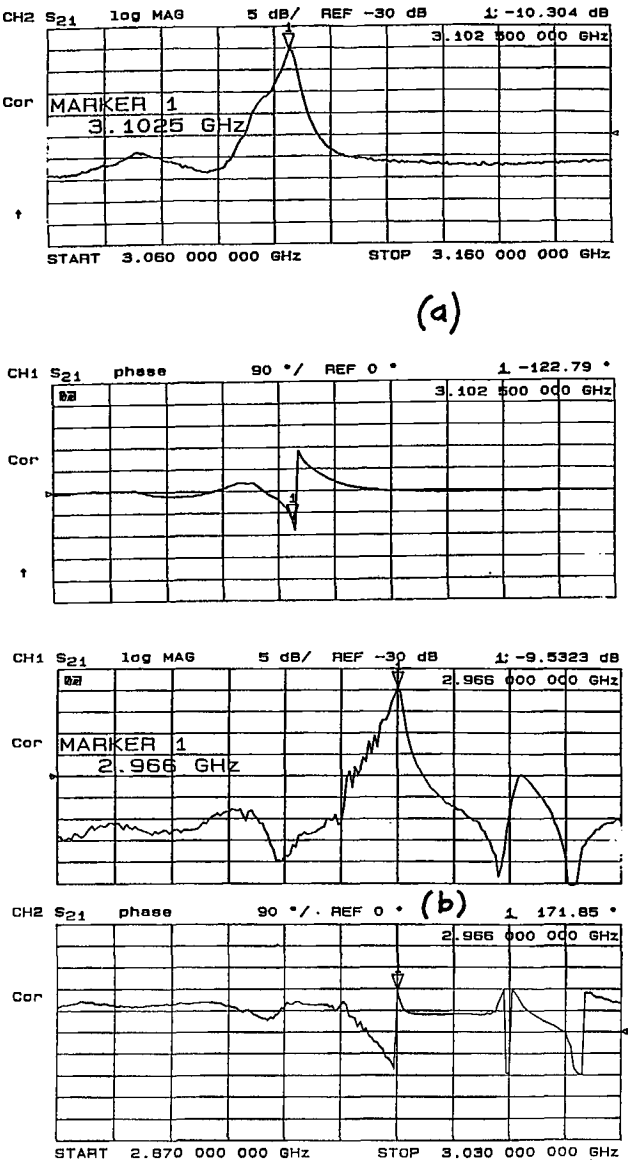


Fig. 8 :
Best results obtained for type # 1 resonators (a) and type # 2 resonators (b)

Discussion

Best values for the unloaded Q-factor were obtained with type # 1 (quasi synchronous) : Q_u = 2 420 (Fig. 8a), whereas the best value for insertion losses was obtained with type # 2 with $I.L.$ = −9.5 dB (Fig. 8b). Type # 2 exhibits additional ripples on the left-hand side of the main resonant peak. These ripples do not exist in type # 1 and may be associated with the change of periodicity inside the cavity of type # 2. However, the characteristic length $L = V/2\Delta F$ (where V is the velocity of the wave and ΔF the periodicity of the spectrum) associated to the periodic ripples is here equal to 1.3 mm, i.e. the length of the resonator. Then, it is likely that parasitic reflections occur at the end of the reflectors in type # 2, these reflections being smoothed and in the # 1 design.

For both kinds of design, best results were obtained for a finger to gap ratio higher than .5, namely 62 %. This fact, already known for high frequency Rayleigh wave resonators [11] confirms that a similar behavior is observed for STW, i.e. increased resonator performances at higher finger to gap ratios.

As for the behavior of the Q-factor and insertion losses versus metal thickness are concerned, it is interesting to compare our results (Fig. 10) to already published data by Avramov et al [7, 8]. The behavior versus metal thickness of the curves is exactly the same, with the coïncidence of a minimum of Q with a maximum of insertion loss.

Conversely, a maximum of Q-factor may be found at higher metal thickness, and then the insertion losses are best. At 620 and 650 MHz, Avramov's results show on optimum relative thickness h/Λ of 1.6-1.7 % whereas at 3 GHz the optimum is found at 2.5 % for the same kind (quasi synchronous) design.

It appears that at higher frequencies, the relative metal thickness should be increased over 2 % to ensure the best performances. In terms of the factor of merit $Q_u.f$, Table IV summarizes the results obtained here and compared with already published data.

Table IV
Comparison of different factors of merit

Reference	Operating frequency	Qu	Qu.f
Bagwell & Bray [6]	1,72 GHz	5 600	9.6×10^{12}
Denissenko & Avramov [1]	2 GHz	3 800	7.6×10^{12}
This work type # 1	3 GHz	2 400	7.5×10^{12}
This work type # 2	3 GHz	1 700	5.3×10^{12}

Conclusion

STW resonators have been successfully designed, fabricated on (Y+36°) quartz at 3 GHz by e-beam lithography. Results have demonstrated a factor of merit $Q_u.f = 7.5 \times 10^{12}$ comparable to already published data. The present technology available at LMM-CNRS will be able to achieve smaller linewidths and higher operating frequencies. To improve the present performances, a more detailed analysis is required to find what are the limits of STW resonators at high frequencies i.e. the influence of the design, ohmic losses in the fingers and intrinsic material losses of quartz. For the moment, the reported $Q.f$ product for STW resonators is in the range of 5 to 10×10^{12}, whereas bulk wave resonators can exceed this value (e.g. : 13.5×10^{12} for 10 MHz AT-cut resonators [12], and intrinsic limits might be in the range of $20\text{-}30\times10^{12}$ according to the values of the motional time constant τ_1 reported for AT and BT quartz [13].

since the STW mode on (Y+36°) quartz has in fact acoustic properties close to a BT-cut bulk mode, one may expect improvements in the $Q.f$ product if other sources of losses (imperfections of the design and fabrication, ohmic losses) can be mastered properly.

Acknowledgements

This work was supported by CNRS under the form of a visiting scientist position in France for 3 months. The authors wish to express their gratitude to Mme H. Launois, director of LMM/CNRS and all the people at LMM, LPMO (France) and ISP (Novossibirsk-Russia) who helped this collaboration to be successful.

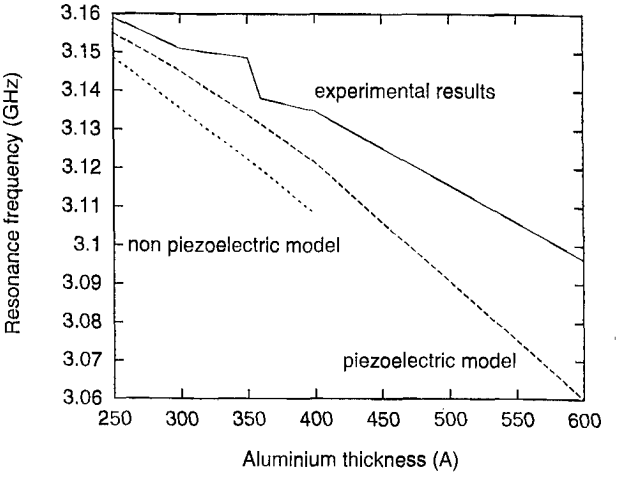


Fig. 9 : *Comparison of predicted resonant frequencies according to a nonpiezoelectric model [3], the present model [5] with measured data at 3 GHz*

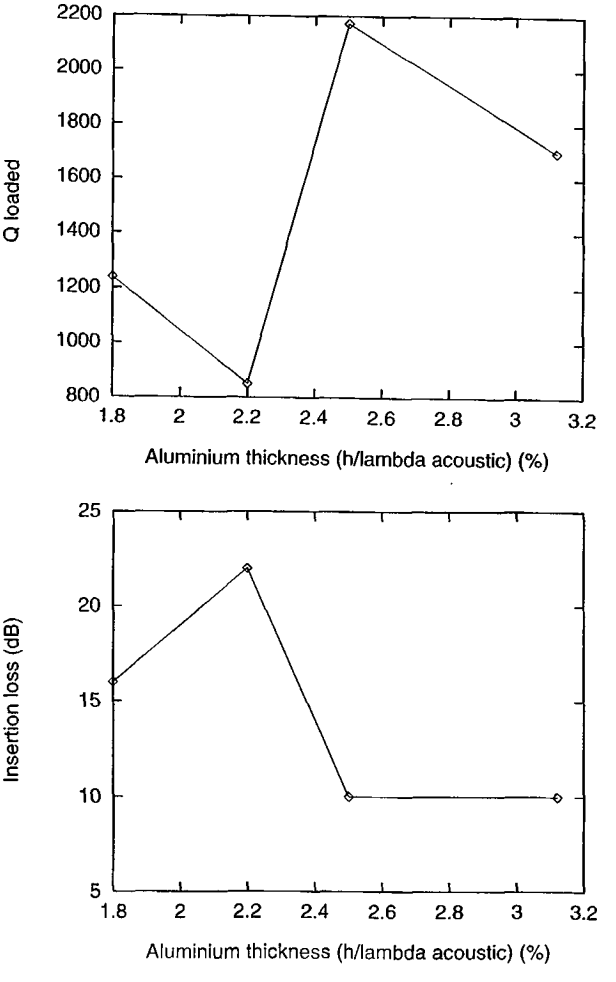


Fig. 10 :
Experimental results for the loaded Q-factor (a) and insertion losses (b) for type # 1 resonators

References

[1] S. Denissenko, I.D. Avramov, "Design, fabrication and performance of 2 GHz surface transverse wave resonators", Proc. 1993 I.F.C.S., IEEE cat. 93CH3244-1, pp. 645-649.

[2] I.D. Avramov, F.L. Walls, T.E. Parker, G.K. Montress, "Surface transverse wave oscillators with extremely low thermal noise floors", Proc. 1994 I.F.C.S., Boston (USA), june 1-3, IEEE cat. 94CH2446-2, pp.379-394.

[3] E. Bigler and E. Gavignet, B.A. Auld, E. Ritz and E. Sang, "Surface transverse wave (STW) quartz resonators in the GHz range", Proc. 6th European Frequency and Time Forum, Noordwijk, Netherlands, March 1992, pp. 219-222.

[4] T. Thorvaldsson, V.P. Plessky, S. Muckenhirm, M. Joray, "GHz range STW resonators and narrow band filters", Proc. 1994 Ultrasonics Symp., Cannes, France, nov.. 1994.

[5] E. Gavignet, S. Ballandras, E. Bigler, "Theoretical analysis of surface transverse waves propagating on a piezoelectric substrate under shallow-groove or thin metal strip grating", to be published in J. Appl. Phys.

[6] T.L. Bagwell and R.C. Bray, "Novel surface transverse wave resonators with low loss and high Q", Proc. 1987 Ultrasonics Symp., IEEE cat. n° 87CH2492-7, pp. 319-324.

[7] I.D. Avramov, "High Q metal strip SSBW resonators using a SAW design", IEEE Trans. Ultr. Ferro. Freq. Contr., vol. UFFC-37, n° 6, pp. 530-534 (nov. 1990).

[8] I.D. Avramov, D.F. Thompson, "Influence of the metallisation on the electrical performance of STW two-port resonators", Proc. 1993 Acoustoelectronics Conference, L. Spassov *et al* editors, sept. 19-25 1993, Varna (Bulgaria).

[9] V.M. Bright and W.D. Hunt, "Analysis of Bleustein-Gulyaev wave propagation under thin periodic metal electrodes", J. Appl. Phys., vol. 70, n° 2, pp. 594-602 (1991).

[10] E. Bigler, B.A. Auld, E. Ritz and E. Sang, "An analysis of the influence of design parameters on the resonant frequency and Q-factor of surface transverse wave (STW) resonators", Proc. 45th Ann. Freq. Cont. Symp., Los Angeles, june 1991.

[11] L.L. Pendergras, L.G. Studebaker, IEEE Trans. Ultrasonics, Ferro. Freq. Control, vol. UFFC-35, n° 3 (1988), pp. 372-379.

[12] M. Mourey, R.J. Besson, "A space oscillator with cylindrical oven and symmetry", Proc. 45th A.F.C.S., IEEE cat. # 91CH2965-2, pp. 431-441.

[13] A. Ballato, J.J. Gualtieri, "Piezoelectric resonator materials", Proc. 48th I.F.C.S., IEEE cat. # 94CH3446-2, pp. 22-34.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

AN OSCILLATOR DESIGN USING LOW g-SENSITIVITY, LOW PHASE NOISE STW DEVICES

MARK S. CAVIN AND RODOLFO C. ALMAR

Sawtek Incorporated
P. O. Box 609501, Orlando, Florida 32860-9501

Abstract

This paper will focus upon results of integrating Surface Transverse Wave (STW) technology into an oscillator product line. Phase noise of <-100 dBc/Hz at 100 Hz, noise floor of <-180 dBc/Hz, and g-sensitivity of $<5 \times 10^{-10}$ per g have been achieved at 1 GHz using STW technology. Select units have exhibited g-sensitivities below 1×10^{-10} per g. These 1/f and g-sensitivity results represent state of the art for 1 GHz STW oscillators. The above low phase noise and low g-sensitivity ($<5 \times 10^{-10}$ per g) have been achieved in a 1.5" x 1" x .2" package. In addition, this g-sensitivity performance is attained in an industry-standard, 14-pin dual in-line package (DIP) measuring .87" x .5" x .24" (See Figure 1). Bare hybrid oscillator products with hermetically-sealed resonators have been developed at a size of .715" x .380" x .110" for subsequent hermetic module integration. Design techniques have resulted in production of standard oscillators from 300 MHz to nearly 1200 MHz and over temperature ranges from -55 degrees Celsius to +110 degrees Celsius using the exact same hybrid substrate design. The design of this hybrid circuit will be discussed.

STW technology has played a key role in developing oscillators at the upper end of the nearly two octave range of frequency application. The design of 1 GHz range STW resonators used in the product line will be discussed. STW resonators have been designed using combined-mode and in-line acoustically cascaded techniques. Performance parameters for both techniques will be discussed.

Phase noise, g-sensitivity variation with cut angle, and electrical parameter repeatability statistics on multiple oscillators in the 1 GHz range will be presented.

Introduction

Sawtek has produced a line of standard catalog oscillators since 1991. These are chip and wire hybrid circuits using alumina substrates and spiral inductors. These oscillators are of the 50 ohm interface loop type, which allows for characterization and optimization of functional modules. Loop oscillators have been the subject of much work over the last decade, as evidenced by numerous publications on the subject [2],[3],[4],[5],[6].

Key components of such oscillators will be summarized and tradeoffs to implement such a design in the size of a 14-pin DIP will be discussed. This includes careful allocation of substrate area for inductance used in resonator matching, phase shift, power splitting, and RF choke functions as well as substrate area allocated to amplifier module, resonator surface mount package, tuning capacitors, and attenuators. A functional module loss/gain budget for a very common application will be detailed to illustrate some of the tradeoffs made in the design. The mechanical structure of the resulting standard oscillator as applicable to the vibration performance will be described.

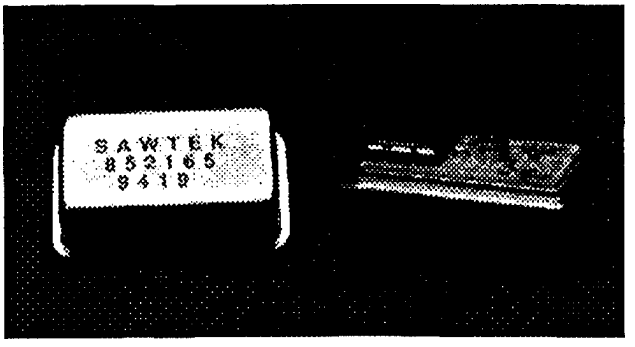


Figure 1. Sawtek Standard Fixed Frequency Oscillator

0-7803-2500-1/95 $4.00 © 1995 IEEE

STW Resonator Design

The 1 GHz STW frequency control elements used in this study are Combined-Mode Resonator Filters (CMRF) and In-Line Acoustically Coupled Resonator Filters (RFI). The CMRF topology is similar to that of a standard two-port, one-pole SAW resonator filter. Basically, it consists of a pair of reflector gratings that form a resonant cavity into which two interdigital transducers (IDT) are placed to generate and detect the waves.

Unlike the one-pole SAW resonator, the CMRF uses two STW modes and a non-trapped Surface Skimming Bulk Wave (SSBW) mode to provide bandwidths larger than obtainable with a one-pole resonator. Their theory of operation has been described previously in [1] and [4]. The larger bandwidths obtained with this type of structure is used to absorb frequency set errors and frequency drifts due to temperature and aging while maintaining a reasonably constant group delay over the resonator passband.

The STW RFI design includes a coupling grating between the transducers as is the case in conventional in-line acoustically coupled SAW resonators. Although the residual phase noise for both type of devices has been measured to be very similar, about -140 dBc/Hz at 1 Hz, larger values of group delay have been obtained with the RFI structure. This is an important parameter when close in oscillator phase noise is critical. In these cases, an RFI design is preferred.

Table 1 lists some average measurements taken on STW resonators. This data was taken in fixtures used to match the STW resonators to 50 ohms. The bandwidths are measured about the point of minimum insertion loss. The low loss of these devices is key for low noise floor designs because this allows the use of amplifiers with as low as 10 dB of gain. Higher loss frequency control elements would necessitate the use of higher gain amplifiers, thus degrading the noise floor. This, of course, assumes all other circuit parameters are equivalent.

Table 1. STW Resonator Data

Frequency	Loss	Delay	1 dB BW	3 dB BW	Topology	Lot Sample
MHz	dB	ns	kHz	kHz		
960	3.5	326	900	1258	CMRF	102
960	3.7	318	896	1282	CMRF	233
1030	3.6	304	981	1458	CMRF	210
1090	3.8	298	1170	1642	CMRF	132
1160	3.7	536	346	736	RFI	186
1160	3.5	547	349	694	RFI	226

STW Fixed Frequency Oscillator Design

As discussed above, 50 ohm loop oscillators have many advantages over negative resistance designs. One key advantage is less stringent requirements for two port device characterization of the resonator and amplifier to evaluate stability criteria. Functional modules are matched to 50 ohms to minimize reflections at module interfaces. Figure 2 shows a schematic of a fixed frequency oscillator (FFO). Similar oscillators have been the subject of extensive published work as illustrated in [2]. This circuit topology has been integrated into a 14-pin DIP hybrid oscillator package and proven to be effective over the frequency range of 300 MHz to nearly 1200 MHz.

The resonator is encapsulated in its own hermetically-sealed surface mount package (SMP). This allows for preselection of resonators and protection from contaminants. Figure 3 shows the hybrid oscillator mechanical profile from the oscillator base to the quartz resonator die. This mechanical structure has been shown to provide g-sensitivities less than 5×10^{-10} per g for STW oscillators at 1 GHz. This work will be presented in a subsequent section of this paper.

The switch shown in the resonator area illustrates a polarity reversal of one of the transducers in the resonator. This alleviates the need for a full 360 degree phase shift capability by the loop electronics, thus saving extensive substrate area.

Provisions exist for trimming excess loop gain and output power through the use of chip attenuators. Loop trimming can improve phase noise and harmonic levels while output power trimming increases load isolation, therefore decreasing load pull.

The switch at the input of the amplifier illustrates the capability of performing an open loop test to measure the loop insertion phase and excess loop gain. Although not a necessity in production, this is an invaluable tool during the developmental stage of any prototype oscillator.

The low-pass topology of the Wilkinson divider obviously aids in harmonic suppression, making it an effective option for the power divider. It can be designed with an unequal splitting ratio when less loop loss or more coupled power is required. Isolation can still be achieved but VSWR of the splitter ports will degrade.

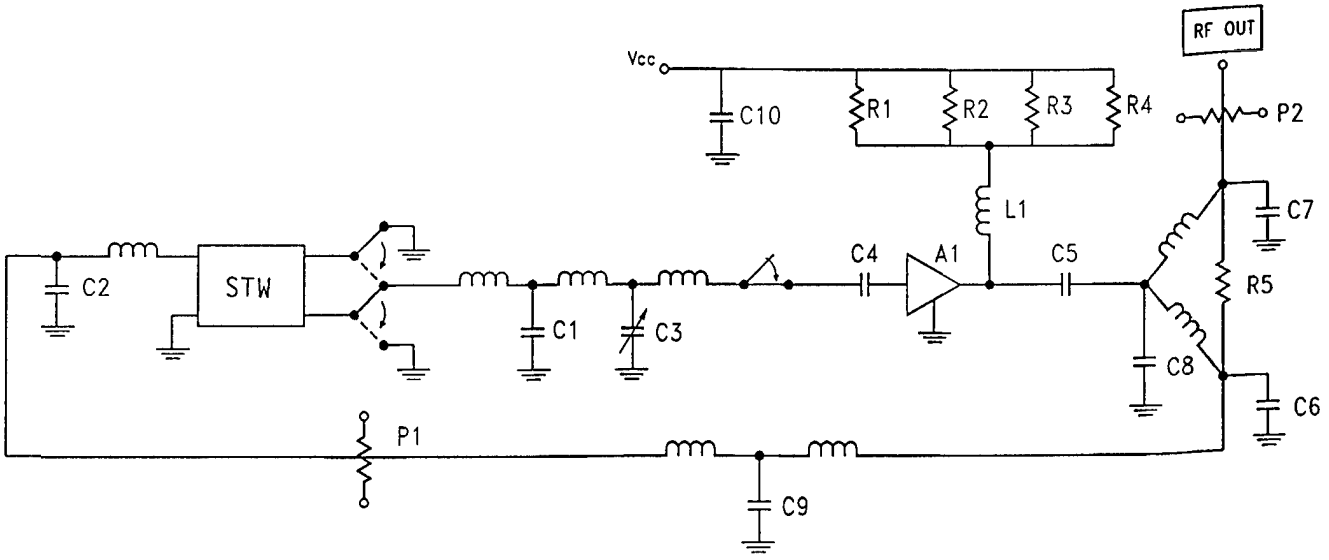


Figure 2. Schematic of Fixed Frequency Oscillator

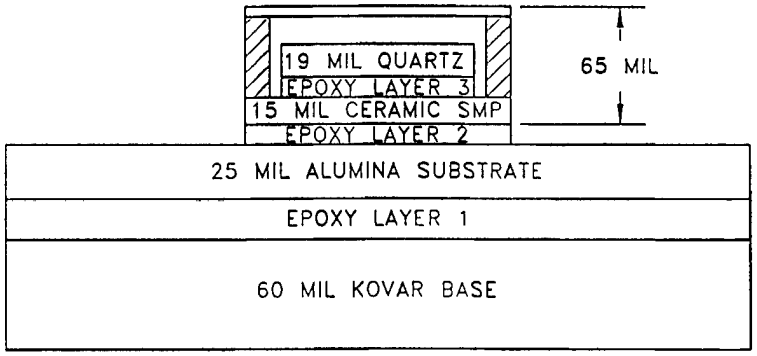


Figure 3. Hybrid Oscillator Mechanical Profile

Allocation of substrate area for all the required functions is perhaps the most challenging aspect of implementing the circuit of Figure 2 in a 14-pin DIP. At low frequencies, larger inductors are required for resonator matching, RF choke, Wilkinson inductors, and phase shift inductors. RF choke inductances, power splitter inductances, and phase shift inductances can be readily calculated at the lowest frequency of potential circuit operation.

Spiral inductors were characterized and modeled to determine optimum number of turns, metal thickness, line width, and series resonant frequency. The result is the oscillator shown in Figure 1.

Table 2 shows a loss/gain budget for a typical hybrid oscillator shown in Figure 1. Figure 4 shows a test fixture used to test up to five unpackaged oscillators. This allows

Table 2. STW FFO Loss/Gain Budget at 1 GHz

Functional Module	*Loss/Gain, dB*
STW Device	-4.5
Amplifier	11
Power Splitter	-3.5
Fixed Phase Shift 1	-0.5
Fixed Phase Shift 2	-0.5
Excess Loop Gain	2

for denser integration into hermetic subsystem modules. Novel fixturing and test techniques have been developed which make the bare oscillator hybrids extremely producible. Table 3 shows some statistics for one hundred oscillators of identical design.

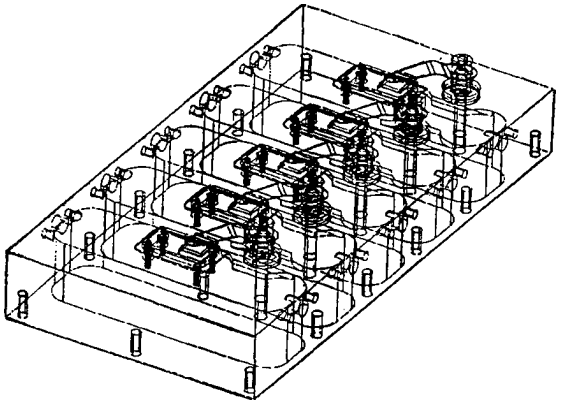


Figure 4. Bare Substrate FFO test fixture

Table 3. STW Oscillator Statistics

	P25,dBm	I25, mA	P110, dBm	I110, mA	P-55, dBm	I-55, mA
Avg	-6.8	21.4	-7.5	23.2	-7.5	19.8
Std	0.4	0.5	1.6	0.4	0.6	0.4

	2fo, dBc	3fo, dBc
Avg	28.0	55.7
Std	1.4	2.1

g-Sensitivity

The increasing demand on the performance of military systems is leading to very stringent requirements for oscillators. Some of these systems operate in very hostile conditions, such as high vibration environments; examples of which are missiles, helicopters, and high performance aircraft.

Under these conditions the dynamic close in phase noise of an oscillator is ultimately determined by the vibration environment and the oscillator g-sensitivity. When the oscillator is subjected to vibration, sidebands appear at the oscillator output as a consequence of the modulation of the oscillator frequency by the frequency of vibration. The level of the sidebands, which appear as phase noise degradation, is related to the level of vibration to which the oscillator is exposed.

It has been observed that the frequency change of the oscillator due to vibration is proportional to the magnitude of the acceleration and is also dependent on the direction of the acceleration. Therefore, the scalar change in frequency can be written as a dot product of two vectors [7], namely, the acceleration $\boldsymbol{a}$ and the g-sensitivity vector Γ as follows :

$$f(\boldsymbol{a}) = fo\,(1+ \Gamma \boldsymbol{a}) \qquad (1)$$

where fo is the frequency of oscillation under no vibration.

For a small modulation index and under sinusoidal vibration, the acceleration sensitivity vector can be expressed as a function of the vibration-induced sidebands as follows :

$$\Gamma_i = [(2fv)/(a_i fo)]10^{Lv/20} \qquad (2)$$

Where i denotes the ith component of the acceleration sensitivity vector, a_i is the ith component of the acceleration, fv is the frequency of vibration and Lv is the sideband to carrier power ratio.

Equation (2) provides a means for determining the vibration sensitivity of an oscillator by measuring the ratio of the power in the sidebands to the power in the carrier when the oscillator is subjected to a sinusoidal vibration.

g-Sensitivity Measurements

The experiments reported here were conducted using the 1 GHz FFO shown in Figure 1. The STW resonators were housed in ceramic surface mount packages (SMPs), which were hermetically sealed and mounted on the oscillator substrate. The mounting procedure and choice of mounting materials used in the attachment of the die to the SMP and the SMP to the oscillator substrate is of paramount importance for obtaining low g-sensitivity.

All the measurements reported here were conducted using sinusoidal vibration with typical peak acceleration values between 1 g and 4 g. The acceleration sensitivity was then obtained using equation (2).

Most of the measurements were performed on the normal component of the acceleration. Preliminary data on the in-plane g-sensitivity seems to indicate that the normal component of the vector is two to three times larger than the in-plane components.

Table 4 lists data measured in a direction perpendicular to the propagation plane. Typical values are in the 2 to 3 x 10^{-10} per g range, with some devices showing g-sensitivity as low as 0.65 x 10^{-11} per g at certain vibration frequencies.

Table 4. Γ of 1 GHz STW FFO in 10^{-10} per g

Frequency of Vibration f_v *[Hz]*	*SN1*	*SN3*	*SN4*	*SN5*	*SN7*
500	2.2	2.8	5.6	2.5	1.4
800	0.63	1.6	7.1	1.3	2.0
1000	1.3	1.6	3.2	1.1	1.0
1300	0.65	2.1	3.3	1.3	1.3
1500	1.7	2.1	3.8	1.5	1.5
1800	2.8	2.0	3.2	1.6	1.8
2000	3.5	3.2	3.2	1.6	2.0

The above measurements were performed on oscillators that use CMRF STW devices fabricated on 36° Y-rotated quartz as frequency control elements.

The SMP package that houses the resonator is mounted on the oscillator ceramic substrate using a combination of conductive and non-conductive materials. Figures 5 and 6 show two different manners of mounting the SMP onto the ceramic substrate.

The one described in Figure 5, which will be referred to as non-full mount, uses only conductive material to make the appropriate electrical connections between the resonator and the oscillator electronics. The material is placed on the substrate using the dot arrangement shown in the figure to ensure that no shorting between hot and ground electrodes is introduced.

On the other hand, Figure 6 shows a full mount arrangement. In this case, both conductive and non-conductive materials are used.

The non-conductive material serves as a filler to promote a more uniform stress distribution under the SMP.

The g-sensitivity response of oscillators fabricated using non-full mount and full mount techniques is shown in Figure 7, where a significant performance difference can be observed. This particular example was specifically chosen to emphasize the importance of the mounting technique. Therefore, care should be exercised in the choice of material and/or mounting arrangements.

The behavior of g-sensitivity with cut angle had not, to the authors' knowledge, been previously studied, so an experiment was undertaken in which the 1 GHz resonators were built on different cuts of Y-rotated quartz ranging from 32° to 46°. No attempt was made to optimize the design for each cut by changing the aspect ratio of the die, which was approximately 2:1 for the CMRF devices and 2.4:1 for the RFI device. Aspect ratio is here defined as the ratio between the length in the propagation direction to the width of the die. The same design and metal thickness was used for all the different cuts. All the wafers were 3" in diameter and .48 mm thick with the exception of the 42.75° wafer which was .56 mm thick. The mounting material and mounting arrangements were maintained the same for all cases.

The measured g-sensitivity for oscillators that use CMRF devices as frequency control elements fabricated on 32°, 36° and 46° Y-rotated quartz is shown as a function of the vibration frequency in Figures 8, 9, and 10.

Within the frequency range measured, the g-sensitivity is rather independent of frequency. These results clearly indicate the dependance of g-sensitivity with cut angle for a given design.

Experimental results for all cuts at a vibration frequency of 1300 Hz are shown in Figure 11. It is very interesting to note that the g-sensitivity goes through a minimum for a cut angle of 36.5°.

1 GHz STW VCO Design

The Standard FFO discussed above utilized silicon MMIC die due to size and cost constraints. As a result the 1/f phase noise was dominated by the residual phase noise contribution of the MMIC. This section will summarize work on the addition of a low residual phase noise amplifier to a 1 GHz STW hybrid VCO.

Figure 12 shows Sawtek's Standard Voltage Controlled Oscillator. It contains a packaged MMIC amplifier module which can be seen in the mid to upper

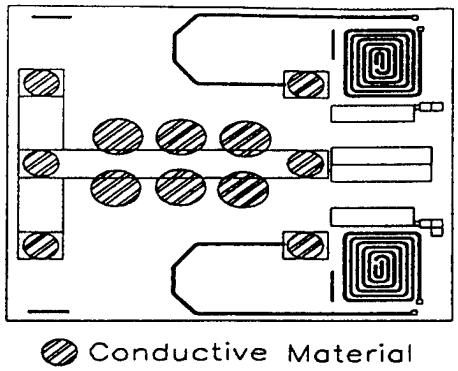


Figure 5. SMP Non-Full Mount

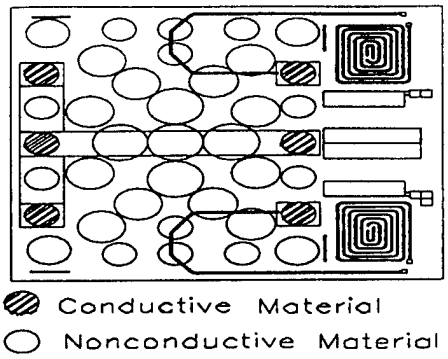


Figure 6. SMP Full Mount

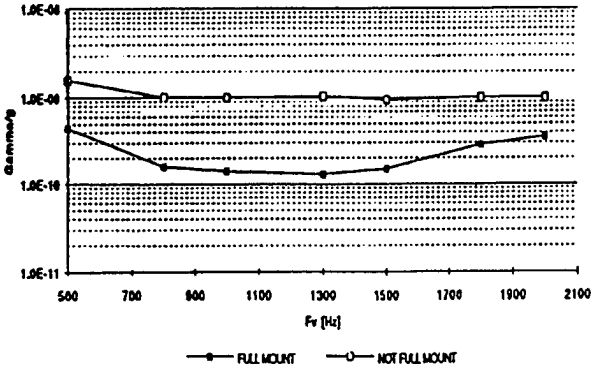


Figure 7. g-Sensitivity Dependence on Mount

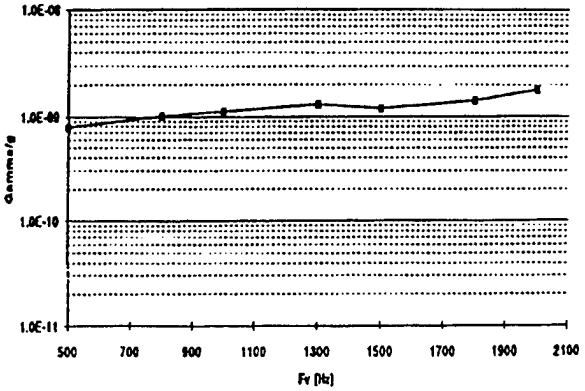


Figure 8. FFO g-Sensitivity, 32° Quartz

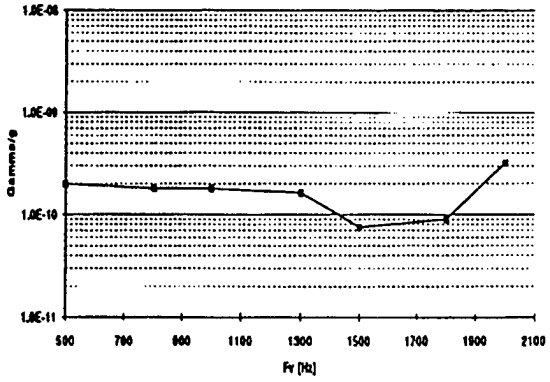


Figure 9. FFO g-Sensitivity, 36° Quartz

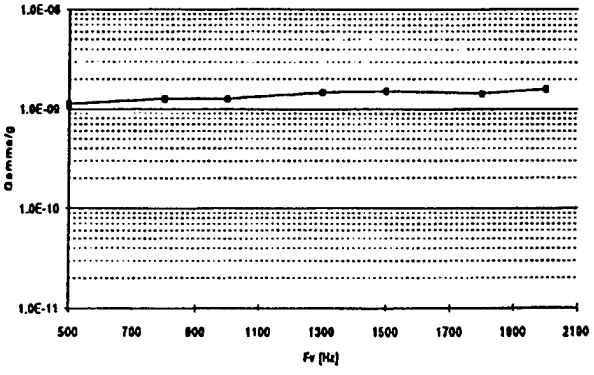


Figure 10. FFO g-Sensitivity, 46° Quartz

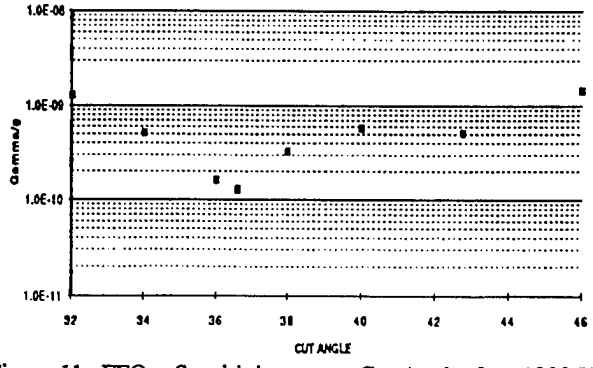


Figure 11. FFO g-Sensitivity versus Cut Angle, fv = 1300 Hz

left portion of the hybrid circuit. This amplifier module was replaced with a low residual phase noise amplifier to prototype a 1 GHz STW VCO with state-of-the-art g-sensitivity and static phase noise. The size of this package is larger due to the need for a tunable phase shifter and other options such as a voltage regulator, packaged MMIC, and ECL output option. Its dimensions are 1.5" x 1" x .2".

A STW RFI topology was chosen due to the large group delay achievable. Measurements of the 1 GHz STW hybrid module are shown in Figures 13 through 16. The group delay was approximately 1.3 to 1.4 microseconds and the insertion loss was 4.3 dB. Figure 17 shows a plot of the open loop measurement of the VCO.

Closed-loop frequency and power data was taken and the results are shown in Table 5. Phase noise data is shown in Figures 18-20 and Table 6. Included is a single oscillator test against a Hewlett Packard (HP) 8662A synthesizer (Figure 18), a two oscillator test (Figure 19), and results of calculations to correct for the HP 3048 floor and the second oscillator's contribution as done in [6] (Figure 20 and Table 6). These results indicate 1/f noise of -101 dBc/Hz at 100 Hz and a 1 Hz intercept of approximately -38 dBc/Hz. This is approximately 5 dB better than previously reported state-of-the-art 1 GHz STW 1/f data [5] for 1 Hz intercepts and approximately 8 dB better at 100 Hz offsets. In addition, this oscillator was in approximately 3.4 dB of gain compression, a sufficient amount to operate over military temperature ranges.

The subject oscillator was tested for g-sensitivity and the results are shown in Figure 21 and Table 7. The unit was tested at an acceleration of 2 g to readily show the vibration sidebands on the spectrum analyzer. Calculated g-sensitivity of less than 4 x 10^{-10} out to 2 kHz is comparable with data taken on the Standard FFO in the smaller 14-pin DIP package. Further improvement in g-sensitivity can most likely be made by full mounting the resonator as shown in Figure 6.

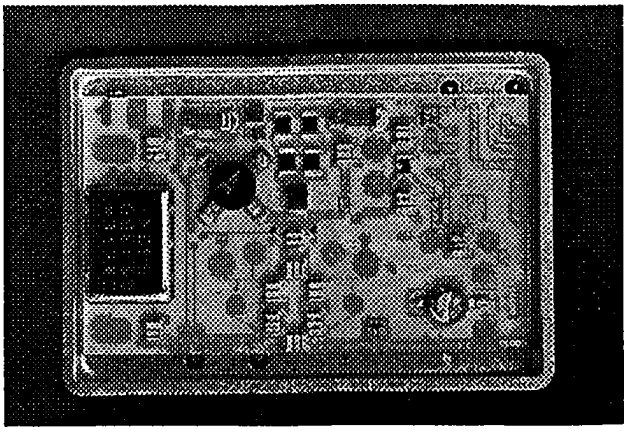


Figure 12. Sawtek Standard VCO

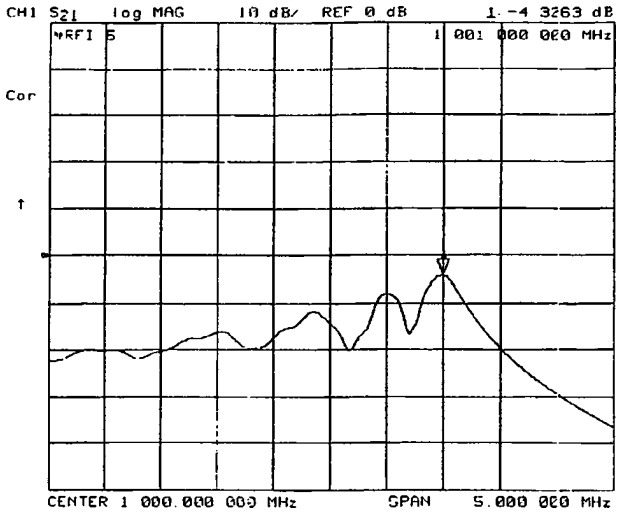


Figure 13. Insertion loss of 1 GHz STW RFI

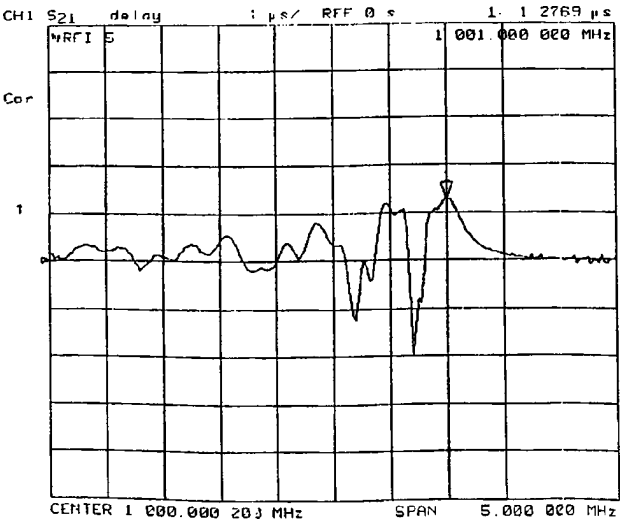


Figure 14. Group delay of 1 GHz STW RFI

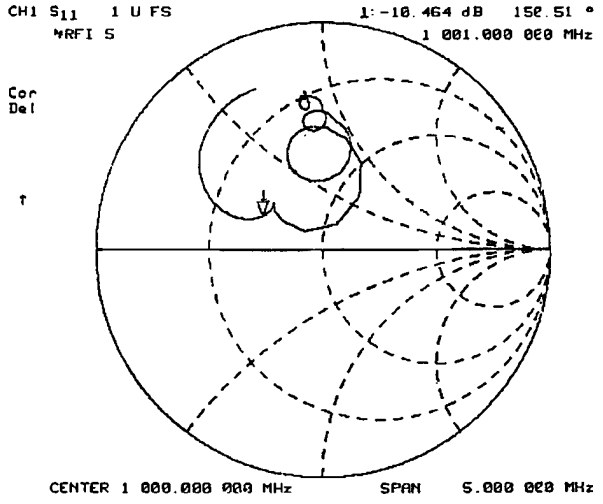


Figure 15. Input Return Loss of 1 GHz STW RFI

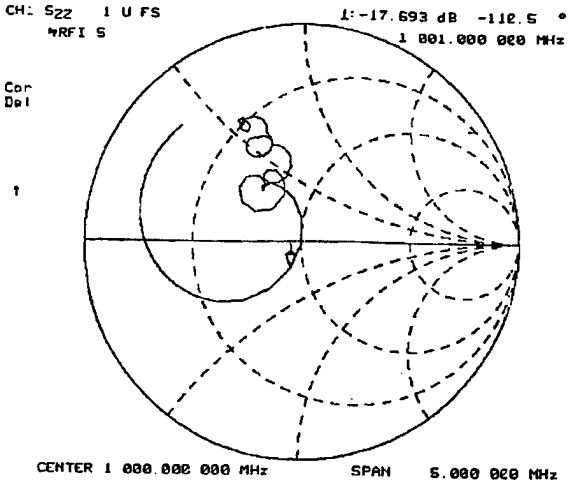


Figure 16. Output Return Loss of 1 GHz STW RFI

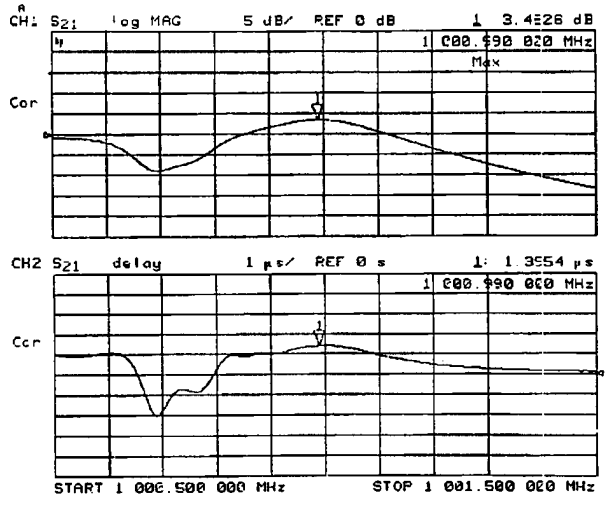


Figure 17. Open Loop Response of 1 GHz STW VCO

Table 5. 1 GHz STW VCO Closed Loop Data

	Unit 1	*Unit 1*	*Unit 2*	*Unit 2*
Tune Voltage Volts	Frequency MHz	Power dBm	Frequency MHz	Power dBm
2	1000.898	21.3	1000.885	23.2
4	1000.913	22.5	1000.908	23.2
6	1000.932	22.7	1000.929	23.1
8	1000.955	22.1	1000.947	22.9

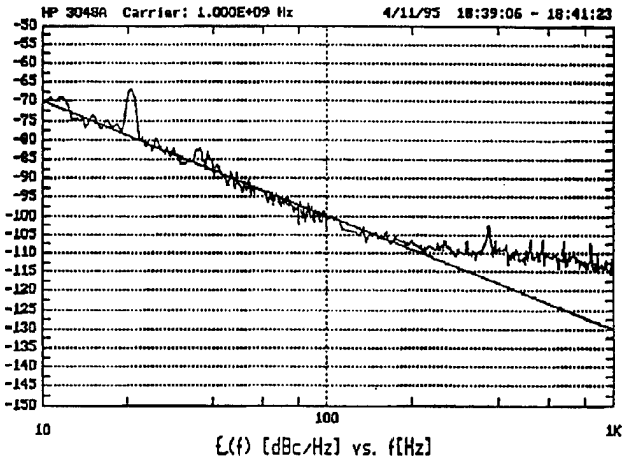


Figure 18. 1 GHz STW VCO versus HP 8662A

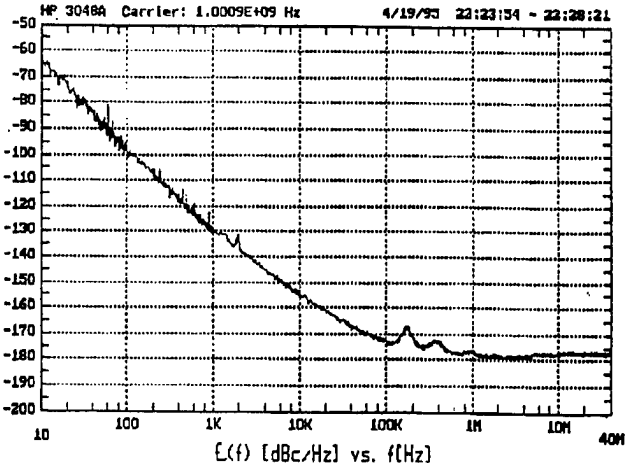
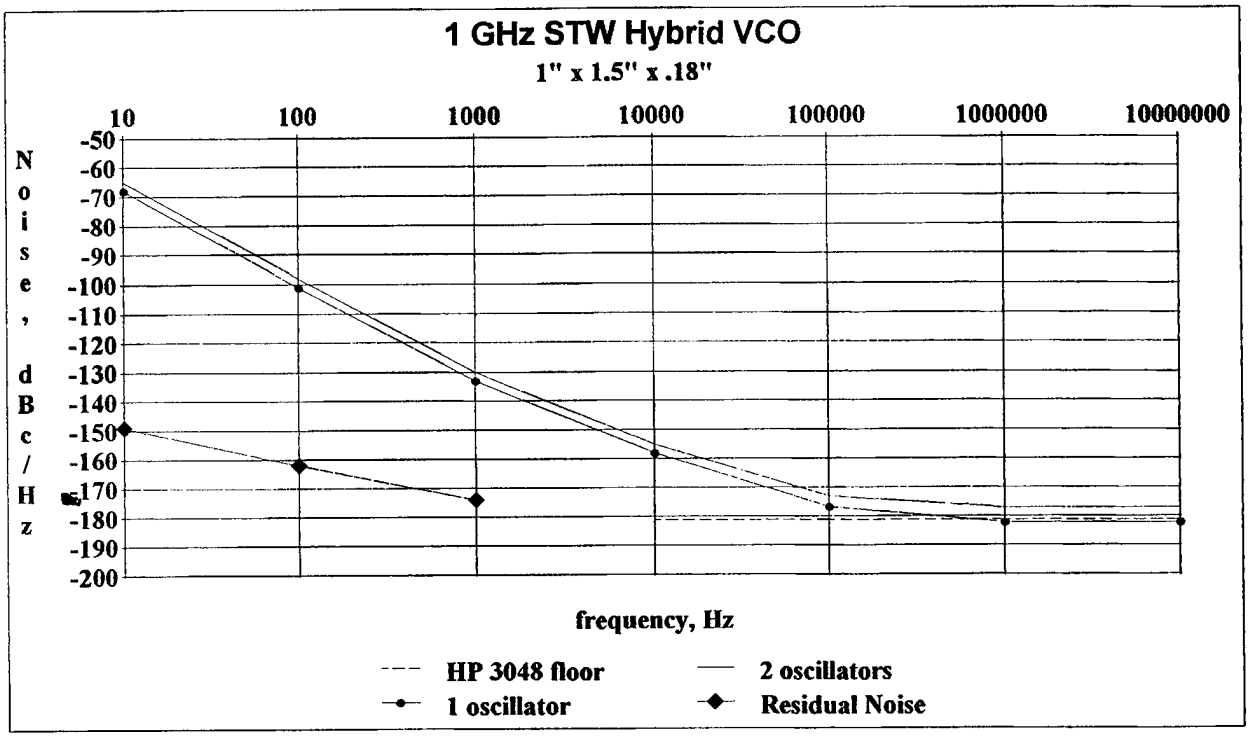


Figure 19. 1 GHz STW VCO (2 oscillator test)



Frequency, Hz	3048 Floor, dBc/Hz	2 Oscillators, dBc/Hz	1 Oscillator, dBc/Hz	Residual Noise, dBc/Hz
10		-65	-68	-149
100		-98	-101	-162
1000		-130	-133	-174
10000	-181	-155	-158	
100000	-181	-173	-177	
1000000	-181	-177	-182	
10000000	-181	-177	-182	

Figure 20 and Table 6. 1 GHz STW VCO Phase Noise

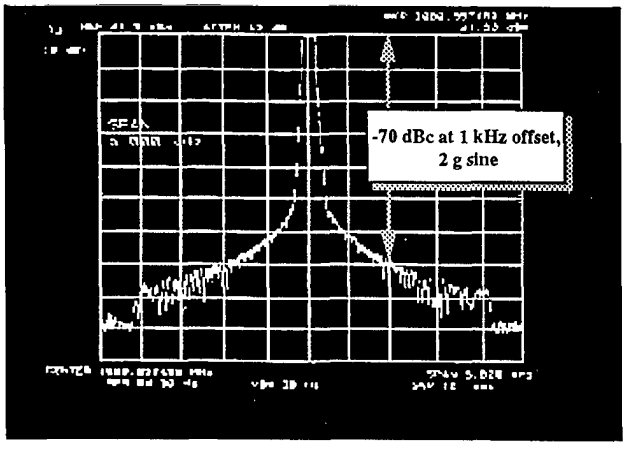


Figure 21. 1 GHz STW VCO under 2 g swept sinusoidal vibration

Table 7. g-Sensitivity of 1 GHz Hybrid STW VCO

fv, Hz	*Sideband level, dBc*	Γ
500	-62	3.97164E-10
1000	-70	3.16228E-10
1500	-75	2.66742E-10
2000	-77	2.82508E-10

Conclusions

The design of a 14-pin DIP hybrid oscillator was discussed. This same design has been used successfully over a frequency range of 300 MHz to nearly 1200 MHz. STW technology is used to achieve fundamental circuit operation at the upper end of this nearly two octave frequency range.

The use of STW technology enables the design of very low g-sensitivity oscillators using reasonably inexpensive techniques. In fact, this work reports values of g-sensitivity comparable to those obtained using significantly more sophisticated and, consequently, higher cost technologies. The same principles of sound mechanical design applied in other technologies can, of course, be applied here, thus opening the possibility of obtaining g-sensitivity values in the mid 10^{-11} range at a reasonable cost and with reasonable yields in the near future. These techniques include but are not limited to finite element analysis and die aspect ratio compensation.

Phase noise of -101 dBc/Hz at 100 Hz, Noise floor of -182 dBc/Hz, and g-sensitivity of $< 4 \times 10^{-10}$ per g were reported for a 1 GHz STW hybrid VCO in a 1.5" x 1" x .2" package. This state-of-the-art oscillator offers the system designer exceptional combined performance in areas of close in phase noise, noise floor, g-sensitivity, size, and cost.

Acknowledgments

The authors would like to thank Sawtek management for the continued support of work in the area of frequency control. Thanks go to B. Barrieau, D. Jordan, H. Patterson, C. Register, and D. Zimmerman for assistance with resonator and oscillator measurements. Thanks also to J. Mink for design of production STW resonators and J. Gore for mechanical design of bare substrate test fixture. Special thanks to Marjorie Buckley and Martha Ellis for assistance in generating figures, tables, and viewgraphs for this work.

References

1. I. Avramov, Electronics Letters, Vol. 27, No. 5, February 1991, pp. 414-415.

2. G. Montress and T. Parker, "High Stability SAW Oscillators: Design and Performance", Short Course Presented at 1991 IEEE Ultrasonics Symposium, December 8, 1991.

3. R. Almar and M. Cavin, "Low g-Sensitivity Fixed-Frequency Oscillators", Microwave Journal, February 1995.

4. R. Almar, B. Horine, and J. Andersen, "High Frequency STW Resonator Filters", 1992 Ultrasonics Symposium Proceedings, pp 51-56.

5. I. Avramov, F. Walls, T. Parker, and G. Montress, "Surface Transverse Wave Oscillators with Extremely Low Thermal Noise Floors", 1994 IEEE International Frequency Control Symposium Proceedings, pp 379-394.

6. G. Montress and T. Parker, "Design and Performance of an Extremely Low Noise Surface Acoustic Wave Oscillator", 1994 IEEE International Frequency Control Symposium Proceedings, pp 365-373.

7. R. Filler, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 35, No. 3, May 1988, pp 297-305.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

AN ANALYSIS OF THE NORMAL ACCELERATION SENSITIVITY OF ROTATED Y-CUT QUARTZ SURFACE TRANSVERSE WAVE RESONATORS SIMPLY SUPPORTED ALONG RECTANGULAR EDGES

John A. Kosinski, James T. Stewart, Arthur Ballato, and Rodolfo Almar*

US Army Research Laboratory, ATTN: AMSRL-PS-ED, Fort Monmouth, New Jersey 07703-5601

*SAWTEK, Inc., P.O. Box 609501, Orlando, Florida 32860-9501

Abstract: An analysis of the normal acceleration sensitivity of rotated Y-cut quartz STW resonators simply supported along rectangular edges is presented. The analysis is based upon the perturbation integral formulation of Tiersten, and makes use of the previously known analytic representation of the substrate deformation for the case considered. An approximate STW mode shape which accounts for spatial variations along all three axes is employed in the perturbation integral, allowing the derivation of an analytic expression relating the normal acceleration sensitivity to such parameters as substrate length, width, thickness, and orientation, mode center offset, mode shape spatial variations, etc.

Introduction

Surface transverse wave (STW) resonators have a number of potential performance advantages over their Rayleigh wave (SAW) counterparts. Firstly, STW resonators feature higher wave velocities leading to either higher frequency operation for a given feature size or less rigorous feature sizes for a given frequency of operation. Secondly, STW resonators feature much higher power handling capability leading to superior noise floor performance. Thirdly, the STW mode features a higher intrinsic Q·f product, with a value 2×10^{14} ($\tau_1 = 4.88$ fs) for the AT-cut.

In light of the substantial performance advantages of STW over SAW in these various areas, theoretical and experimental studies of the acceleration sensitivity of STW resonators have been undertaken. The purpose of the studies has been to understand the nature of STW acceleration sensitivity, and to determine whether the performance advantages seen in other areas extend to the case of acceleration sensitivity.

In this paper, the authors present an analysis of the normal acceleration sensitivity of rotated Y-cut quartz STW resonators simply supported along rectangular edges. The analysis is based upon the perturbation integral formulation of Tiersten, and makes use of the previously known [1] analytic representation of the substrate deformation for the case considered. An approximate STW mode shape which accounts for spatial variations along all three axes is employed in the perturbation integral, allowing the derivation of an analytic expression relating the normal acceleration sensitivity to such parameters as substrate length, width, thickness, and orientation, mode center offset, mode shape spatial variations, etc. While the case considered represents something of an overestimate as compared to the support configurations employed in production, questions regarding aspect-ratio compensation, optimum choice of substrate parameters, etc. are clarified by the case considered.

Basic Approach

The analysis of the acceleration sensitivity of surface transverse wave (STW) resonators is conveniently accomplished using the perturbation integral formulation of Tiersten [2], which we reproduce here for completeness:

$$\omega = \omega_\mu - \Delta_\mu , \tag{1}$$

$$\Delta_\mu = H_\mu / 2\omega_\mu , \tag{2}$$

0-7803-2500-1/95 $4.00 © 1995 IEEE

and

$$H_\mu = -\int_V g^\mu_{\alpha,M} \hat{c}_{L\gamma M\alpha} g^\mu_{\gamma,L} dV, \quad (3)$$

where

$$g^\mu_\gamma = u^\mu_\gamma / N_\mu, \quad (4)$$

and

$$N^2_\mu = \int_V \rho u^\mu_\gamma u^\mu_\gamma dV. \quad (5)$$

Equation (1) teaches that the frequency under acceleration, ω, is shifted from the unperturbed frequency, ω_μ, by a small amount Δ_μ which can be calculated using equation (2) and "the perturbation integral" of equation (3). The perturbation integral given by equation (3) looks quite complicated but actually has a rather simple interpretation: it is essentially a weighted average of the acceleration induced bias throughout the volume of the quartz plate, where the weighting factor for the averaging is determined by the acoustic mode shape. Equations (4) and (5) provide the necessary normalization of the acoustic mode shape.

The acceleration-induced baising state is most conveniently written using the material cofactor representation [3]

$$\hat{c}_{L\gamma M\alpha} = k_{L\gamma M\alpha KN} w_{N,K}, \quad (6)$$

where

$$k_{L\gamma M\alpha KN} = c_{L\gamma M\alpha KN} + c_{LMKN}\delta_{\gamma\alpha} + c_{L\gamma KM}\delta_{\alpha N} + c_{LKM\alpha}\delta_{\gamma N}. \quad (7)$$

The $w_{N,K}$ term in equation (6) represents the nine acceleration-induced biasing deformation gradients (N and K take values 1,2,3), and contains all of the required information on how the shape of the crystal plate is deformed by the acceleration. Which c-hat terms are required is determined by the mode of operation, and each c-hat term is the sum of the deformation gradients, each multiplied by the relevant material cofactors as defined in equation (7). Besides being computationally convenient, equations (6) and (7) clarify two fundamentally important concepts, namely 1) the frequency shifts because the plate is deformed and 2) the response to the deformation involves both linear and nonlinear elastic constants [3]. As applied to the calculation of the normal acceleration sensitivity, equations (1) through (3) readily yield

$$\Gamma_2 = \frac{1}{8\pi^2 \nu^2 a_2} \int_V g^\mu_{\alpha,M} \hat{c}_{L\gamma M\alpha} g^\mu_{\gamma,L} dV. \quad (8)$$

Equation (8) is typically evaluated numerically, often using finite element techniques to determine the acceleration-induced biasing state. While this approach has advantages in the analysis of complicated support structures and resonator geometries, it does not provide the designer with insight into the basic nature of the problem. As a consequence, we have chosen to analyze the problem at hand from an analytical standpoint. The goal of this work has been not simply to calculate the acceleration sensitivity, but to do so in such a fashion as to obtain an understanding of the functional dependencies of the acceleration sensitivity upon real-world design and fabrication parameters.

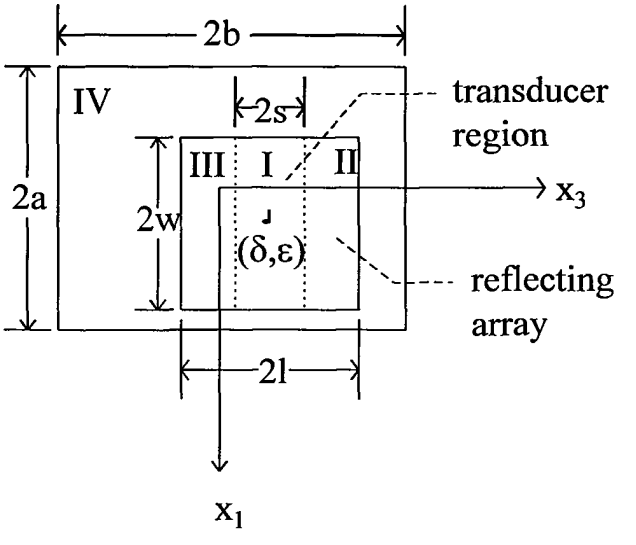


Figure 1. STW device plan view.

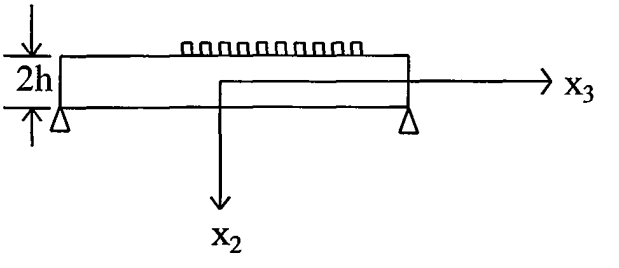


Figure 2. STW device cross-section.

Sample Geometry

The sample geometry for the case considered here is shown in Figures 1 and 2. The substrate is a rotated Y-cut quartz plate with width 2a along the crystallographic X-axis, thickness 2h along the rotated Y-axis, and length 2b along the rotated Z-axis, with the axes denoted X_1, X_2, and X_3 respectively in the coordinate system of the plate. The substrate is considered to be simply supported along the edges, and an acceleration a_2 is applied parallel to the X_2 axis.

The STW device is confined to the portion of the substrate denoted by regions I, II, and III in Figure 1. The STW metallization is considered to be on the face of the plate located at X_2=-h. The center of the STW region is displaced from the center of the substrate by some small amounts denoted as (δ,ε) along (X_1,X_3).

STW Acoustic Mode Shape

The STW mode considered here is a horizontal shear wave propagating along the X_3 direction with particle displacements polarized along the X_1 direction. The amplitude of the STW in the transducer region is taken to be uniform along the propagation direction, cosine along the transverse direction, and exponentially decaying into the substrate. The amplitude in the reflector regions is taken to include an exponential decay beginning at the transducer/reflector boundary, and is taken to be zero at the ends of the reflectors. For the u_1 particle displacement, we thus consider four regions:

Region I (transducers):

$$u_1^I = M\cos\left[\chi(x_1-\delta)\right]\exp\left[-\xi(x_2+h)\right]\cos\left[\beta(x_3-\varepsilon)\right],$$

$$\delta-w\le x_1\le\delta+w,-h\le x_2\le h,\varepsilon-s\le x_3\le\varepsilon+s. \quad (9)$$

Region II (r.h. reflector):

$$u_1^{II} = u_1^I\exp\left[-\zeta\left(x_3-\{\varepsilon+s\}\right)\right],$$

$$\delta-w\le x_1\le\delta+w,-h\le x_2\le h,\varepsilon+s\le x_3\le\varepsilon+l. \quad (10)$$

Region III (l.h. reflector):

$$u_1^{III} = u_1^I\exp\left[+\zeta\left(x_3-\{\varepsilon-s\}\right)\right],$$

$$\delta-w\le x_1\le\delta+w,-h\le x_2\le h,\varepsilon-l\le x_3\le\varepsilon-s \quad (11)$$

Region IV:

$$u_1^{IV} = 0, \text{ elsewhere.} \quad (12)$$

In addition, we take u_2=u_3=0 everywhere.

Flexural Biasing State In The Rotated Y-Cut Quartz Plate Simply Supported Along Rectangular Edges

The flexural biasing state in the rotated Y-cut quartz plate simply supported along rectangular edges has been determined by Shick and Tiersten [1]. The zeroth order, X_2-component of the flexural deformation is given by a Fourier series expansion:

$$w_2^{(0)} = \sum_m^{\text{odd}}\sum_n^{\text{odd}} A_{mn}\cos(\alpha_m x_1)\cos(\kappa_n x_3), \quad (13)$$

where

$$\alpha_m = \frac{m\pi}{2a}, \quad (14)$$

$$\kappa_n = \frac{n\pi}{2b}, \quad (15)$$

$$A_{mn} = \frac{48\rho a_2}{h^2 mn\pi^2 R_{mn}}, \quad (16)$$

and

$$R_{mn} = -\left[\gamma_{11}\alpha_m^4+\gamma_{33}\kappa_n^4+(2\gamma_{13}+4\gamma_{55})\alpha_m^2\kappa_n^2\right], \quad (17)$$

with

$$\gamma_{RS} = c_{RS} - c_{RW}c_{WV}^{-1}c_{VS}. \quad (18)$$

Note that equation (18) is a matrix equation, with R,S=1,3,5 and W,V=2,4,6. Analytic expressions for all of the biasing deformation gradients are then obtained from the zeroth order X_2 deformation as

$$w_{1,1} = -\left[w_{2,11}^{(0)}\right]\cdot x_2, \quad (19)$$

$$w_{1,2} = -w_{2,1}^{(0)} - 2\left[P_{31}w_{2,11}^{(0)} + P_{32}w_{2,33}^{(0)} + P_{33}w_{2,13}^{(0)}\right]\cdot x_2 \quad (20)$$

$$w_{1,3} = -\left[w_{2,13}^{(0)}\right]\cdot x_2, \quad (21)$$

$$w_{2,1} = w_{2,1}^{(0)}, \quad (22)$$

$$w_{2,2} = -\left[P_{11}w_{2,11}^{(0)} + P_{12}w_{2,33}^{(0)} + P_{13}w_{2,13}^{(0)}\right]\cdot x_2, \quad (23)$$

$$w_{2,3} = w_{2,3}^{(0)}, \quad (24)$$

$$w_{3,1} = -\left[w_{2,13}^{(0)}\right]\cdot x_2, \quad (25)$$

$$w_{3,2} = -w_{2,3}^{(0)} - 2\left[P_{21}w_{2,11}^{(0)} + P_{22}w_{2,33}^{(0)} + P_{23}w_{2,13}^{(0)}\right]\cdot x_2 \quad (26)$$

and

$$w_{3,3} = -\left[w_{2,33}^{(0)}\right]\cdot x_2, \quad (27)$$

where we have introduced the matrix P, defined as

$$P \equiv -c_{WV}^{-1}c_{VS} \quad (28)$$

with S=1,3,5 and W,V=2,4,6.

Net Biasing Terms (c-hat's)

For the STW mode considered here, six c-hat terms must be evaluated. Equations (13) through (28) are substituted into equation (6), and after a few algebraic manipulations yield simple expressions of two basic types.

TYPE I: when $L\gamma M\alpha = 3131,\ 3121,\ 2121,$ or 1111

$$\hat{c}_{L\gamma M\alpha} = \sum_{m}^{odd}\sum_{n}^{odd} A_{mn}\left[E_{L\gamma M\alpha}\alpha_m^2 + F_{L\gamma M\alpha}\kappa_n^2\right]\{\cos(\alpha_m x_1)\}\{x_2\}\{\cos(\kappa_n x_3)\} \quad (29)$$

where we have defined the quantities

$$E_{L\gamma M\alpha} = k_{L\gamma M\alpha 11} + P_{11}k_{L\gamma M\alpha 22} + 2P_{21}k_{L\gamma M\alpha 23} \quad (30)$$

and

$$F_{L\gamma M\alpha} = k_{L\gamma M\alpha 33} + P_{12}k_{L\gamma M\alpha 22} + 2P_{22}k_{L\gamma M\alpha 23}. \quad (31)$$

TYPE II: when $L\gamma M\alpha = 3111$ or 2111

$$\hat{c}_{L\gamma M\alpha} = -\sum_{m}^{odd}\sum_{n}^{odd} A_{mn}\alpha_m G_{L\gamma M\alpha}\sin(\alpha_m x_1)\cos(\kappa_n x_3) - \sum_{m}^{odd}\sum_{n}^{odd} A_{mn}\alpha_m\kappa_n H_{L\gamma M\alpha}\sin(\alpha_m x_1)x_2\sin(\kappa_n x_3) \quad (32)$$

where we have defined

$$G_{L\gamma M\alpha} = k_{L\gamma M\alpha 12} - k_{L\gamma M\alpha 21} \quad (33)$$

and

$$H_{L\gamma M\alpha} = k_{L\gamma M\alpha 13} + k_{L\gamma M\alpha 31} + 2P_{33}k_{L\gamma M\alpha 21}. \quad (34)$$

Net Normal Acceleration Sensitivity

The c-hat terms given by equations (29) through (34), along with the relevant normalized mode shape displacement gradients calculated using equations (4), (5), and (9) through (12), are substituted into equation (8) in order to obtain an expression for the normal acceleration sensitivity. After some tedious but tractable

integral calculus, followed by some algebraic manipulations, equation (8) yields

$$\Gamma_2 = \sum_{m}^{odd} \sum_{n}^{odd} \gamma_{mn}, \tag{35}$$

where

$$\gamma_{mn} = -\frac{1}{v^2} * \frac{24}{\pi^2} * \cos(\alpha_m \delta) * \cos(\kappa_n \varepsilon)$$

$$* \left[\frac{a^2 b^2 \left[(m/n) E_{3131} b^2 + (n/m) F_{3131} a^2 \right]}{\gamma_{11} m^4 b^4 + \gamma_{33} n^4 a^4 + (2\gamma_{13} + 4\gamma_{55}) m^2 n^2 a^2 b^2} \right]$$

$$* \left[\frac{2\sin(\alpha_m w)}{\alpha_m w} + \frac{\sin(\{2\chi + \alpha_m\} w)}{(2\chi + \alpha_m) w} + \frac{\sin(\{2\chi - \alpha_m\} w)}{(2\chi - \alpha_m) w} \right]$$

$$* \left[\left(\frac{1}{h} \right) \cdot \left(\frac{1 - 2\xi h \cdot \coth(2\xi h)}{2\xi h} \right) \right]$$

$$* \left\{ \left[\frac{2\sin(\kappa_n s)}{\kappa_n s} + \frac{\sin(\{2\beta + \kappa_n\} s)}{(2\beta + \kappa_n) s} + \frac{\sin(\{2\beta - \kappa_n\} s)}{(2\beta - \kappa_n) s} \right] \right.$$

$$+ \left[\frac{2\{2\zeta \cos(\kappa_n s) - \kappa_n \sin(\kappa_n s)\}}{\left[4\zeta^2 + \kappa_n^2 \right] s} \right.$$

$$- \frac{2\zeta \cos[(2\beta + \kappa_n) s] - (2\beta + \kappa_n) \sin[(2\beta + \kappa_n) s]}{\left[4\zeta^2 + (2\beta + \kappa_n)^2 \right] s}$$

$$\left. \left. - \frac{2\zeta \cos[(2\beta - \kappa_n) s] - (2\beta - \kappa_n) \sin[(2\beta - \kappa_n) s]}{\left[4\zeta^2 + (2\beta - \kappa_n)^2 \right] s} \right] \right\}$$

$$* \left[1 + \frac{1}{2\zeta s} + \left(\frac{\beta \zeta}{\beta^2 + \zeta^2} \right) \frac{\cos(2\beta s)}{2\beta s} + \left(\frac{\zeta^2}{\beta^2 + \zeta^2} \right) \frac{\sin(2\beta s)}{2\beta s} \right]^{-1}$$

+ 12 other smaller terms. (36)

Dependencies of Γ_2 Upon SubstrateParameters

The functional form given by equation (36) clarifies the dependencies of the normal acceleration sensitivity on various quantities of importance in the design and fabrication of STW resonators. Two distinct design paradigms are found which lead to low normal acceleration sensitivity: 1) the aspect-ratio compensation paradigm in which control of the substrate length to width ratio is used to obtain low normal acceleration sensitivity, and 2) a non-aspect-ratio compensation paradigm wherein the plate area is minimized and plate thickness is maximized to obtain low normal acceleration sensitivity.

Substrate Material Properties:

The acceleration sensitivity of an STW resonator is determined in part by the choice of substrate material through the values of various material constants.

Γ_2 is directly proportional to a sum of linear and nonlinear stiffnesses, which essentially govern the sensitivities to effective cavity length and wave velocity changes respectively.

Γ_2 is directly proportional to material mass density, as this contributes to the magnitude of the body force developed in response to the acceleration.

As expected, Γ_2 is inversely proportional to the linear stiffnesses controlling the flexure of the plate.

Substrate Orientation:

Γ_2 reflects the angular dependence of the linear and nonlinear elastic stiffnesses involved. As a consequence, the optimum design for low acceleration sensitivity varies slightly with cut angle.

Substrate Length:

For non-optimum aspect ratio, Γ_2 has a net proportionality to the square of the lateral dimensions (essentially, proportional to substrate area).

Substrate Width:

Similar to substrate length.

Substrate Thickness:

As expected, Γ_2 is inversely proportional to substrate thickness. Note that this is essentially the same effect seen in, for example, Figure 7 of [4].

Aspect Ratio (a/b):

Aspect ratio compensation arises from the fact that certain cofactor combinations have opposite signs. The required aspect ratio is dominated by the first term of equation (36) and may be estimated using

$$\sum_{m}^{odd}\sum_{n}^{odd}\left[(m/n)E_{3131}b^2+(n/m)F_{3131}a^2\right]\approx 0. \qquad (37)$$

The exact value of the optimum aspect ratio for a given substrate orientation, substrate thickness, etc. depends in detail upon the other smaller terms (several pages worth) whose functional dependencies are not the same as those of the first term.

The net normal acceleration sensitivity achievable using aspect ratio compensation is directly proportional to the dimensional tolerances achievable in production.

Dependencies of Γ_2 Upon Mode Shape Parameters

It is by now well known that the acceleration sensitivity has an important dependence upon the offset of the mode shape center with respect to the center of the plate [5]. The functional form given by equation (36) provides important insights into this and other dependencies. In particular, the acceleration sensitivity of an acoustic resonator is determined in part by the type of acoustic mode chosen, as the choice of mode type determines the relevant material constants.

Wave Velocity:

Γ_2 is inversely proportional to the square of the wave velocity. For a given substrate with all other factors being equal, higher velocity modes have inherently superior acceleration sensitivity.

Mode Shape Envelope:

The acoustic standing wave of the STW resonator exists within an amplitude envelope as described in equations (9) through (12). The functional form given by equation (36) indicates that spatial variations of the mode shape envelope do not dominate and hence cannot be used to minimize the net normal acceleration sensitivity.

Γ_2 has a moderately strong (factor of 1/3) but complicated dependence upon the X_3-axis (propagation direction) cross-section of the mode shape envelope, which in turn depends upon the details of transducer, cavity, and reflector design.

Γ_2 is independent of the X_1-axis cross-section of the mode shape envelope (transverse cosine order).

Γ_2 has a slight dependence on the ratio of acoustic aperture to substrate width.

Γ_2 has a weak dependence on the penetration depth into the bulk of the substrate.

Mode Center Offset:

The dominant term given in equation (36) is proportional to the cosines of the in-plane offsets along both X_1 and X_3 axes. In the design paradigm where aspect-ratio compensation is not used, the largest acceleration sensitivity occurs for the mode shape exactly on center.

The other smaller terms noted in equation (36) have various combinations of sine and cosine proportionalities to the in-plane offsets. As a consequence, when aspect-ratio compensation is employed, the net normal acceleration sensitivity is minimized for the mode shape exactly on center.

In the aspect-ratio compensation paradigm, the residual sensitivities to offsets along the X_1 and X_3 axes are of different signs and of different magnitudes, leading to an X-shaped locus of allowed mode shape centers for minimum normal acceleration sensitivity. For a given non-zero Γ_2, the direction in which the mode shape center needs to be moved to minimize Γ_2 is readily determined from the sign of the vector component.

Experimental Verification

The utility of equation (36) for calculating the normal acceleration sensitivity of STW resonators has been verified in part by comparing the calculated and measured performances for a variety of devices from multiple sources as listed in Table I. The substrate and

metallization dimensions for the various resonator types are listed in Table II. The ARL devices were mounted, unpackaged, onto a microstrip oscillator via "picture frames" and double-sided tape as described in [6], the Raytheon devices were typical all-quartz package (AQP) devices as described in [7], and the SAWTEK devices were uniformly supported inside ceramic hybrid packages. It is important to note that none of the devices listed in Tables I and II are aspect-ratio compensated. In this paradigm, the utility of equation (36) in providing design guidance clearly extends far beyond the simply-supported, thin-plate case.

The utility of equation (36) in precise calculations using aspect-ratio compensation is somewhat less as illustrated in Figure 3. In this Figure, the measured values of $|\Gamma_2|$ at various orientation angles are plotted along with the corresponding values calculated using equation (36) in conjunction with the third-order elastic stiffnesses of Thurston [8] and Stern [9] as noted. The maximum data and minimum data curves represent respectively the largest and smallest observed values of $|\Gamma_2|$ over the range of vibration frequencies from 100Hz to 2kHz.

The measured data indicate a minimum acceleration sensitivity magnitude for a uniformly supported STW with an aspect ratio of 0.465 at $\theta \approx 36°$. Calculations using equation (36) in conjunction with the third-order elastic stiffnesses of Thurston and Stern place the minimum at $\theta \approx 28.5°$ and $\theta \approx 20.5°$ respectively. Equivalently, for $\theta \approx 36°$, the experimentally determined optimum aspect ratio for a uniform support is approximately 0.465, while the optimum values calculated using equation (36) in conjunction with the third-order elastic stiffnesses of Thurston and Stern are 1.15 and 10.8 respectively.

The exact reason for these discrepancies remains unclear, but it seems likely that they arise from defects in the published values of third-order elastic stiffnesses in combination with substantial deviations from the simply-supported, thin-plate assumptions. We note that the discrepancies involved are quite similar to those encountered elsewhere for the Rayleigh wave case [1, 10-12].

Conclusions

A closed-form expression for the normal acceleration sensitivity of STW resonators using thin,

TABLE I
MEASURED AND CALCULATED STW $|\Gamma_2|$

	Source	θ deg.	v MHz	v m/sec	Calc. 10^{-10}/g	Meas. 10^{-10}/g
1	ARL	42.75	356	4984	117	200-400
2	Raytheon	36.7	1001	5072	13.5	10-20
3	SAWTEK	36	500	5000	8.08	3-6
4	SAWTEK	36	500	5000	5.68	2-30

TABLE II
STW DEVICE SUBSTRATE PARAMETERS

	2a (mm)	2b (mm)	2h (mm)	2s (mm)	2l (mm)	2w (mm)
1	14.00	14.00	0.51	3.39	17.38	1.40
2	3.81	7.62	0.89	0.79	5.35	0.50
3	1.85	5.82	0.48	1.57	5.15	1.40
4	1.37	5.82	0.48	1.57	5.15	0.92

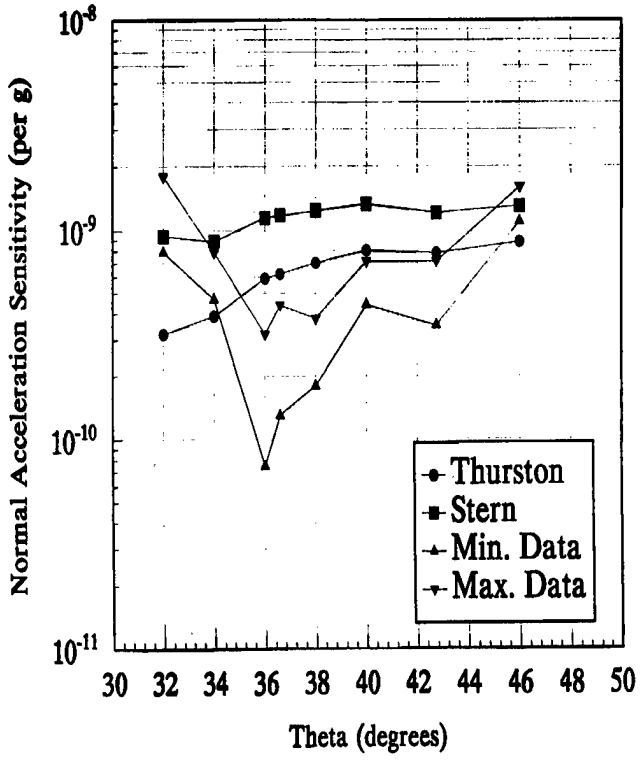


Fig. 3. Calculated and measured oreintation angle dependence of $|\Gamma_2|$ for STW resonators with a/b=0.465 [13].

rotated Y-cut quartz plates simply supported along rectangular edges has been derived and validated through comparison with experimental data. From the closed-form expression, the functional dependencies of the normal acceleration sensitivity have been determined. The most important parameters for obtaining low acceleration sensitivity are found to be the substrate dimensions and dimensional tolerances.

Acknowledgment

The authors wish to thank Dr. Gary Montress for providing the acceleration sensitivity data and design parameters for the Raytheon devices as listed in Tables I and II.

References

[1] D. V. Shick and H. F. Tiersten, "An Analysis of the Acceleration Sensitivity of ST-cut Quartz Surface Wave Resonators Supported Along the Edges," in Proceedings of the 40th Annual Frequency Control Symposium, May 1986, pp. 262-268.

[2] H. F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," Journal of the Acoustical Society of America, vol. 64, no. 3, pp. 832-837, September 1978.

[3] J. A. Kosinski and A. Ballato, "Designing for Low Acceleration Sensitivity," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 40, no. 5, pp. 532-537, September 1993.

[4] P. C. Y. Lee and M. S. H. Tang, "Acceleration Insensitivity of Thickness Frequencies of Doubly Rotated Quartz Crystal Disks," in Proceedings of the 42nd Annual Frequency Control Symposium, June 1988, pp. 14-18.

[5] E. P. EerNisse, R. W. Ward, M. H. Watts, R. B. Wiggins, and O. L. Wood, "Experimental Evidence for Mode Shape Influence on Acceleration-Induced Frequency Shifts in Quartz Resonators," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 37, no. 6, pp. 566-570, November 1990.

[6] D. Huynh, R. McGowan, J. Kosinski, J. Stewart, R. Piekarz, and C. D. Mulford, "Experimental Studies of SAW and STW Acceleration Sensitivity," elsewhere in these Proceedings.

[7] T. E. Parker, J. Callerame, and G. K. Montress, "A New All Quartz Package for SAW Devices," in Proceedings of the 39th Annual Frequency Control Symposium, May 1985, pp. 519-525.

[8] R. N. Thurston, H. J. McSkimin, and P. Andreatch, "Third-Order Elastic Coefficients of Quartz," Journal of Applied Physics, vol. 37, no. 1, pp. 267-275, January 1966.

[9] R. Stern and R. T. Smith, "On the Third-Order Elastic Moduli of Quartz," Journal of the Acoustical Society of America, vol. 44, pp. 640-641, 1968.

[10] H. F. Tiersten and D. V. Shick, "An Analysis of the Normal Acceleration Sensitivity of ST-cut Quartz Surface Wave Resonators Rigidly Supported Along the Edges," in Proceedings of the 41st Annual Frequency Control Symposium, May 1987, pp. 282-288.

[11] H. F. Tiersten and D. V. Shick, "An Analysis of the Normal Acceleration Sensitivity of ST-cut Quartz Surface Wave Resonators with the Substrate Extending Beyond the Supports," in Proceedings of the 2nd European Frequency and Time Forum, March 1988, pp. 753-765.

[12] J. A. Greer and T. E. Parker, "Improved Vibration Sensitivity of the All Quartz Package Surface Acoustic Wave Resonator," in Proceedings of the 42nd Annual Frequency Control Symposium, June 1988, pp. 239-251.

[13] M. S. Cavin and R. C. Almar, "An Oscillator Design UsingLow G-Sensitivity, Low Phase Noise STW Devices," elsewhere in these Proceedings.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

EXPERIMENTAL STUDIES OF SAW AND STW ACCELERATION SENSITIVITY

D. Huynh, R. McGowan, J. A. Kosinski, J. T. Stewart, R. Piekarz, and C. D. Mulford

US Army Research Laboratory, ATTN: AMSRL-PS-ED, Fort Monmouth, New Jersey 07703-5601

Abstract: Comparative studies of the normal acceleration sensitivities of 360 MHz SAW and STW two-port resonators have been performed. The resonators were fabricated on the same ST-cut wafer to insure the clearest comparison. The results show that the normal acceleration sensitivities of both types of devices can easily be improved one order magnitude by selecting the appropriate aspect-ratio of the device substrate and/or mounting system. Based on these results, the normal acceleration sensitivities of the STW resonators have been further improved through modifications to the existing resonators, resulting in normal acceleration sensitivity as low as 2.93×10^{-10}/g. The measured values of the STW normal acceleration sensitivities are compared to that calculated for the case of simple supports along rectangular edges.

Introduction

Surface transverse wave (STW) resonators are currently being proposed as the basis for high performance oscillators in a variety of military electronic systems. In these applications, phase noise under vibration, and hence acceleration sensitivity, are critical parameters. Unlike the SAW case for which a significant body of literature exists, little has been published to date on the acceleration sensitivity of STW resonators. As a consequence, comparative studies of the normal acceleration sensitivities of 360 MHz SAW and STW two-port resonators have been performed. The primary focus of the studies has been to determine whether SAW or STW has an intrinsic advantage over the other in regard to acceleration sensitivity, and to evaluate the use of aspect-ratio compensation for reducing the normal acceleration sensitivity of STW resonators.

Experimental Procedure

Two-Port Resonators

The two-port SAW and STW resonators used in the comparative studies were designed and fabricated at the U.S. Army Research Laboratory's Physical Sciences Directorate facility located at Fort Monmouth, New Jersey. All of the devices were fabricated on ST-cut (θ=42.75°) quartz using three inch diameter, 20 mil (0.508 mm) thick wafers, taking care to avoid the seed region. For this study, two wafers were processed containing a total of eight devices.

To insure the clearest comparison, both SAW and STW resonators were fabricated on each wafer. All of the devices were designed for a 360 MHz nominal center frequency. Using ST-cut quartz, the SAW devices were temperature compensated while the STW devices were not. The thickness of the Al metallization (1400 Å) was chosen as a compromise between the desired 1% metallization for SAW and the desired 2% metallization for STW. Again, to insure the clearest comparison, the wafers were diced such that the substrates for both SAW and STW resonators were 20 mm x 20 mm squares.

The two-port SAW resonators were driven and sampled by 70½ finger pair transducers spaced approximately 100λ apart. The reflective gratings consisted of 1000 shorted metal strips with a period of 4.36μm and an acoustic aperture of 100λ. The reflective gratings were left floating.

The design of the two-port STW resonators was as close as possible to that of the SAW resonators. The STW resonators were driven and sampled by 70½ finger pair transducers spaced 100λ apart. The reflective gratings consisted of 1000 shorted metal strips with a period of 7.00 μm and an acoustic aperture of 100λ. The reflective gratings were left floating. An inter-transducer grating with a period of 6.80 μm was used to trap the wave in the cavity.

Test Oscillator

A basic feedback loop oscillator circuit compatible with both resonator types was developed for the comparative studies. The oscillator circuit consists

U.S. Government work not protected by U.S. copyright

of, in order, the acoustic resonator, a multistage electronic phase shifter, a single op amp, and a 20 dB output coupler. The physical implementation of the oscillator is shown in Figure 1. The oscillator circuit was implemented in microstrip using a 0.254 mm thick Duroid board with 1/2 oz. of rolled copper on the top surface. The 9.8 cm x 8.3 cm circuit board was rigidly attached to a 6.35 mm thick aluminum backing on the bottom surface. For the vibration tests, the oscillator was operated at the frequency of minimum insertion loss of the resonator being tested.

Resonator Mounting System

In order to facilitate the comparative studies, a simple, interchangeable resonator mounting system was developed. The mounting system consisted of a set of "picture frames" (see Figure 2) which were attached

TABLE I
SUPPORT FRAME INNER DIMENSIONS

Frame Number	Inner Length	Inner Width	Aspect-Ratio
1	0.000 mm	14.000 mm	0
2	4.666 mm	14.000 mm	1:3
3	5.600 mm	14.000 mm	2:5
4	7.000 mm	14.000 mm	1:2
5	9.333 mm	14.000 mm	2:3
6	14.000 mm	14.000 mm	1:1
5	14.000 mm	9.333 mm	3:2
4	14.000 mm	7.000 mm	2:1
3	14.000 mm	5.600 mm	5:2
2	14.000 mm	4.666 mm	3:1
1	14.000 mm	0.000 mm	∞

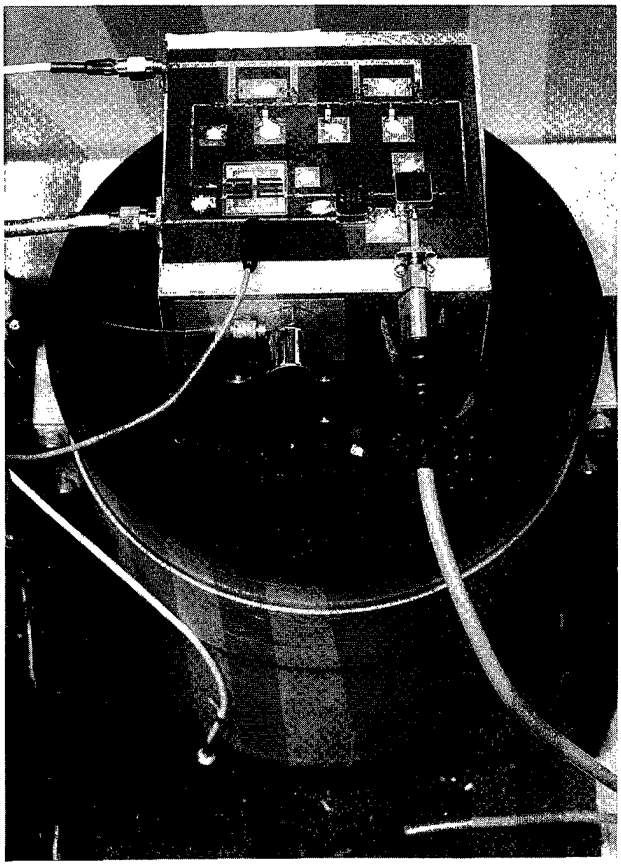


Figure 1. Test oscillator and shake-table apparatus. An STW resonator is mounted in the test oscillator.

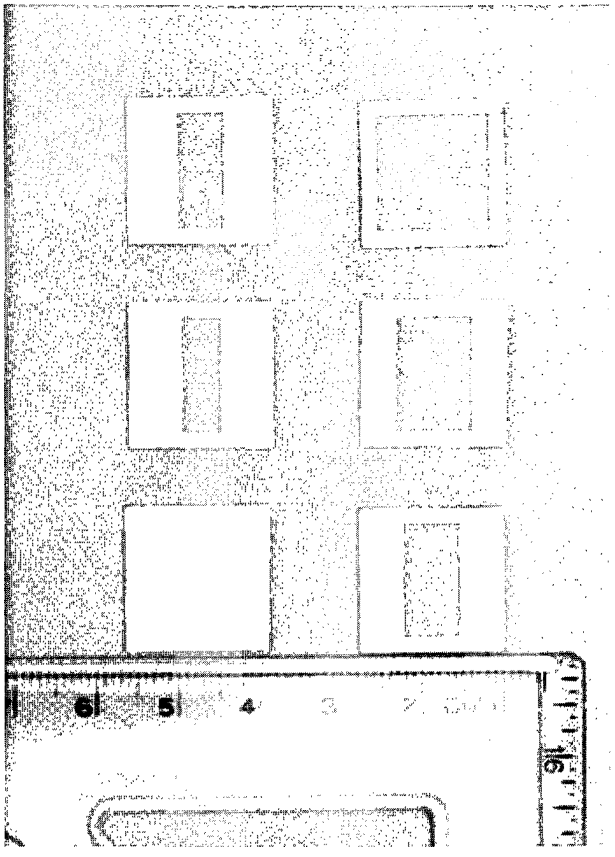


Figure 2. Support frames for mounting the resonators.

between the oscillator circuit board and the acoustic resonator using a double-sided adhesive tape. The support frames were sheared and machined from a sheet of 0.508 mm thick molybdenum material. The outer dimensions of all of the support frames were fixed at 18 mm x 18 mm. The inner width dimension of each frame was cut and machined to 14 mm, while the inner length dimensions of the frames were cut and machined to yield a variety of aspect-ratios as listed in Table I. The inverse of the intended aspect-ratios were obtained by rotating each frame by 90°, and taking the inner length to be fixed. A solid plate (no inner opening) was used for the limiting aspect-ratio values of zero and infinity.

Shake Table Test System

The normal acceleration sensitivities of the various SAW and STW resonators were measured using a shake table test system described in detail elsewhere [1]. A four inch aluminum cube was bolted directly to the shake table, and served as the support for the test oscillator as shown in Figure 1. The test oscillator was uniformly attached to the aluminum cube using double-sided adhesive tape. The dc lines for both power and tuning voltage were shielded.

The test oscillator was subjected to a sinusoidal acceleration normal to the plane of the resonator. An accelerometer mounted on the test oscillator circuit was used in a feedback loop to maintain a peak acceleration level of 1 g along the desired test axis. A second accelerometer was used to monitor any unwanted transverse acceleration. The tests were performed over the range of vibration frequencies from 90 Hz to 10 kHz. The vibration-induced FM sideband levels were measured directly using an rf spectrum analyzer. At each vibration frequency, measurements were taken of the power levels and frequencies of the carrier and the first upper and lower sidebands. The acceleration sensitivity was calculated by using the average of the upper and lower sideband power levels, the applied vibration frequency, the carrier frequency, and the peak acceleration along the desired vibration axis. The effects of temperature-induced frequency drift were overcome by using a single sweep mode to record the measurement at each vibration frequency.

In most cases, the small modulation index approximation of Filler [2] was used to interpret the measured data. For certain aspect-ratios using both SAW and STW resonators, the first sideband-to-carrier power level ratio $\mathcal{L}^{1}_{v}(f)$ exceeded -26 dBc, resulting in a modulation index greater than 0.1. Under these conditions, the "moderate" modulation index approximation was utilized to accurately calculate the acceleration sensitivity [1].

Experimental Results

SAW Resonators

The measured normal acceleration sensitivities of the SAW resonators tested are plotted as a function of aspect-ratio in Figure 3. The acceleration sensitivity data are presented in terms of the maximum and minimum normal acceleration sensitivities observed over the range of vibration frequencies from 90 Hz to 990 Hz. Note that the SAW devices are taken to be of length 2a along the X_1 axis and width 2b along the X_3 axis, in which case the aspect-ratio a/b corresponds to that defined by Tiersten [3]. For graphical purposes, the normal acceleration sensitivities measured using the solid plate configuration (aspect-ratios of zero and infinity) are plotted as aspect-ratios of zero and four.

The largest acceleration sensitivities are obtained using the support frame with an aspect-ratio of 1:1, while the smallest acceleration sensitivities are obtained using the solid plate configuration. A local minimum appears to be indicated at an aspect-ratio of 2.5:1.

STW Resonators

The measured normal acceleration sensitivities of the STW resonators tested are plotted as a function of aspect-ratio in Figure 4. The acceleration sensitivity data are presented in terms of the maximum and minimum normal acceleration sensitivities observed over the range of vibration frequencies from 90 Hz to 990 Hz. Note that, for consistency with the analytical results of [4], the STW device is taken to be of width 2a along the X_1 axis and length 2b along the X_3 axis, i.e., the dimensions a and b are assigned to the X_1 and X_3 axes respectively and not to the length and width directions. For graphical purposes, the normal acceleration sensitivities measured using the solid plate configuration (aspect-ratios of zero and infinity) are plotted as aspect-ratios of zero and four.

The largest acceleration sensitivities are obtained using the support frame with an aspect-ratio of 1:1, while the smallest acceleration sensitivities are obtained using the solid plate configuration. The presence of any local minima are not clearly indicated.

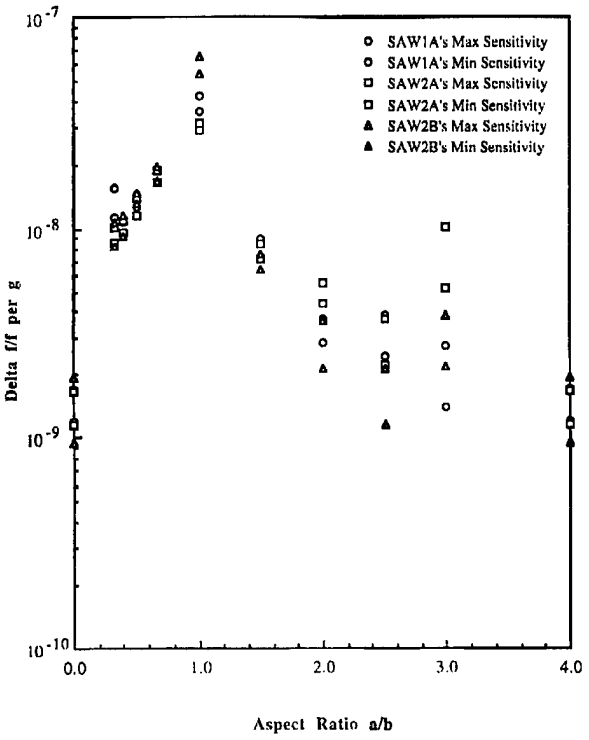


Figure 3. Measured SAW resonator normal acceleration sensitivity versus aspect-ratio.

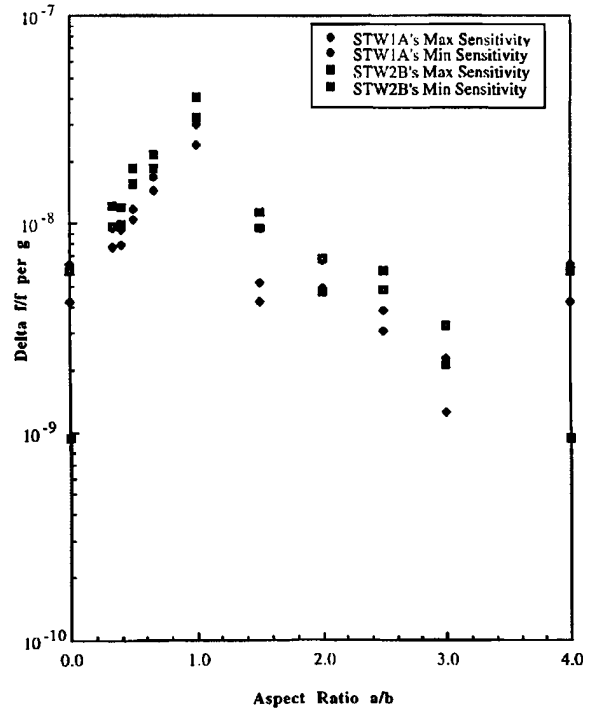


Figure 4. Measured STW resonator normal acceleration sensitivity versus aspect-ratio.

Discussion

The solid plate mounting configuration yields the lowest normal acceleration sensitivities for both SAW and STW resonators. For the SAW resonators, the measured data using the solid plate mount ranged from 9.57×10^{-10}/g to 1.94×10^{-9}/g, while for the similarly mounted STW resonators the data ranged from 9.34×10^{-10}/g to 5.99×10^{-9}/g. The data indicated essentially identical best case performance for the two types of devices. The smaller scatter in the SAW data using the solid plate mount are attributed to a somewhat more uniform adhesion of the double-sided tape achieved by pressing with a pair of tweezers in the inter-transducer gap when mounting the devices. The same could not be done with the STW devices due to the inter-transducer grating.

Using the 1:1 aspect-ratio support frame mounting, the STW resonators performed marginally better than the SAW resonators. The apparent local minimum at or near an aspect-ratio of 2.5:1 for the SAW case is in general agreement with the value of 1.8:1 reported in [5].

In Figure 5, the measured data for the STW resonators are compared with an analytic calculation of the normal acceleration sensitivities of comparable STW resonators simply supported along rectangular edges with the support dimensions corresponding to the inner dimensions of the mounting support frames for the various cases considered. The calculation appears to provide a reasonable estimate of the minimum observed normal acceleration sensitivity data. This is consistent with the lessons of practical experience that acceleration sensitivity is readily made worse but rarely, if ever, made better by accident. The calculation anticipates an aspect-ratio compensation of the normal acceleration sensitivity for an aspect-ratio of 1.8:1. The data may indicate a local minimum near 1.5:1, however additional measurements with finer aspect-ratio resolution will be required to provide a definitive answer.

After completing the comparative studies, the STW devices were subjected to a series of modifications (different modifications for each device) to further verify the understanding gained from the experimental and analytical studies. The results to date are shown in Figure 6. The normal acceleration sensitivities of the 360 MHz STW resonators were reduced to as low as 2.93×10^{-10}/g after the series of modifications.

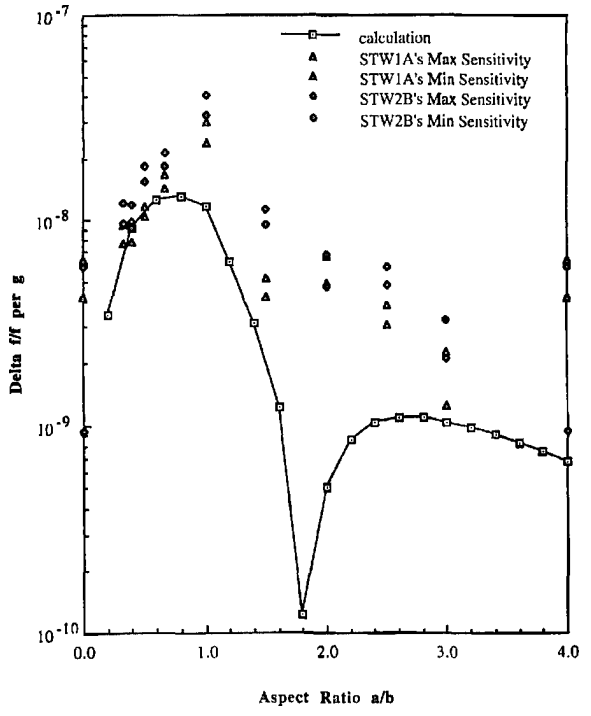


Figure 5. Comparison of measured and calculated STW normal acceleration sensitivity.

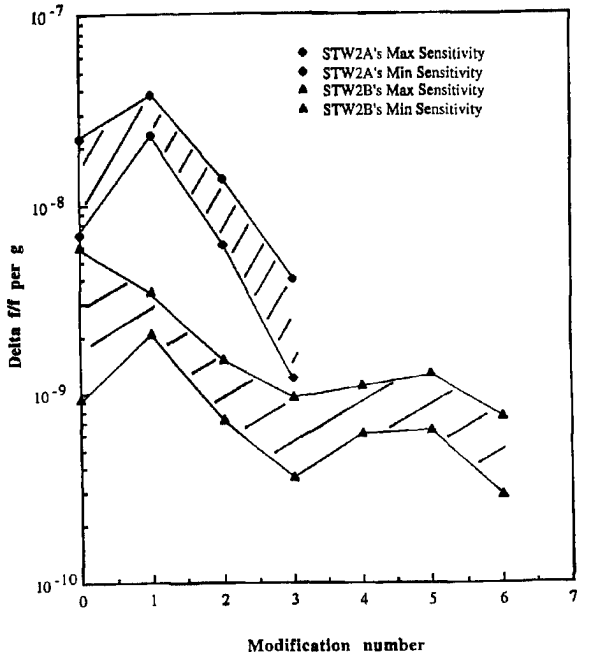


Figure 6. Measured STW resonator normal acceleration sensitivity versus modification number.

Conclusion

Comparative studies of the acceleration sensitivities of 360 MHz SAW and STW two-port resonators have been performed. Neither SAW nor STW exhibited any significant advantage over the other in regard to normal acceleration sensitivity.

Acknowledgment

The authors wish to thank Sally Laffey of Vitronics, Inc. for assistance with the resonator mounting/demounting and cleaning procedures, and Frank McBride of CECOM for assistance in developing the support frame mounting structure.

References

[1] J. A. Kosinski and A. Ballato, "Advances in Acceleration Sensitivity Measurement and Modeling," Proceedings of the 1992 IEEE Frequency Control Symposium, May 1992, pp. 838-848.

[2] R. L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: A Review," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 35, No. 3, pp. 297-305, May 1988.

[3] D. V. Shick and H. F. Tiersten, "An Analysis of the Acceleration Sensitivity of ST-cut Quartz Surface Wave Resonators Supported Along the Edges," Proceding of the 40th Annual Frequency Control Symposium, May 1986, pp. 262-268.

[4] J. A. Kosinski, J. T. Stewart, A. Ballato, and R. Almar, "An Analysis of the Normal Acceleration Sensitivity of Rotated Y-cut Quartz Surface Transverse Wave Resonators Simply Supported Along Rectangular Edges," elsewhere in these Proceedings.

[5] J. A. Greer and T. E. Parker, "Improved Vibration Sensitivity of the All Quartz Package Surface Acoustic Wave Resonator," Proceeding of the 42nd Annual Frequency Control Symposium, June 1988, pp. 239-251.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SEMI-ANALYTICAL FINITE ELEMENT ANALYSIS OF ACCELERATION-INDUCED FREQUENCY CHANGE IN SAW RESONATORS

J.T. Stewart, R. McGowan, J. A. Kosinski and A. Ballato

US Army Research Laboratory, ATTN: AMSRL-PS-ED, Fort Monmouth, NJ 07703-5601

Abstract: A quantitative analysis tool has been developed which determines the acceleration induced frequency change for surface acoustic waves (SAW) traveling in an arbitrary piezoelectric crystal resonator. This analysis tool uses a finite element approach to solve the static bias problem and Sinha & Tiersten's analytical technique to determine the SAW mode shape. Combining this data with the spatially varying effective elastic constants, the numerical perturbation integral derived by Tiersten determines the acceleration sensitivity of the SAW resonator. The unique aspect of this software is the determination of the spatially varying effective elastic constants using finite element shape functions. The versatility of the tool will be demonstrated by modeling three mounting structures that have been experimentally verified. These mounting structures include two plane stress problems and a plane strain problem.

Introduction

The current emphasis on minimizing development time by incorporating the principles of the system engineering process and concurrent engineering has lead to increased use of computer aided design (CAD) and computer aided analysis (CAA) tools to achieve first run design success. In the field of acoustics, especially when surface acoustic wave devices are considered, many CAD and CAA tools have been developed to aid the engineer in simulating proposed designs. These tools allow the designer to go from a set of design requirements for a SAW device to both a physical layout of the device and a prediction of expected device performance [1-3]. Theoretical device performance calculations use the physical dimensioning of the device combined with the material constants of the piezoelectric substrate to generate the expected device performance data. These analysis tools provide excellent simulations of device characteristics such as quality factor, input and output impedance, and frequency versus amplitude response.

Most of these critical design criteria are well understood for SAW. However, there are many specified requirements for systems that contain acoustic devices which are not well understood and result in iterative engineering solutions where several prototype devices are fabricated for evaluation. Some of these effects include processing, meteorology (temperature, pressure, humidity), residual noise, aging, force, and vibration. Although all of these effects are intriguing and warrant investigation, this paper focuses primarily on the effects of vibration and mode shape on the output frequency of a SAW resonator.

The solution algorithm that was selected to solve the acceleration sensitivity of a SAW device on an arbitrary piezoelectric substrate has four key elements. First, the substrate is parameterized and a finite element mesh of the physical problem is generated. Next, the forces are applied to the finite element mesh and the nodal displacements are calculated which forming the solution of the static biasing problem. The material constants and physical characteristics of the SAW device are then used in Sinha and Tiersten's analytical technique to determine the mode shape of the generated SAW. Finally, the mode shape and the spatially varying effective elastic constants determined by the static bias are employed in Tiersten's perturbation integral, and the acceleration sensitivity of the SAW device is evaluated numerically.

The versatility and power of this tool has been verified by using this algorithm to model three examples from the literature [4,5,6]. Good agreement was observed between the results presented in the references and those obtained using this CAA tool.

Acceleration Sensitivity Algorithm

Solution of The Biasing State

The development of the static finite element equations for the solution of the biasing state begins with the general three dimensional equations of elasticity

U.S. Government work not protected by U.S. copyright

$$T_{ij,i} + \rho b_j = 0, \tag{1}$$

$$s_{ij} = \frac{1}{2}(w_{i,j} + w_{j,i}), \tag{2}$$

and

$$T_{ij} = c_{ijkl} s_{kl}, \tag{3}$$

where T_{ij} are the components of the stress tensor, b_i are the components of body force per unit volume, s_{ij} are the components of the infinitesimal strain tensor with w_i representing the components of displacement, c_{ijkl} are the components of the 2nd order elastic constants, and ρ is the mass density of the material. The variational form, or weak form, of equations (1) - (3) is formulated for a body occupying a volume V bounded by a surface S as

$$\delta\Pi = \int_V T_{ij}\delta s_{ij} dV - \int_V b_i \delta w_i dV - \int_S t_i \delta w_i dS = 0 \tag{4}$$

The finite element disctretization process is applied by interpolating the displacements with a set of shape functions, N^q as follows:

$$w_j = N^q \bar{w}_j^q \tag{5}$$

where $\bar{w}_j^q$ are the nodal displacements. In equation (5) the superscripts are intended to imply a sum over nodes within an element or an entire mesh, depending upon context. This notation will be employed throughout to save space. The shape functions N^q may take on several forms and will not be explicitly defined here. The reader is referred to [7] and [8] for these and other omitted finite element definitions. Using equation (5) in the functional (4) gives

$$\delta\Pi = \left[\int_\Omega \frac{\partial N^p}{\partial x_i} c_{ijkl} \frac{\partial N^q}{\partial x_l} d\Omega \bar{w}_k^q - \int_\Omega b_j N^p d\Omega - \int_\Gamma t_j N^p d\Gamma\right] \delta\bar{w}_j^p = 0 \tag{6}$$

Here Ω represents the discretized domain. For arbitrary variations $\delta\bar{w}_j^p$ equation (6) reduces to

$$K_{jk}^{pq} \bar{w}_k^q - F_j^p = 0 \tag{7}$$

with the elemental equation:

$$k_{jk}^{pq} = c_{ijkl} G_{il}^{pq}, \tag{8}$$

where K_{jk}^{pq} and F_j^p are the global stiffness matrix and load vector, with k_{jk}^{pq} and f_j^p the corresponding element or local quantities, and

$$G_{il}^{pq} = \int_{\Omega_e} \frac{\partial N^p}{\partial x_i} \frac{\partial N^q}{\partial x_l} d\Omega_e, \tag{9}$$

and

$$f_j^p = \int_{\Omega_e} b_j N^p d\Omega_e + \int_{\Gamma_e} t_j N^p d\Gamma_e. \tag{10}$$

Here Ω_e represents the a single element domain with bounding surface Γ_e. The global system is assembled in the usual way giving rise to the matrix problem

$$\underset{\sim}{K}\underset{\sim}{w} = \underset{\sim}{F} \tag{11}$$

with solution

$$\underset{\sim}{w} = \underset{\sim}{K}^{-1}\underset{\sim}{F}. \tag{12}$$

Calculation of the SAW Mode Shape

The general three dimensional surface acoustic wave mode shape is obtained from the straight crested solution obtained by Sinha and Tiersten [11,12,13]

$$\hat{\alpha}(x_2) = \sum_{n=1}^{4} C^{(n)} A_j^{(n)} e^{-\beta_n \xi x_2} \tag{13}$$

where $C^{(n)}$ and $A_j^{(n)}$ are amplitude ratios, β_n are the decay constants along the x_2 direction, and ξ is the straight crested propagation number along the x_1 direction. With this solution, the transformed variably crested solution is obtained by replacing ξ with a modified wave number, ζ, such that

$$\zeta = \sqrt{\kappa^2 + \xi^2} \tag{14}$$

where κ is the approximate wave number along x_3 for the m^{th} transverse mode given by

$$\kappa = \frac{m\pi}{2w} \tag{15}$$

with 2w denoting the width of a strip. With this

transformed variably crested solution, the acoustic field, $u_j(x_1, x_2, x_3)$, in the transmission path can be written as a purely real function as

$$u_j(x_1, x_2, x_3) = f(x_3) \sum_{n=1}^{4} [\bar{A}_j^{(n)} \cos(\zeta x_1) + \bar{B}_j^{(n)} \sin(\zeta x_1)] e^{-\nu_n x_2} \quad (16)$$

where $\bar{A}_j^{(n)}$ and $\bar{B}_j^{(n)}$ are functions of $C^{(n)}$ and $A_j^{(n)}$, $\nu_n = \beta_n \zeta$, and $f(x_3)$ is defined for odd and even harmonics as

$$f(x_3) = \begin{cases} \cos(\kappa x_3) & n = 0,2,4,... \\ \sin(\kappa x_3) & n = 1,3,5,... \end{cases} \quad (17)$$

The general solution in the reflector arrays is obtained by solving the approximate two dimensional surface wave equations obtained from the variational formulation of Sinha and Tiersten [11,12,13] and using the resulting transmission matrix in the difference equation solution by Tiersten, et al., [14] to obtain the effective decay constants along x_1. Using this, the surface wave solution in the reflector arrays may be written as a purely real function of the form

$$u_j(x_1, x_2, x_3) \quad (18)$$
$$= f(x_3) \sum_{n=1}^{4} \{\left(\tilde{A}_j^{(n)} e^{-\alpha_1 x'_1} + \tilde{C}_j^{(n)} e^{-\alpha_2 x'_1}\right) \cos\zeta x'_1 + \left(\tilde{B}_j^{(n)} e^{-\alpha_1 x'_1} + \tilde{D}_j^{(n)} e^{-\alpha_2 x'_1}\right) \sin\zeta x'_1 \} e^{-\nu_n x_2}$$

where α_1 and α_2 are decay constants, $\tilde{A}_j^{(n)}$, $\tilde{B}_j^{(n)}$, $\tilde{C}_j^{(n)}$, and $\tilde{D}_j^{(n)}$ are constants related to $\bar{A}_j^{(n)}$ and $\bar{B}_j^{(n)}$, as well as scale factors derived from the difference equation solutions [12]. The variable x'_1 is simply a translated x_1 for each reflector array. The normalized mode shape is then obtained as

$$g_j^{\mu} = u_j / N \quad (19)$$

where N is a normalizing constant given as

$$N^2 = \int_V \rho u_j u_j dV \quad (20)$$

The amplitude envelopes of the first three transverse harmonics of a typical surface wave mode are plotted in Figures 1 through 3.

Calculation of the Frequency Shift

The frequency shift under a given static biasing state is computed using Tiersten's perturbation method [10] for small fields superposed on a bias [9]. The change in resonant frequency of the μ^{th} eigen-mode is given as

$$\Delta_{\mu} = H_{\mu} / 2\omega_{\mu}, \quad (21)$$

where

$$H_{\mu} = -\int_V \hat{c}_{L\gamma M\alpha} g^{\mu}_{\alpha,M} g^{\mu}_{\gamma,L} dV \quad (22)$$

and

$$\hat{c}_{L\gamma M\alpha} = T^1_{LM} \delta_{\gamma\alpha} + c_{L\gamma\alpha MAB} E^1_{AB} + c_{L\gamma NM} w_{\alpha,N} + c_{LN\alpha M} w_{\gamma,N} \quad (23)$$

The components $\hat{c}_{L\gamma M\alpha}$ are spatially varying effective elastic constants derived from the biasing state with T^1_{LM} the biasing stresses, E^1_{AB} the biasing strains, and $w_{\gamma,N}$ the biasing deformation gradients. The constants $c_{L\gamma\alpha MAB}$ denote the third order elastic constants for the material.

The perturbation integral (22) is evaluated as a sum over N elements as

$$H_{\mu} = -\sum_{i=1}^{N} \int_{\Omega_i} \hat{c}_{L\gamma M\alpha} g^{\mu}_{\alpha,M} g^{\mu}_{\gamma,L} d\Omega_i \quad (24)$$

where Ω_i denotes the volumetric domain of the i^{th} element. To evaluate equation (24) the necessary quantities need to be sampled at a specific point, (η_1, η_2, η_3), in the interior of the element domain. This can be achieved by using the solution vector with equation (5) to obtain

$$E^1_{ij}(\eta_1, \eta_2, \eta_3) = \frac{1}{2}\left(\frac{\partial N^p}{\partial x_j} \bar{w}^p_i + \frac{\partial N^p}{\partial x_i} \bar{w}^p_j\right) \quad (25)$$

and

$$\Omega^1_{ij}(\eta_1, \eta_2, \eta_3) = \frac{1}{2}\left(\frac{\partial N^p}{\partial x_j} \bar{w}^p_i - \frac{\partial N^p}{\partial x_i} \bar{w}^p_j\right) \quad (26)$$

where Ω_{ij}^1 are the biasing rotations. It should be noted that E_{ij}^1 and Ω_{ij}^1 are sampled in the interior of the element domain where they are accurate. In general these quantities should not be considered accurate at nodal points ar along element edges. Using equations (25) and (26), the deformation gradients at the element domain point (η_1, η_2, η_3) are given as:

$$w_{i,j}(\eta_1, \eta_2, \eta_3) = E_{ij}(\eta_1, \eta_2, \eta_3) + \Omega_{ij}(\eta_1, \eta_2, \eta_3) \quad (27)$$

The biasing stresses, T_{LM}^1, at this point are then computed using equation (3) and all are combined via equation (23) to produce a values of the components of $\hat{c}_{L\gamma M\alpha}$ at a point in the interior of the element under consideration.

A single element integral of equation (24) is computed by first subdividing the element into $m_1 \times m_2 \times m_3$ smaller sub-elements and sampling the $\hat{c}_{L\gamma M\alpha}$ terms at the centroid of each of sub-domain. For a small enough domain, it may be assumed that the $\hat{c}_{L\gamma M\alpha}$ terms are approximately constant, which leads to the approximation

$$\int_\Omega \hat{c}_{L\gamma M\alpha} g^\mu_{\alpha,M} g^\mu_{\gamma,L} d\Omega \approx \quad (28)$$
$$\sum_i^{m_1}\sum_j^{m_2}\sum_k^{m_3} \hat{c}_{L\gamma M\alpha}\left(\eta_1^{(i)}, \eta_2^{(j)}, \eta_3^{(k)}\right) \int_\Omega g^\mu_{\alpha,M} g^\mu_{\gamma,L} d\Omega$$

where $(\eta_1^{(i)}, \eta_2^{(j)}, \eta_3^{(k)})$ denotes the centroid of the (i^{th}, j^{th}, k^{th}) sub-element in the domain under consideration. The integral $\int_\Omega g^\mu_{\alpha,M} g^\mu_{\gamma,L} d\Omega$ in equation (28) is evaluated exactly using the forms given by equations (16) and (18). The resulting formulas are very long and cumbersome and are therefore not listed here.

Algorithm Verification

The capabilities of the CAA tool been verified by performing acceleration sensitivity analyses on three reference problems and then comparing our calculated results with the experimental or theoretical results obtained in the references. The referenced examples could all be modeled as 2-dimensional problems. Two examples were plane stress problems, while the third example represented a plane strain problem.

The first reference problem analyzed was a classic force-frequency experiment carried out using a circular ST-cut substrate which was 31.75 mm in diameter. A 60 MHz SAW device was centered on the substrate which had an active area that was approximately 1.0 mm by 1.0 mm [4]. In this example a diametrical in-plane force was placed on the substrate and the change in resonant frequency of the SAW device was measured as a function of the azimuthal angle of the force with respect to the propagation direction of the SAW. Symmetry was used to reduce the size of the finite element mesh by a factor of two because only half of the disc needed to be modeled to accurately define the stresses generated near the active region of the device. In all cases, a convergence study was performed on the mesh to determine the optimal configuration for modeling the mechanical system. The finite element mesh at which convergence occurred was comprised of 432 quadratic serendipity quadrilaterals with a total of 1347 nodes. A representation of the mesh can be seen in Figure 4. Figure 5 shows a comparison of the results published by Sinha, et al., with those produced by the analysis tool. The percentage of error between the calculated and experimental results ranged from 9% to 23%. The primary source of error in modeling this example was the inability to accurately model a point force using the serendipity quadrilateral.

The second reference problem modeled the in-plane acceleration sensitivity of a SAW resonator on a rectangular substrate that was rigidly supported on all edges [5]. For this problem, the acceleration-induced frequency change had been theoretically predicted by Shick, et al. The variable in this experiment is the aspect ratio of the substrate. These devices operated at approximately 400 MHz and the active area was 1.0 mm. The initial substrate was a 19.05 mm square. Symmetry reduced the complexity of the finite element mesh by a factor of four, because only a quarter of the plate needed to be modeled to accurately portray the mechanical stress biasing the active area of the device. The finite element mesh of the quarter plate contained 300 quadratic serendipity quadrilaterals and contained 981 nodes. A model of the mesh for the entire plate can be seen in Figure 6. Beginning with a square substrate, aspect ratio (a/b) of 1:1, where 2a is the length of the substrate and 2b is the width, seven in-plane acceleration sensitivities were calculated with the aspect ratio of the substrate varying from 1:1 to 4:1. The results of this analysis are in Figure 7. Comparing the calculated results from the analysis tool with Shick's theoretical data at an aspect ratio of 1:1 the calculated data differed by 5%, while at 2:1 the calculated data differed by 17%. The difference in the 2:1 result occurred because Shick modeled a larger substrate in his

analysis (25.4 x 12.7, instead of 19.05 x 9.525).

The third example is a plane strain problem. In this experiment by Andres and Parker, it was shown that by properly massloading the backside of an all quartz package (AQP) SAW substrate that the bending stresses caused by a normal acceleration could be minimized [6]. Modeling this situation meant that the dimensions of concern were the length and thickness of the substrate which were 10.16 mm and 0.89 mm respectively, the length of the active SAW area was 0.5 mm. Since no dimensions were given for the masses, they were assumed to be 1.2 mm in length and three times as dense as quartz. The finite element mesh of the SAW substrate, masses, and glass frit can be seen in Figure 8. Symmetry reduced the complexity of the analyzed model by a factor of two. The mesh contained 270 quadratic serendipity quadrilaterals which had 929 nodes. The experiment was carried out as follows, two masses were attached to one of four symmetric positions on backside of the AQP SAW substrate and a normal acceleration was applied to the SAW device. The acceleration induced frequency shift was measured in the output of the oscillator under test and the normal vibration sensitivity was calculated. Once this process was completed, the masses were detached, repositioned in another of the sites and the process was repeated. Figure 9 shows a comparison between the experimental data and the calculated results. Although the quantitative results were not accurate because of insufficient data to model the physical problem, qualitatively the analysis of the problem yielded similar results with regard to the optimal placement of the weights to minimize acceleration sensitivity.

Conclusion

A quantitative analysis tool has been developed and demonstrated which calculates the acceleration sensitivity of a SAW traveling in an arbitrary piezoelectric substrate. This computer aided analysis tool performs four major functions to determine the acceleration sensitivity for any given problem. First, it generates a finite element mesh which accurately models the physical problem. Next, the static biasing problem is solved by calculating the nodal displacements caused by the forces applied to the mechanical system. The third function that the analysis tool performs is a calculation of the SAW mode shape based on Sinha and Tiersten's analytical technique. Finally, data from the previous steps are combine with the spatially varying effective elastic constants to calculate the acceleration of the SAW device.

One of the more interesting aspects of this analysis tool is that it allows the user to sample the stresses, strains, and rotations within any given element and then use that data in sophisticated integration schemes to improve model accuracy. The software has been proliferated on several different platforms including, a Cray supercomputer, a Sun workstation and a 486 laptop PC, demonstrating its flexibility.

The CAA tool achieved good agreement (5%-10%) with the experimental and theoretical examples where an accurate model of the physical example could be generated. All of the examples analyzed in this study were 2-D problems modeled on a 486 laptop PC. The CAA tool can also evaluate 3-dimensional problems, however, a workstation or larger CPU would be required.

References

[1] S. Richie, B. Abbott and D. Malocha, "Description and Development of a SAW Filter CAD System," IEEE Trans. MTT, Vol. 36, pp. 456-465, 1988.

[2] D. Malocha, S. Richie and B. Abbott, "Methods of Computer Aided Analysis and Design for Surface Acoustic Wave Devices," IEEE Ultrasonics Symposium Proceedings, 1988, pp. 11-21.

[3] O. Schwelb, E. Adler and J. Slaboszewicz, "Modeling, Simulation and Design of SAW Grating Filters," IEEE Trans. UFFC, Vol. 37, pp. 205-214, May 1990.

[4] B. Sinha, W. Tanski, T. Lukaszek and A. Ballato, "Influence of Biasing Stresses on the Propagation of Surface Waves," J. Appl. Phys., Vol. 57, No. 3, pp. 767-776, Feb. 1983.

[5] D. Shick, Y. Zhou and H.F. Tiersten, "An Analysis of the In-Plane Acceleration Sensitivity of Quartz Surface-Wave Resonators Rigidly Supported Along the Edges," J. Appl. Phys., Vol. 65, No. 1, pp 35-40, Jan. 1989.

[6] D. Andres and T. Parker, "Designing Smaller SAW Oscillators for Low Vibration Sensitivity," Proceedings of the 1994 IEEE International Frequency Control Symposium, June 1994, pp. 352-358.

[7] T. J. R. Hughes, The Finite Element Method, Linear Static and Dynamic Analysis, Prentice Hall, 1987.

[8] O.C. Zienkiewicz and R.L. Taylor, The Finite Element Method, Vols. 1 &2, McGraw-Hill, 1989.

[9] J.C. Baumhauer and H.F. Tiersten, "Nonlinear Electroelastic Equations for Small Fields Superposed on a Bias", J. Acoust. Soc. Am., Vol. 54, No. 4, 1973, pp. 1017-

1034.

[10] H. F. Tiersten, "Perturbation Theory For Linear Electroelastic Equations for Small Fields Superposed On a Bias", J. Acoust. Soc. Am., Vol. 64, No. 3, Sept., 1978, pp. 832-837.

[11] B.K. Sinha and H.F. Tiersten, "Elastic and Piezoelectric Surface Waves Guided By Thin Films", J. Appl. Phys., Vol. 44, No. 11, Nov. 1973, pp 4831-4854.

[12] B.K. Sinha and H.F. Tiersten, "Variational Analysis of the Reflection of Surface Waves by Arrays of Reflecting Strips", J. Appl. Phys., Vol. 47, No. 7, July 1976, pp 2824-2832.

[13] B.K. Sinha and H.F. Tiersten, "An Analysis of Transverse Modes in Acoustic Wave Resonators", J. Appl. Phys., Vol. 51, No. 6, June 1980, pp 3099-3112.

[14] H.F. Tiersten, J.T. Song, and D.V. Shick, "A Continuous Representation of the Acoustic Surface Wave Mode Shape in Arrays of Reflecting Grooves", J. Appl. Phys., Vol. 62, No. 4, August, 1987, pp 1154-1161.

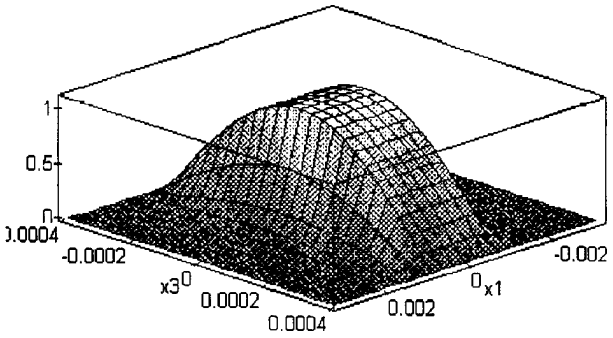


Figure 1. Amplitude envelope for first transverse mode in SAW transmission path and reflecting arrays

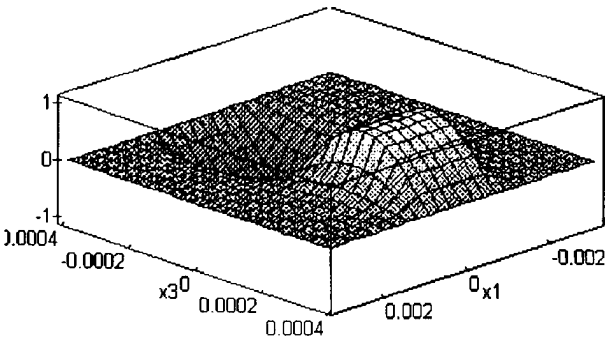


Figure 2. Amplitude envelope for second transverse mode in SAW transmission path and reflecting arrays

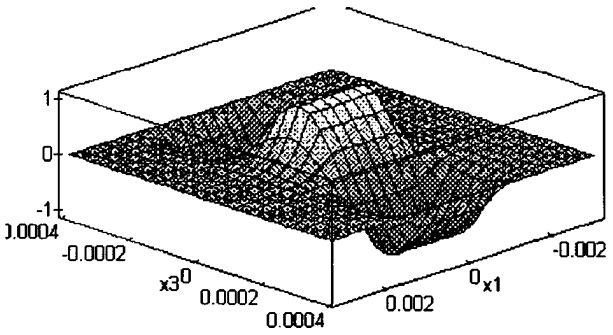
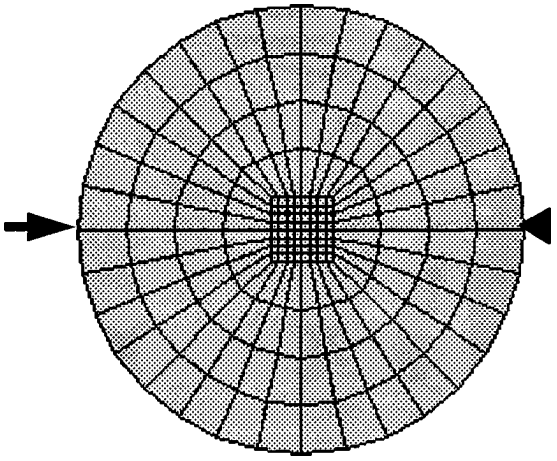
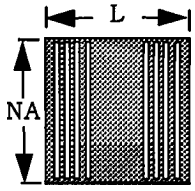


Figure 3. Amplitude envelope for third transverse mode in SAW transmission path and reflecting arrays



Finite Element Mesh of Resonator
on Circular Disk:
# of Elements - 432
# of Nodes - 1347

Substrate Diameter - 31.75 mm



SAW Transducers:
F_o -60 MHz
λ -52.63 um
# of electrodes/transduce
NA - 1.0 mm
L - 1.0 mm

Figure 4. Example problem for force-frequency experiment.

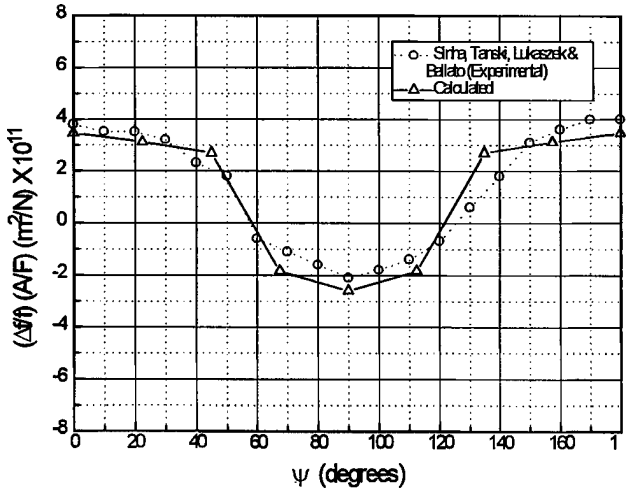
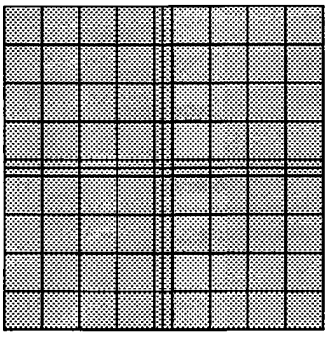
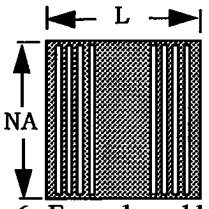


Figure 5. Comparison of calculated and experimental results for force-frequency experiment.



Finite Element Mesh of Resonator
on Square Plate:
# of Elements - 300
# of Nodes - 981



SAW Transducers:
F_o - 401.5 MHz
λ - 7.846 um
# of electrodes/transduce
NA - 1.0 mm
L - 1.0 mm

Figure 6. Example problem for in-plane acceleration using plane stress analysis.

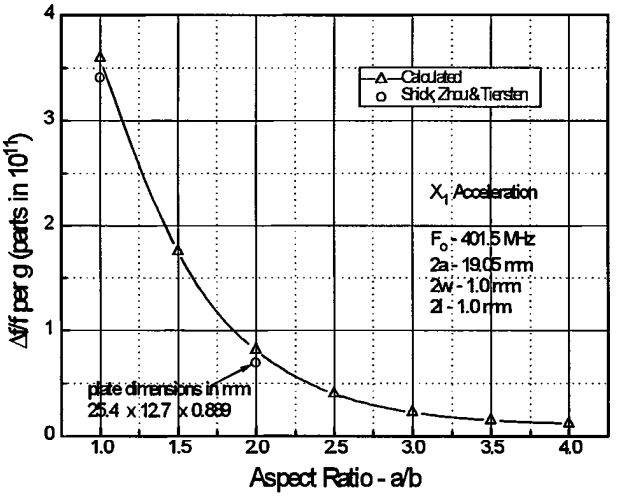
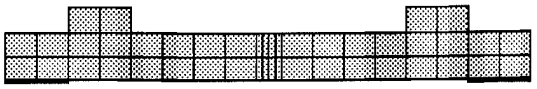


Figure 7. Comparison of calculated and analytical results for in-plane acceleration plane stress problem.



Finite element mesh of SAW substrate,
frit & masses:
# of Elements - 270
# of Nodes - 929

Substrate Length - 10.16 mm & Thickness - 0.89

SAW Transducers:
F_o - 900 MHz
λ - 3.51 um
Length - 0.5 mm

Figure 8. Example problem for the mass on substrate experiment using plane strain analysis.

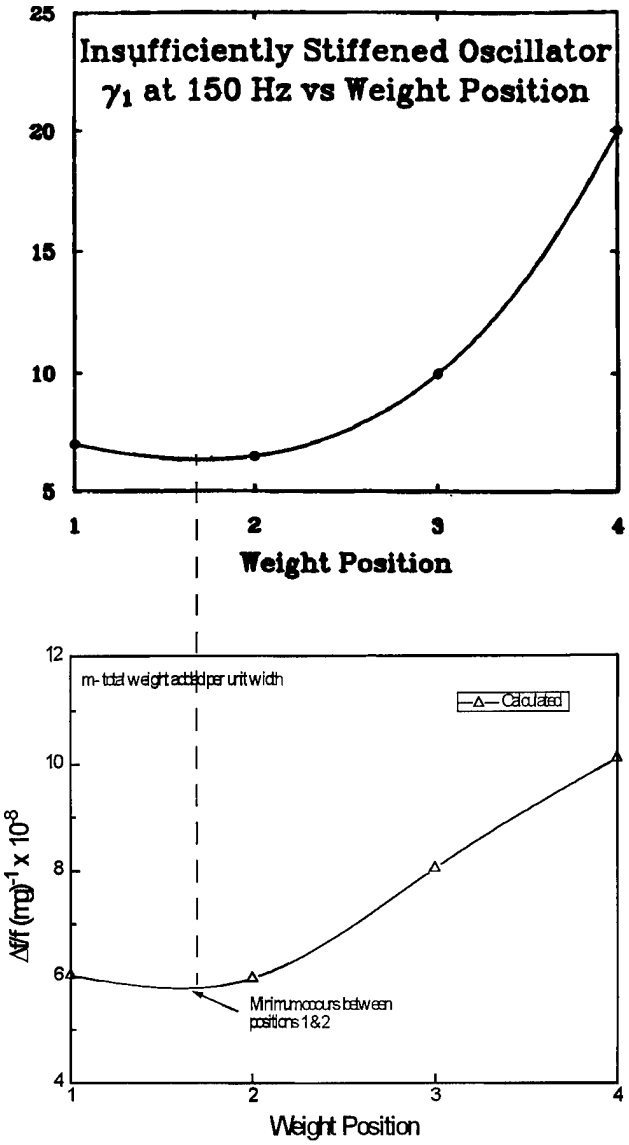


Figure 9. Comparison of calculated and experimental results for the mass on substrate experiment [6].

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

AN ANALYSIS OF THE DYNAMIC BEHAVIOR AND ACCELERATION SENSITIVITY OF A SAW RESONATOR SUPPORTED BY FLEXIBLE BEAMS

J. T. Stewart, J. A. Kosinski, and A. Ballato

US Army Research Laboratory, ATTN: AMSRL-PS-ED, Fort Monmouth, NJ 07703-5601

Abstract: A numerical analysis is performed of the normal acceleration sensitivity and the low frequency dynamic vibration response of a SAW resonator supported by flexible beams of various geometries. The acceleration sensitivity problem is solved using Tiersten's perturbation theory for small fields superposed on a static bias in combination with a very accurate solution to the general two dimensional surface wave equations of Sinha and Tiersten. The static biasing states under normal acceleration are obtained using three dimensional finite element models which are generated and solved using recently developed FEA software. The frequency shift under normal acceleration is computed using a rigorous and precise integration algorithm which allows an analytical surface wave mode shape to be combined with a discrete finite element solution to obtain the first perturbation of the eigen-value. A modal analysis of the SAW substrate and support beam assembly is also performed using the FEA software.

Introduction

The normal acceleration sensitivity and low frequency dynamic behavior of a surface acoustic wave resonator supported by flexible beams is numerically analyzed using recently developed semi analytical finite element tools [7]. The general perturbation theory of Tiersten [1,2] is used to calculate the frequency shift under normal acceleration when the biasing state is solved using the finite element method and the surface acoustic wave mode shape is solved analytically using the two dimensional theory of Sinha and Tiersten [3,4,5]. The theoretical development of the finite element solution method and the integration algorithm are given in reference [7] and are not listed here. The modeling software is used to study problems of three different geometries and it is shown that each of these configurations leads to different observed behavior.

Numerical Results

The first example considered is that of a rectangular ST-Cut quartz plate with four symmetrically placed beam supports, also ST-Cut quartz, along the edges parallel to the x_1 axis. A 360 MHz SAW resonator of the dimensions and characteristics indicated in reference [6] is placed in the center of the plate and the entire assembly is subjected to a unit normal acceleration ($a_2 = 1g$, $a_1 = a_3 = 0$). Figure 1 shows the general geometry of the structure with the basic design variables labeled. Figure 2 shows a plotting of the finite element model used for this study with numerical values for the fixed design variables given. The calculations are carried out for various beam lengths and substrate thicknesses to obtain a detailed picture of the static acceleration performance of this particular structure. Since the SAW mode resides primarily in the center of the substrate, it is appropriate to assume that the deformations in this area under gravity induced bending are the important quantities to observe and manipulate during the design process. Figure 3 shows a three dimensional distortion plot of the structure under normal acceleration and shown in Figures 4 and 5 are curves which represent the deflection of the middle plane of the plate along the x_1 and x_3 axes, respectively. The deflection plots of Figures 4 and 5 are shown for a plate of thickness 0.508 mm with beam lengths of 2, 4, and 6 mm., and the "slices" are taken along the coordinate axes, which conveniently coincide with the center-lines of the plate. In Figure 5, it is observed that for each beam length the curvature in the center of the plate increases slightly, while in Figure 4 it is seen that as the beam length is increased, the curvature at the middle point decreases, tending to flatten along this direction. This suggests a shift in the distribution of biasing stresses and strains which can be controlled by the lengths of the beams. It is observed for this particular case that an apparent zero acceleration sensitivity is achieved at a beam length of approximately 3.5 mm. Figure 6 shows the

U.S. Government work not protected by U.S. copyright

computed acceleration sensitivity for this structure for plate thicknesses varying from 0.254 mm to 1.016 mm and beam lengths varying from 1 mm to 8 mm. It is seen that for each thickness an optimal point exists where the normal acceleration sensitivity vanishes.

The second example considers a rectangular plate supported at its corners by flexible beams. Figure 7 shows the general geometrical layout of the structure and Figure 8 shows the model used for the calculations with pertinent dimensions listed. As before, ST-Cut quartz is used throughout and the same 360 MHz SAW resonator [6] is assumed. Figure 9 shows the general three dimensional deformation of a typical structure and Figures 10 and 11 show the deflections of the middle plane along the coordinate axes. These deflection plots are shown for a beam angle of $\theta = 45°$ and for planar aspect ratios of $L_p/W_p = 0.5$, $L_p/W_p = 0.7$, and $L_p/W_p = 1.0$. Once again it is observed that while the x_3 flexure characteristics are somewhat constant, the x_1 behavior exhibits some variation for the different dimensions. As shown in figure 11, the curvature of the plate along x_1 is almost flat at $L_p/W_p = 0.5$ and increases with increasing L_p/W_p. It is observed that for this structure a point of aspect ratio compensation occurs at approximately $L_p/W_p = 0.56$. Figure 12 shows the results of normal acceleration sensitivity calculations for beam angles ranging from 20° to 70° and for planar aspect ratios from 0.3 to 1.8. It is observed that for each beam angle, there is one point of aspect ratio compensation. It is also observed that away from these optimal points, the performance of this structure degrades rapidly with nominal values in parts in 10^8 and below.

The third example considered is that of a rectangular ST-Cut quartz plate supported by three beams along each of two edges parallel to the x_1 axes. The same 360 MHz resonator [6] is again used and is centered with respect to the plate's major edges. Figure 13 shows the general description of the geometry and Figure 14 shows the details of the model used. The general three dimensional deformation of this structure is shown in Figure 15 and the deflections along the center-lines of the plate are shown in Figures 16 and 17. In this study, the beam separation distance is varied along with the planar aspect ratio. Figure 18 shows the normal acceleration sensitivity verses the beam separation distance for planar aspect ratios of $L_p/W_p = 0.5$, $L_p/W_p = 1.0$, and $L_p/W_p = 1.5$. It is interesting to note that in this case, no optimal point is noticed and the performance is generally poor with nominal values in parts in 10^8 to 10^7.

The final example explores the dynamic behavior of the plate-beam structure. Central to the design of the types of structures considered here is the calculation of the fundamental resonance of the overall device. This package resonance defines the limiting dynamic range that the system may operate in. To study the dynamic behavior of these plate structures and package assemblies, an efficient eigen-value solver has been incorporated into the finite element code. Using this eigen-solver, a modal analysis has been carried out for the three-beam model for a planar aspect ratio of 1.0. Figure 19 shows the first three vibration modes of the 3-beam supported structure, and Figure 20 shows a parametric study of beam length verses plate thickness for the fundamental resonance. It has been found that these resonant frequencies are most sensitive to these parameters. Equivalent studies on the other geometries yield similar results. It is observed that for the most desirable beam lengths, i.e. 3.0 mm to 5.0 mm, the fundamental resonant frequency is reduced by approximately one half, generally reaching a value of about 8 KHz to 10 KHz.

Conclusion

The numerical experiments carried out on the beam supported SAW resonator structures reveal some very interesting results. From the first two examples it can be concluded that by varying the geometry and the flexibility of the larger supporting structure it is possible to achieve a theoretical zero normal acceleration sensitivity. This is achieved by proportioning critical dimensions to manipulate the biasing stresses and strains. It can also be concluded that when such precise dimensioning can not be achieved, the performance will generally be poor. The results obtained for the optimal values are dimensionally realistic, as all calculations were carried out for standard wafer thicknesses and the dimensions used are in proportion with typical SAW resonator dimensions. The dynamic performance of these structures appears to be compatible with many consumer applications and some military applications, where higher frequency noise is not present.

References

[1] J.C. Baumhauer and H.F. Tiersten, "Nonlinear Electroelastic Equations for Small Fields Superposed on a Bias", J. Acoust. Soc. Am., Vol. 54, No. 4, 1973, pp. 1017-1034.

[2] H. F. Tiersten, "Perturbation Theory For Linear Electroelastic Equations for Small Fields Superposed On a Bias", J. Acoust. Soc. Am., Vol. 64, No. 3, Sept., 1978, pp. 832-837.

[3] B.K. Sinha and H.F. Tiersten, "Elastic and Piezoelectric Surface Waves Guided By Thin Films", J. Appl. Phys., Vol. 44, No. 11, Nov. 1973, pp 4831-4854.

[4] B.K. Sinha and H.F. Tiersten, "Variational Analysis of the Reflection of Surface Waves by Arrays of Reflecting Strips", J. Appl. Phys., Vol. 47, No. 7, July 1976, pp 2824-2832.

[5] B.K. Sinha and H.F. Tiersten, "An Analysis of Transverse Modes in Acoustic Wave Resonators", J. Appl. Phys., Vol. 51, No. 6, June 1980, pp 3099-3112.

[6] D. Huynh, R. McGowan, J. A. Kosinski, J. T. Stewart, R. Piekarz, and C. D. Mulford, "Experimental Studies of SAW and STW Acceleration Sensitivity", elsewhere in these Proceedings.

[7] J. T. Stewart, R. McGowan, J. A. Kosinski, and A. Ballato, "Semi-Analytical Finite Element Analysis of Acceleration-Induced Frequency Shifts in SAW Resonators", elsewhere in these Proceedings.

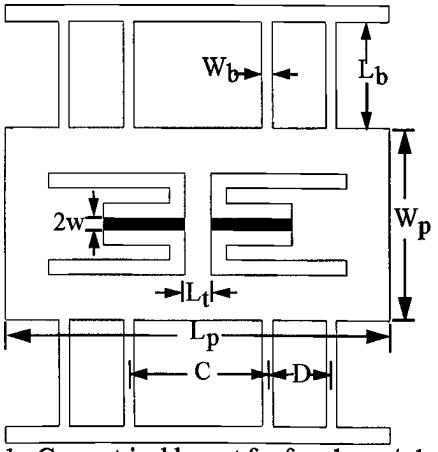


Figure 1. Geometrical layout for four beam/edge configuration

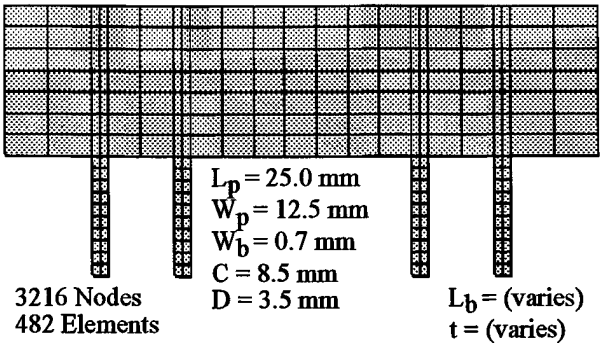


Figure 2. Finite element mesh and dimensions for the four beam/edge configuration.

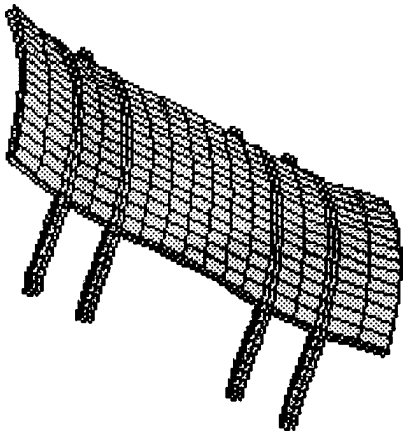


Figure 3. Three dimensional deformation plot of the four beam/edge plate under normal acceleration.

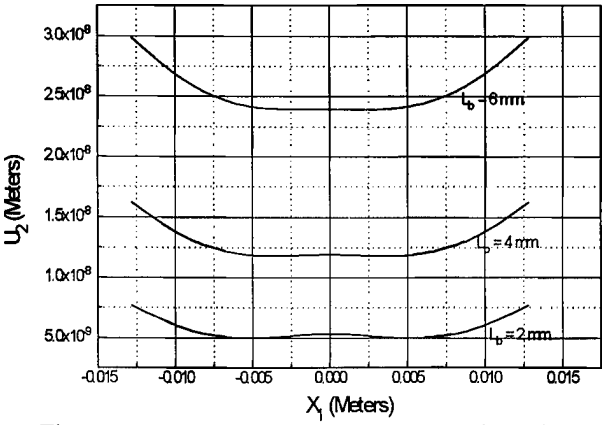


Figure 4. Flexural deformation of the middle plane along the x_1 direction for four beam/edge configuration.

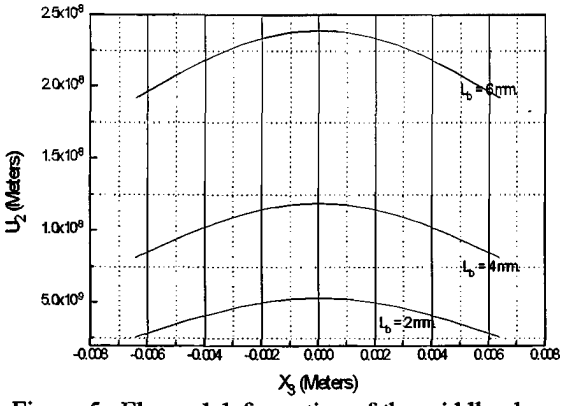


Figure 5. Flexural deformation of the middle plane along the x_3 direction for four beam/edge configuration.

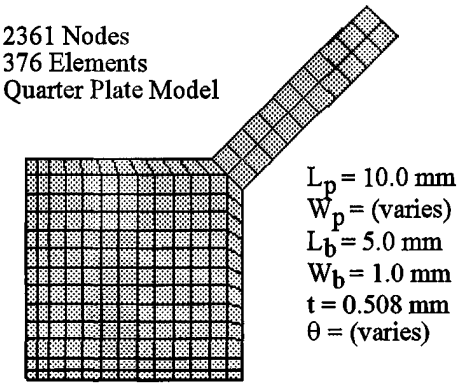


Figure 8. Finite element mesh and dimensions for the corner supported configuration.

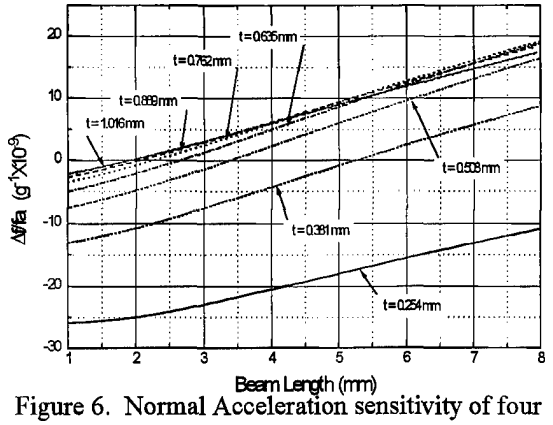


Figure 6. Normal Acceleration sensitivity of four beam/edge configuration as function of plate thickness and support beam length.

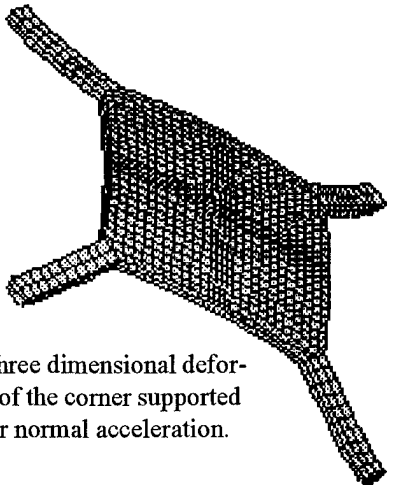


Figure 9. Three dimensional deformation plot of the corner supported plate under normal acceleration.

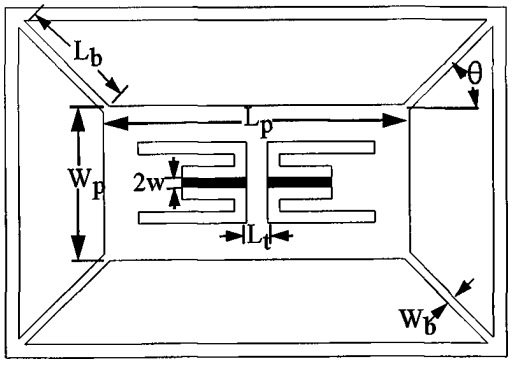


Figure 7. Geometrical layout for corner supported configuration

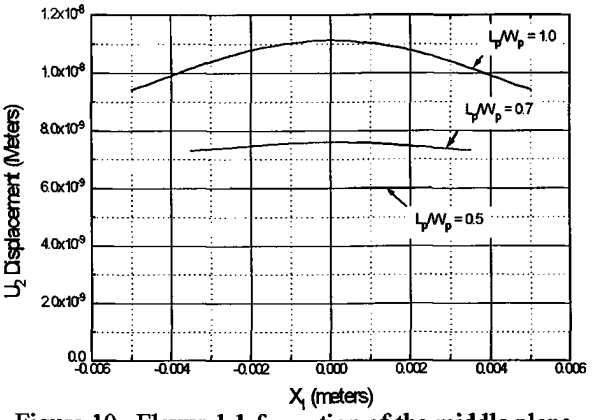


Figure 10. Flexural deformation of the middle plane along the x_1 direction for the corner supported configuration

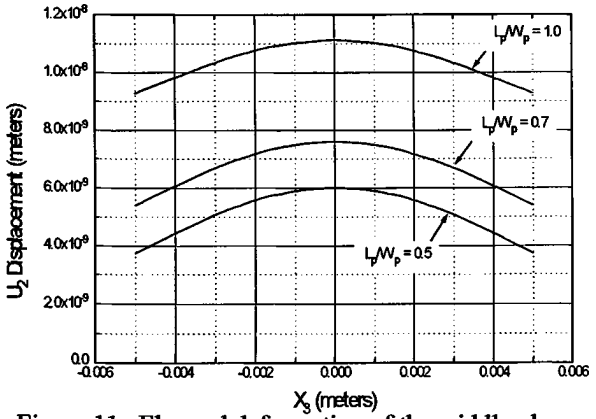


Figure 11. Flexural deformation of the middle plane along the x_3 direction for the corner supported configuration.

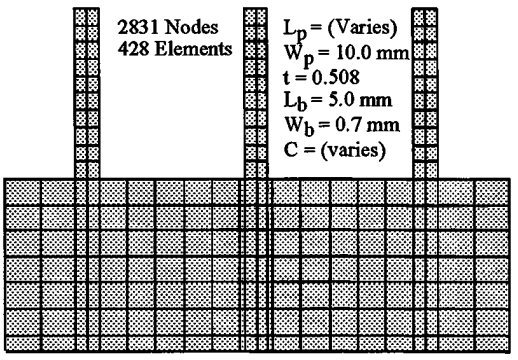


Figure 14. Finite element mesh and dimensions for the three beam/edge configuration.

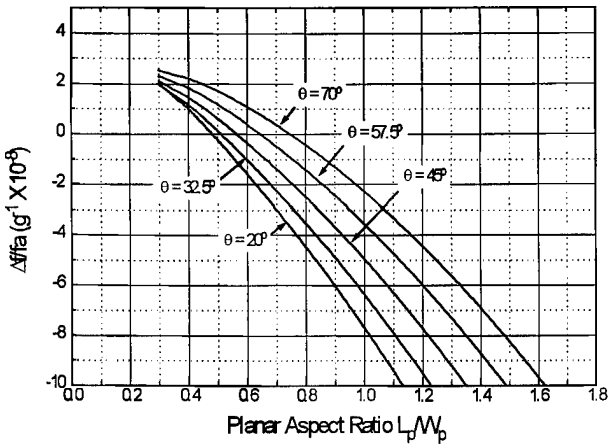


Figure 12.

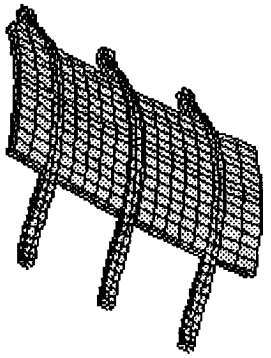


Figure 15. Three dimensional deformation for the three beam/edge configuration under normal acceleration.

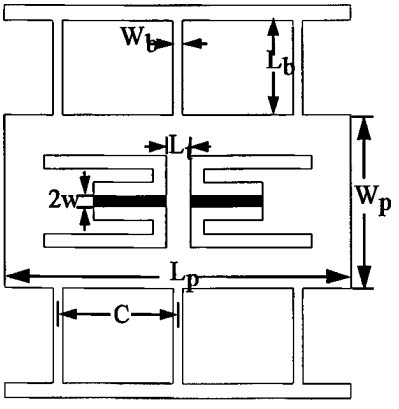


Figure 13. Geometrical layout for three beam/edge configuration

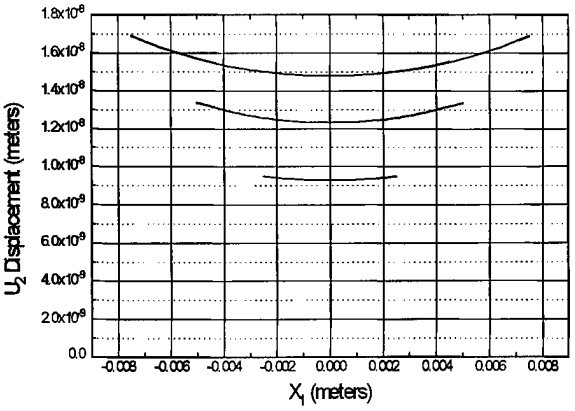


Figure 16. Flexural deformation of the middle plane along the x_1 direction for three beam/edge configuration

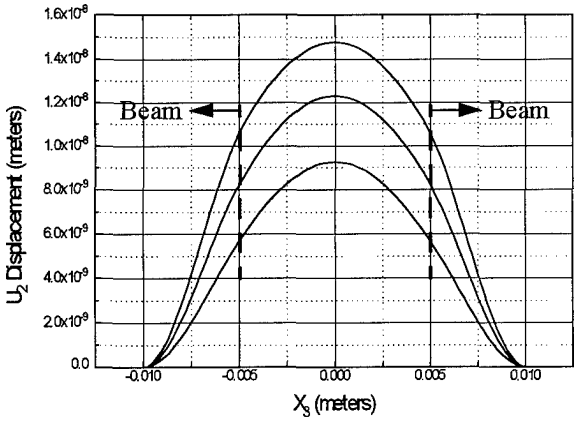


Figure 17. Flexural deformation of the middle plane along the x_3 direction for four beam/edge configuration

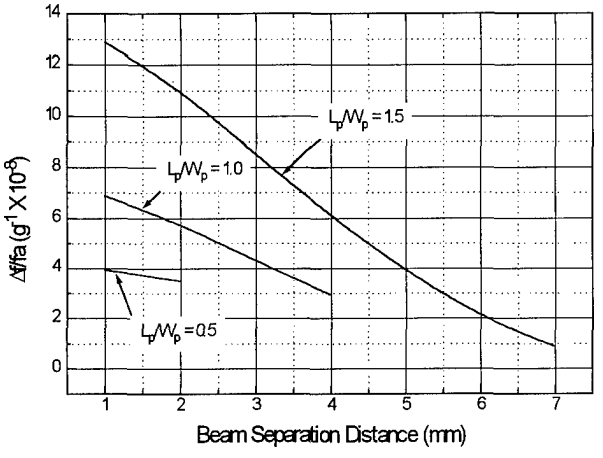


Figure 18.

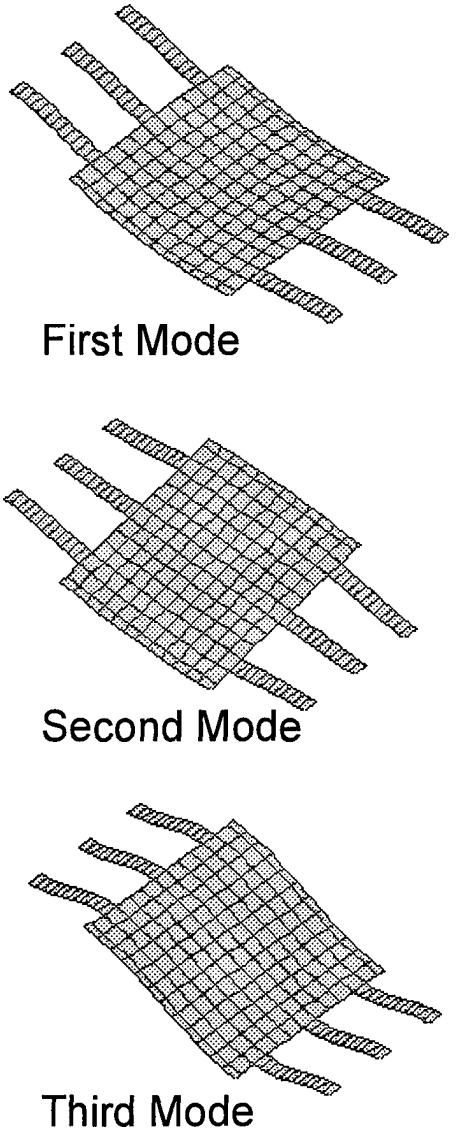


Figure 19. First three modes of the 3-Beam configuration.

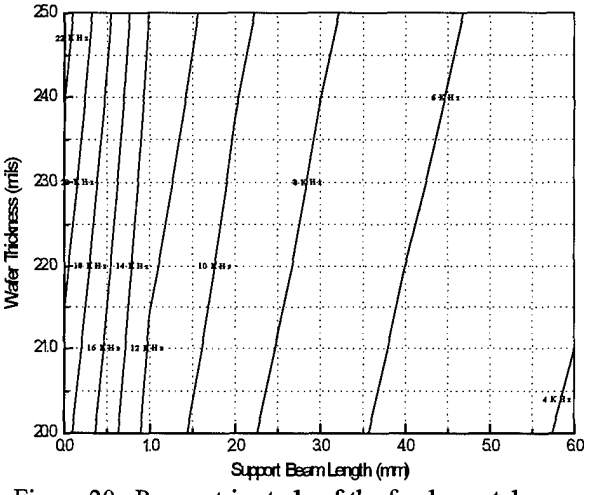


Figure 20. Parametric study of the fundamental resonance of the 3-Beam plate as a function of beam length and thickness.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A SAWR Oscillator Vibration Sensitivity and Phase Noise Reduction Technique Using Multiple Resonators and RF Outputs

M. M. Driscoll

Westinghouse Electronic Systems Group, Baltimore, MD 21203

Abstract

A technique previously described by the author for reduction of quartz crystal oscillator vibration sensitivity and phase noise has been modified and applied to the design of a UHF, SAWR oscillator. Use of four SAW resonators mounted in an in-plane orientation provides 3-axis vibration sensitivity cancellation. Because of variations in individual SAWR vibration sensitivity vector magnitude and direction, the cancellation is non-exact. However, a 14:1 reduction in vibration sensitivity has been achieved using resonators having fairly uniform, individual sensitivities in the range $44X10^{-9}$ to $55X10^{-9}$ per g. The novel oscillator loop circuitry consists of cascaded stages of SAWR/modular amplifier/power divider combinations, providing multiple, per-stage RF output signals whose noise phase floors are uncorrelated as a result of individual SAW resonator frequency selectivity. Thus, output signal amplitude (power combiner) or frequency (mixer) summation of M oscillator outputs yields a 10logM noise floor improvement. A minimum, 10logN improvement in flicker-of-frequency noise is obtained via use of N resonators due to the uncorrelated nature of the individual resonator frequency instabilities and N times increase in oscillator loop group delay. Phase noise floor levels on the order of -185dBc/Hz and flicker-of-frequency noise levels given by £(100Hz)=-124dBc/Hz have been demonstrated in an oscillator using 320MHz SAW resonators manufactured by SAWTEK, Inc., using the technique (with M=N=4), which is also applicable to oscillators incorporating other types of resonators.

Introduction

A comparison between quartz crystal bulk acoustic wave resonator (BAWR) and surface acoustic wave resonator (SAWR) oscillator performance would show that, in general, SAWR oscillators operate at higher frequency and exhibit lower phase noise floor level. Quartz BAWR oscillators, on the other hand, generally exhibit lower near-carrier (i.e., flicker-of-frequency) noise level and lower frequency sensitivity to vibration. The oscillator performance differences are directly related to corresponding differences in resonator characteristics. With regard to SAWR oscillators, low phase noise floor level results from relatively higher SAWR drive level capability [1-2]. Non-uniform and, in general, higher SAWR vibration sensitivity appears to be primarily associated with the design and control of the resonator mounting structure and materials [3]. Exceptionally low sensitivities have been reported for specific SAWR designs [4].

Resonator loaded Q (i.e., transmission response group delay) determines the degree to which oscillator sustaining stage open loop phase noise is converted to closed loop frequency noise [5]. The frequency-Q product for both quartz BAWR and SAWR Q is fairly constant. As a result, the Q factors exhibited by UHF SAW resonators are proportionally lower than those of HF and VHF BAW resonators. Higher relative SAWR oscillator signal flicker-of-frequency or near-carrier noise level can result from either: (1) relatively higher resonator short-term frequency instability and/or lower resonator Q factor (increased conversion of sustaining stage phase instability to oscillator signal frequency instability) [6].

In 1992, the author described a patented technique for improving quartz crystal oscillator vibration sensitivity, near-carrier noise level, and noise floor level through the use of multiple, series connected resonators [7,8]. The cited technique used four, series-connected, 80MHz, 3rd overtone, SC-cut resonators having a net, effective Q identical to that of a single resonator. The resonators themselves consituted the primary contributor to oscillator output signal flicker-of-frequency noise, and a single sustaining stage amplifier was used to provide 400mW total output signal power, 30mW of which was dissipated in the four resonators. This paper descibes use of a similar technique (with notable differences) applied to the design of a UHF SAWR oscillator.

0-7803-2500-1/95 $4.00 © 1995 IEEE

SAWR Oscillator Configuration

Basic Oscillator Block Diagram

Figure 1 shows the block diagram for a SAWR oscillator conceived to exhibit improved flicker-of-frequency noise and noise floor performance as well as reduced frequency sensitivity to vibration. As shown in the figure, the oscillator circuit consists of four, identical, cascaded stages of SAWR/modular amplifier/power divider combinations. The simpler use of four, series-connected SAWRs was not used for several reasons: (1) use of individual SAWR drive levels on the order of 100mW each (400mW total) in order to minimize signal noise floor level would have required use of a sustaining stage amplifier exhibiting exceptionally high compression point performance; (2) use of the cascaded stage approach results in four times higher oscillator loop group delay (four times less conversion of net amplifier phase-to-frequency noise); and (3) use of the figure 1 configuration allows extraction of four, separate RF outputs. The white phase noise (floor) component of each of the outputs is uncorrelated, due to the frequency selectivity of each (per-stage) SAWR. Thus, vector amplitude summation (using power combiners) or frequency summation (using mixers) of each RF output provides a means of obtaining an additional 6dB reduction in combined, output signal noise floor level. In addition, if each output feeds identical load circuits (i.e., frequency multipliers), then signal summation at the load circuit outputs results in 6dB reduction in effective load circuit additive phase noise as well due to the uncorrelated nature of each load circuit's additive noise spectrum. Unlike the aforementioned use of multiple, series-connected resonators, the figure 1 technique requires a reasonable degree of frequency tracking between individual SAWRs. For example, an allowable phase shift change between stages on the order of 0.1 radians would correspond to a SAWR tracking requirement (for the 320MHz resonators) of $\delta f/f$=5ppm. This implies use of ovenized, low aging rate resonators. The degree of allowable mistrack is related to resonator group delay and is larger for higher frequency (lower delay) devices. For example, three times higher (15ppm) mistrack would be allowed between 1GHz resonators typically exhibiting ten times lower group delay (three times lower Q), compared to the 320MHz devices [9].

As shown in figure 1, the input and output IDT capacitance of each resonator has been anti-resonated using shunt inductance, and individual SAWR frequency set-on is accomplished using an appropriate value of series reactance. Figure 2 shows the results of automatic network analyzer measurement of oscillator loop transmission response for 1, 2, 3, and all 4 cascaded, resonator/amplifier/power divider stages. The per-stage and total (4 stage) small signal excess gain was on the order of 1.5dB and 6dB, respectively. Similarly, figure 3 shows the results of measurement of cascaded stage group delay, with the per-stage delay on the order of 8.5usec and the total (4 stage) delay=34usec. An appropriate length of coaxial cable was used to connect the oscillator positive feedback loop in order to set $2N\pi$ radians loop phase shift at the oscillator operating frequency.

Resonator Orientation

The individual oscillator resonators were mounted in an in-plane orientation providing partial cancellation of vibration sensitivity in all three axes. The SAWR mounting orientation is shown in figure 4. The degree of vibration sensitivity reduction achieved depends on the uniformity of the vibration sensitivity vector (magnitude and direction) in each resonator. The overall, effective vibration sensitivity vector for the four SAWRs is the vector sum of 1/4th of each individual SAWR vibration sensitivity vector. (If 3 of the 4 SAWRs had zero magnitude vibration sensitivities, the overall vibration sensitivity of the four SAWR combination would be 1/4th that of the single, vibration-sensitive resonator.)

Measurement Results

Vibration Sensitivity Reduction

Table 1 shows the results of measurement of the vibration sensitivity of each of four, unscreened SAW resonators. The tests consisted of measurement of the vibration-induced, phase modulation sideband appearing on a carrier frequency signal passed through the two-port SAWR, mounted on the shake table in a passive, transmission test fixture. CW test tones of 50Hz, 100Hz, and 200Hz were used. The measurement results indicate that, for all four resonators, the vibration sensitivity vector direction was virtually orthogonal to the SAWR die surface, and the vector magnitude was fairly uniform. The results of use of the four resonators in the oscillator oriented as shown in figure 4 are also shown in the table, and indicates a significant reduction in net vibration sensitivity has been achieved. For the four resonator measurement, vibration-induced, carrier signal FM sideband level was measured at the test tone frequencies for the closed loop oscillator.

Phase Noise Performance

Determination of SAWR oscillator signal flicker-of-frequency and white phase noise (floor) levels

were made at 640MHz via comparison with a low noise, reference signal generator. The SAWR oscillator/frequency doubler block diagram is shown in figure 5. As shown in the figure, pairs of SAWR oscillator outputs were amplitude-summed using zero degree, power combiners. Each of the two, 23dBm summed outputs were fed to (WJFD25HC) high level, diode frequency doublers. The two doubler outputs were then also amplitude-summed using a power combiner. Net 640MHz output signal power level was 14dBm.

The reference signal generator consists of a low noise 80MHz (four-crystal) oscillator plus times-eight frequency multiplier incorporating an intermediate stage, SAWR bandpass filter at 320MHz and a (300KHz bandwidth, 10dB insertion loss) ceramic dielectric, bandpass filter at 640MHz. The reference signal generator output power level is also on the order of 14dBm. The results of phase noise measurement of two, near-identical, reference signal generators is shown in figure 6. Figure 7 shows the level of the HP noise measurement system noise with the system quadrature phase detector operated from a common signal at (14dBm) actual measurement drive levels.

The results of phase noise measurements made via phaselock of the reference signal generator to the SAWR oscillator-doubler are shown in figure 8. Comparison of figures 6, 7, and 8 indicate that in the carrier offset frequency region ($fm \leq 100Hz$), the SAWR oscillator is the dominant source of flicker-of-frequency noise. At moderate carrier offset frequencies ($300KHz \geq fm \geq 100Hz$), the phase noise of the reference signal generator dominates. At high offset frequencies (outside the 300KHz bandwidth of the reference signal generator output filter), the measured phase noise floor level (-175dBc/Hz) is partially due to the SAWR oscillator-doubler (-179dBc/Hz) and partially due to the resolution of the HP Noise measurement system (-177dBc/Hz). The figure 8 results were obtained using SAW resonators individually screened for lowest flicker-of-frequency noise. Using unscreened (table 1) resonators, resulting oscillator-doubler signal flicker-of-frequency noise was 10dB poorer (-108dBc/Hz at fm=100Hz).

Based on the results depicted in figures 6 through 8, the phase noise of the 320MHz SAWR oscillator, prior to frequency doubling, can be described by a flicker-of-frequency component on the order of $Sy(f)=8X10^{-24}/f$ (-124dBc/Hz at 100Hz offset frequency). The effective phase noise floor level, including doublers (referred to the oscillator operating frequency) is on the order of -185dBc/Hz. This represents state-of-the-art SAWR oscillator signal spectral performance.

Conclusions

A technique for improving SAWR oscillator signal phase noise and vibration sensitivity via use of multiple resonators and RF outputs has been successfully demonstrated, achieving state-of-the-art spectral performance.

The technique is readily adaptable to configuration/performance tradeoff modifications such as number of cascaded stages and/or RF outputs used, as well as output signal summing methods.

Compared to the use of multiple, series-connected resonators, the use of the isolated cascade approach requires a higher degree of frequency tracking between resonators, especially for larger group delay (lower frequency) devices.

References

1. G. K. Montress, et al., "Extremely Low Phase Noise SAW Resonator Oscillator Design and Performance," Proc. 1987 IEEE Ultras. Symp., Oct. 1987, pp. 47-52.
2. I. D. Amramov, "Extremely High Loop Power GHz Range Surface Transverse Wave Oscillators Using AB-Class Amplifiers," Proc. 1993 IEEE Freq. Contr. Symp., June 1993, pp. 728-732.
3. J. Himmel, et al., "Market Survey of Acceleration-Insensitive SAW Oscillators," Proc. 1992 IEEE Freq. Contr. Symp., May 1992, pp. 849-860.
4. T. E. Parker, et al., "SAW Oscillators with Low Vibration Sensitivity," Proc. 1991 Freq. Contr. Symp., May 1991, pp. 321-329.
5. D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," Proc. IEEE, Vol. 54, No. 2, Feb. 1966, pp. 329-330.
6. T. E. Parker and D. Andres, "Dependence of SAW Resonator 1/f Noise on Device Size," Proc. 1993 IEEE Freq. Contr. Symp., June 1993, pp. 178-185.
7. M. M. Driscoll, "Reduction of Quartz Crystal Oscillator Flicker-of-Frequency and White Phase Noise (Floor) Levels and Acceleration Sensitivity via Use of Multiple Resonators," Proc. 1992 IEEE Freq. Contr. Symp., May 1992, pp. 334-339.
8. U. S. Patent Numbers 4,851,790 (July 25, 1989) and 5,250,871 (Oct. 5, 1993)
9. I. D. Avramov, et al., "Surface Transverse Wave Oscillator with Extremely Low Thermal Noise Floors," Proc. 1994 IEEE Freq. Contr. Symp., June 1994.

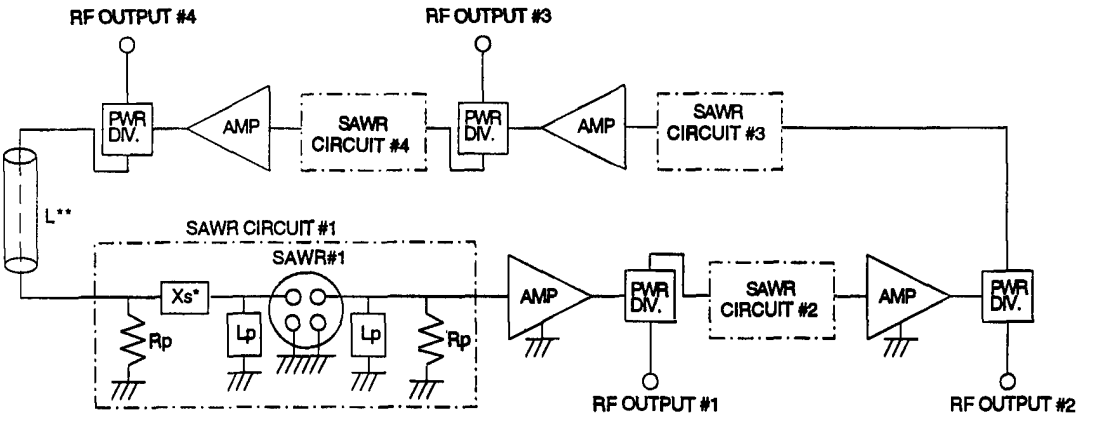


Figure 1. Multiple SAWR Oscillator Block Diagram

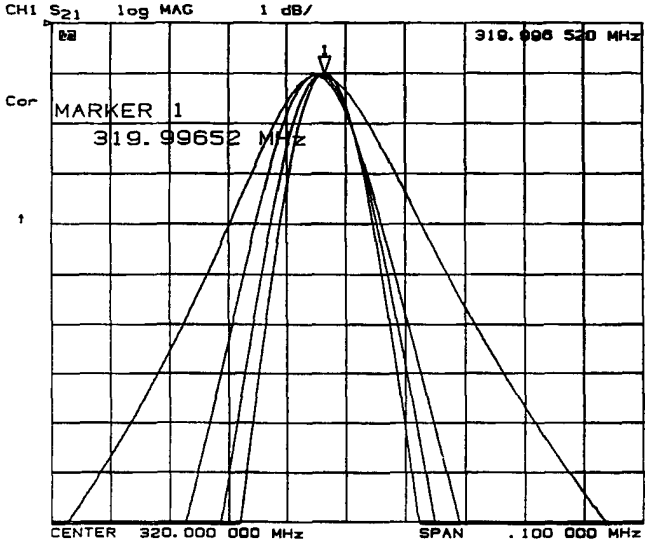


Figure 2. Measured Oscillator Open Loop Transmission Response for 1, 2, 3, and 4 Cascaded SAWR/Amplifier/Power Divider Stages

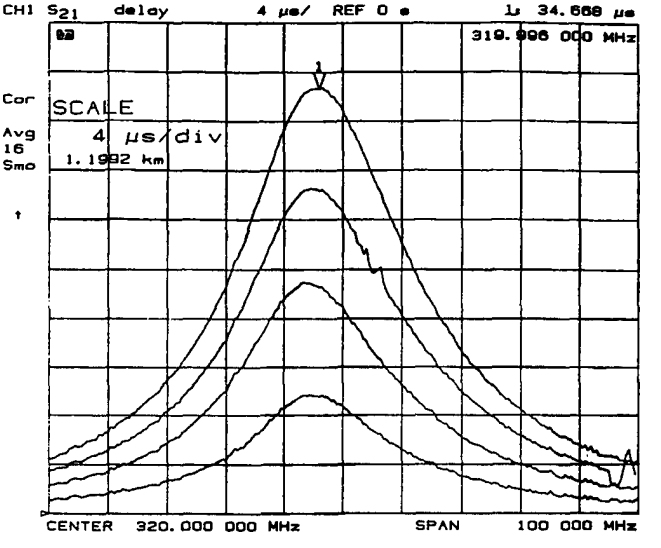


Figure 3. Measured Oscillator Open Loop Group Delay for 1, 2, 3, and 4 Cascaded SAWR/Amplifier/Power Divider Stages

Table 1. SAWR Vibration Sensitivity Measurement Results (Parts in 10^9)

	Vibration Direction (Figure 4)		
	"X"	"Y"	"Z"
Resonator #1	2	3.5	50
Resonator #2	≤0.5	≤0.5	55
Resonator #3	5	5	43
Resonator #4	2.5	3	48
	"A"	"B"	"C"
Oscillator with Resonators #1, 2, 3, 4	≤0.5	≤0.5	3.5

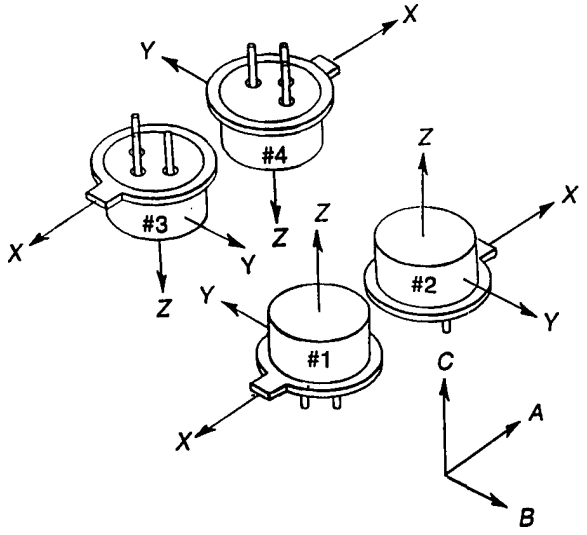


Figure 4. Relative SAWR Orientation for 3-Axis Vibration Sensitivity Reduction

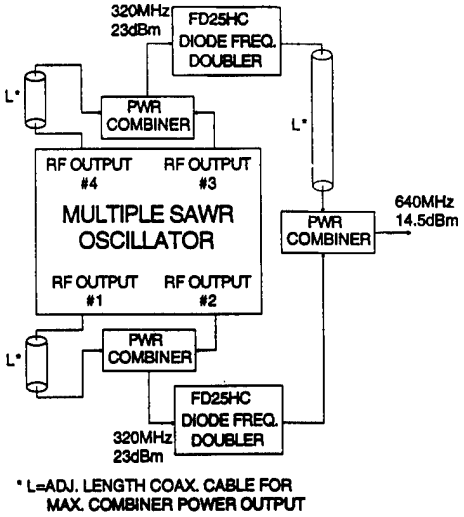


Figure 5. Multiple SAWR Oscillator/Frequency Doubler Block Diagram

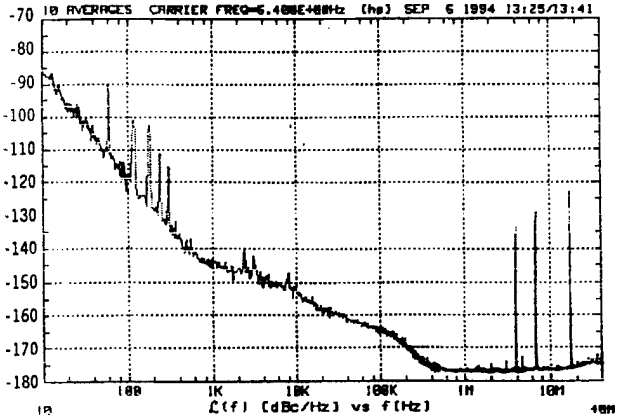


Figure 6. Phase Noise Measurement Results: Two Phaselocked, 640MHz Reference Signal Generators with (300KHz Bandwidth) Output Bandpass Filters (Noise Floor=-177dBc/Hz)

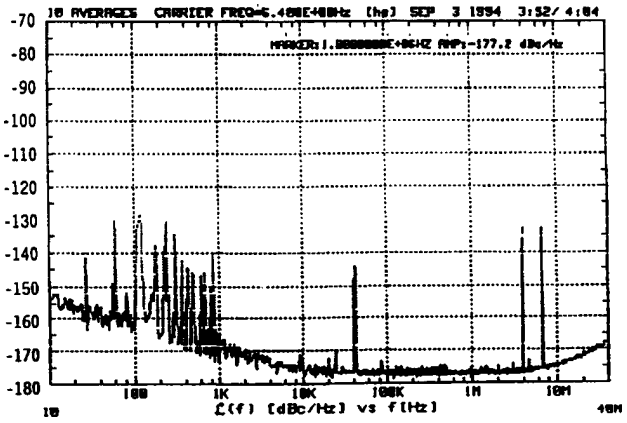


Figure 7. Phase Noise Measurement Results: HP11740A Measurement System Self-Noise with Common 640MHz, 14 dBm System Phase Detector L, R Port Drive (Noise Floor=-177dBc/Hz)

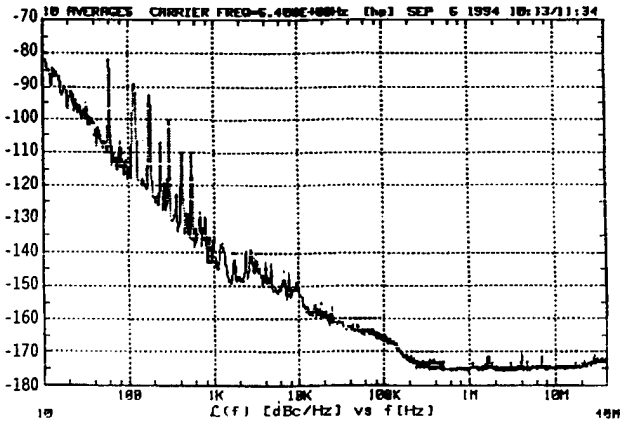


Figure 8. Phase Noise Measurement Results; 640MHz Reference Signal Generator Phaselocked to Multiple SAWR Oscillator/Frequency Doubler (Noise Floor=-175dBc/Hz)

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ACCELERATION CHARGE SENSITIVITY IN AT-QUARTZ RESONATORS

C. L. Anderson*
Motorola
2100 East Elliot Road, Tempe, AZ 85284. e-mail: a87490@email.mot.com

J. S. Bagby, R. L. Barrett and V. Ungvichian
Florida Atlantic University
Boca Raton, FL 33431. e-mail: bagby@ee.fau.edu

ABSTRACT

The behavior of AT-quartz resonators subject to acceleration is studied. For an AT-strip resonator with cantilever mounting, piezoelectric theory predicts that excitation of the pure lowest frequency flexural mode of vibration generates no charge. However, experiments show electromechanical resonance corresponding to this mode for sinusoidal support motion normal to the plane of the crystal with no applied electric field. An amplifier with a voltage follower first stage senses charge output from the crystal. Shielding is required to reduce sixty hertz and drive frequency electromagnetic interference. Measurement of resonant frequency and damping factor for the lowest frequency flexural mode is reported for a group of 17.76 MHz oscillator crystals. This mode is responsible for crystal breakage in some portable communications products. Significant variation in sensitivity among crystals was observed. Charge sensitivity of this mode is attributed in part to mechanical coupling to other modes involving torsion and lateral flexure.

I. INTRODUCTION

Quartz crystal resonators are used as frequency references in portable communications products such as pagers to demodulate the radio frequency signal and to establish clock rate for the microprocessor. These crystals must be designed to withstand acceleration levels of thousands of g's without breaking. Fragility of crystals may be affected by low frequency resonant modes, and as such, determination of their resonance characteristics is useful for crystal design and diagnosis of failure.

One of the crystal configurations used in pagers is the rectangular "AT-strip" quartz resonator mounted as a cantilever plate [2], [3] as shown in Fig. 1. The specific crystal unit used in this study is manufactured by Motorola in a surface mount package with crystal blank dimensions of length $l = 6.10$, width $w = 1.87$ and thickness $t = 1.0$ mm. This results in a design frequency of 17.76 MHz for the first, or lowest frequency, X-thickness shear mode shown in Fig. 2(a), and a resonance of approximately 2.8 kHz corresponding to the first flexural mode as shown in Fig. 2(b).

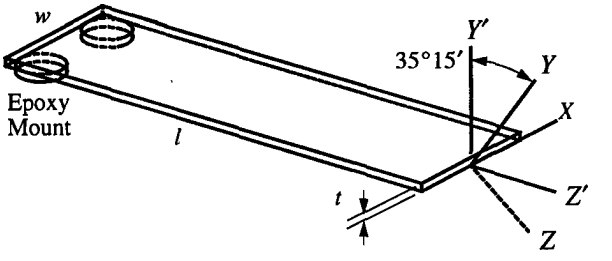


Fig. 1. Quartz AT-strip geometry.

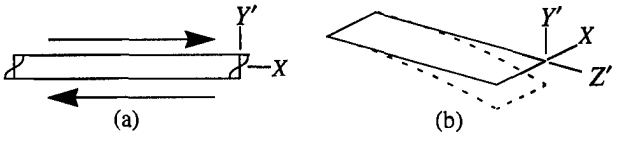


Fig. 2. (a) First X-thickness shear mode of vibration. (b) First flexural mode of vibration.

Imparting a shock pulse to the case of an AT-strip crystal generates a signal that may exhibit any of a number of resonances detectable, for example, using an oscilloscope with a standard probe. The "tap test" as it may be called has been used at Motorola as a method of determining the lowest flexural resonant frequency of AT-quartz resonators. The electrical response is due to charge generated by the crystal itself, and may be termed "acceleration charge sensitivity". Thus, the crystal is acting as a transducer.

*This work is based on a thesis submitted for the Master of Science in Electrical Engineering at Florida Atlantic University [1].

0-7803-2500-1/95 $4.00 © 1995 IEEE

The purpose of this paper is to describe a repeatable technique that isolates the first flexural mode electrical response and to develop a theory to explain why it occurs at all, since this mode by itself theoretically generates no charge. There is little or no mention in the literature regarding piezoelectric coupling for the particular case of flexural displacement in the *Y-Z* plane for AT-quartz resonators, nor for the use of a resonator as a charge-generating accelerometer.

Commonly available acceleration sensors, or accelerometers, are designed to take advantage of the direct piezoelectric effect, that applied stress generates electric displacement. They are normally made of a ferroelectric, polycrystalline ceramic rather than quartz. Charge sensitivity is often expressed in units of pC/g [4]. The indirect piezoelectric effect, that applied electric displacement generates stress, is responsible for operation of resonators used in oscillator circuits. Both effects may be described using the linear piezoelectric constitutive equations [5]-[10]. The lowest frequency resonance of accelerometers is designed to occur well above the frequency spectrum of the input in order to avoid variation of the response with frequency. The resonant frequency of a shock accelerometer is typically about 50 kHz, while calibration data is normally given up to 10 kHz [4].

Acceleration charge sensitivity is not to be confused with acceleration frequency sensitivity, in which oscillator frequency is modulated by acceleration. The latter has received considerable attention lately in the literature [11] - [16]. Filler [11] provides a review of this activity. Frequency shift is due to nonlinear elastic behavior of the crystal subject to body forces resulting in a change in the acoustic velocity. Since the frequency of a resonator is a function of the acoustic velocity and the dimensions of the quartz plate, the forces change the frequency. Typical acceleration frequency sensitivity values reported are on the order of 2 x 10^{-9}, or 2 ppb, per g of acceleration. Because of this affect, random vibration from a mobile platform can cause a large increase in the noise level of a quartz crystal oscillator for frequency sources in applications requiring extremely low noise such as missiles and satellites, Doppler radars, and other systems [11]. For pager applications, oscillator frequency stability is not as critical so that acceleration frequency sensitivity is usually neglected.

Flexural vibration of piezoelectric materials for low-frequency applications has been addressed by Langdon [17], Eernisse, et.al. [18], and Soderkvist [19] - [22]. These works discuss applications where flexure is the desired mode of operation, and where crystal orientation and electrode placement are designed to take advantage of piezoelectric coupling. The most common example of a low-frequency piezoelectric flexural resonator is the quartz crystal used as a frequency reference in watches. More generally, vibration in crystal plates has been addressed, for example, in references [23]-[30].

This paper will present piezoelectric theory for the AT-strip resonator subject to vibration input, an experimental technique to quantify the signal associated with low frequency resonance, and results for a specific crystal.

II. PIEZOELECTRIC PRINCIPLES

II. A. Fundamental Equations

The constitutive equations of linear piezoelectric theory relate stress T, strain S, electric field E, electric displacement D and the material properties elastic stiffness c, piezoelectric stress constant e, and permittivity ε by [5]-[10]

$$T_p = c_{pq}^{E} S_q - e_{kp} E_k \tag{1}$$

$$D_i = e_{iq} S_q + \varepsilon_{ik}^{S} E_k . \tag{2}$$

An alternative form using elastic compliance s and piezoelectric strain constant d is

$$S_p = s_{pq}^{E} T_q + d_{kp} E_k \tag{3}$$

$$D_i = d_{iq} T_q + \varepsilon_{ik}^{T} E_k . \tag{4}$$

The conventions used here follow the IEEE Standard on Piezoelectricity [9]. Superscript E denotes "at constant electric field" and superscript S denotes "at constant mechanical strain." The subscripts p, q are related to coordinate directions as shown in Table 1. The subscripts i, k take on values of 1 (x), 2 (y) or 3 (z). This notation, convenient for matrix manipulation, is actually a reduced notation that is commonly used in the literature [5]-[10] due to symmetry. Elastic stiffness, for example, is actually a fourth-rank tensor. Equations (1) and (3) represent the converse piezoelectric effect: applied electric field generates stress (or strain) in the crystal. Equations (2) and (4) represent the direct piezoelectric effect: applied strain (or stress) results in electric displacement.

Quartz is anisotropic and is classified as a trigonal crystal. The term AT-refers to a specific orientation of the quartz crystal blank relative to the crystallographic axes *X*, *Y* and *Z*. The AT-cut is designed to reduce the effect of temperature on resonator operating frequency. For an initially *Y*-cut plate (thickness in the *Y*-axis), the AT-cut designates a rotation about the X-axis of $\theta = -35°15'$ as shown in Fig. 3 (although the precise angle depends on the application).

The positive sense of θ follows the right-hand rule. The prime symbol (′) is used to designate a rotated value. In Fig. 1, the *X*-axis is not primed because it is not affected by this rotation.

Table 1. Matrix notation.

p or *q*	Description
1	x - Elongation
2	y - Elongation
3	z - Elongation
4	yz - Shear
5	zx - Shear
6	xy - Shear

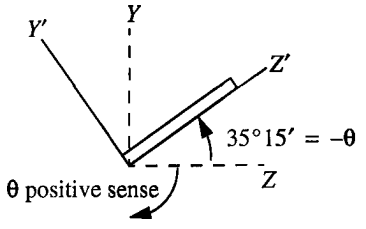


Fig. 3. Orientation of AT-strip.

Salt [8] gives a thorough description of coordinate transformations for crystals. Dropping the subscripts and superscripts for ease of notation, equations (1)- (4) become

$$T = cS - \tilde{e}E \tag{5}$$

$$D = Es + \varepsilon E \tag{6}$$

$$S = sT + \tilde{d}E \tag{7}$$

$$D = dT + \varepsilon E \tag{8}$$

where the symbol "~" denotes transpose. In rotated coordinates equations (5)-(8) become

$$T' = (\tilde{M}cM)\,S' - (\tilde{M}\tilde{e}A)\,E' \tag{9}$$

$$D' = (\tilde{A}eM)\,S' + (\tilde{A}\varepsilon A)\,E' \tag{10}$$

$$S' = (\tilde{N}sN)\,T' + (\tilde{N}\tilde{d}A)\,E' \tag{11}$$

$$D' = (\tilde{A}dN)\,T' + (\tilde{A}\varepsilon A)\,E' \tag{12}$$

where *A*, *M* and *N* are transformation matrices. Thus, the transformation laws for the material constants become $c' = \tilde{M}cM$, $e' = \tilde{A}eM$, $\varepsilon' = \tilde{A}\varepsilon A$, $s' = \tilde{N}sN$ and $d' = \tilde{A}dN$.

For rotation about the coordinate axis *X* (for rotated Y-cuts such as the AT-cut), the transformation matrix *A* is

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c & -s \\ 0 & s & c \end{bmatrix} \tag{13}$$

where $c = \cos(\theta)$ and $s = \sin(\theta)$. *M* and *N* are 6 x 6 matrices whose elements are formed from the elements of the transformation matrix *A*. The matrix *M* for rotated *Y*-cuts is

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & c^2 & s^2 & -cs & 0 & 0 \\ 0 & s^2 & c^2 & cs & 0 & 0 \\ 0 & 2cs & -2cs & c^2 - s^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & c & s \\ 0 & 0 & 0 & 0 & -s & c \end{bmatrix} \tag{14}$$

The matrix *N* is given by

$$\tilde{N} = M^{-1} \text{ or } N = (\tilde{M})^{-1} \tag{15}$$

where M^{-1} denotes the inverse of *M*. The complete set of matrix coefficients for AT-quartz is provided by Anderson [1] based on quartz constants by Bechmann [31], corrected for sign. Of particular interest here is the matrix of piezoelectric strain constants for AT-quartz, which has the form (note $d_{11} = d'_{11}$)

$$\begin{bmatrix} d_{11} & d'_{12} & d'_{12} & d'_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & d'_{25} & d'_{26} \\ 0 & 0 & 0 & 0 & d'_{25} & d'_{26} \end{bmatrix}$$

The meaning of each non-zero component of the piezoelectric strain matrix for AT-quartz is depicted in Fig. 4.

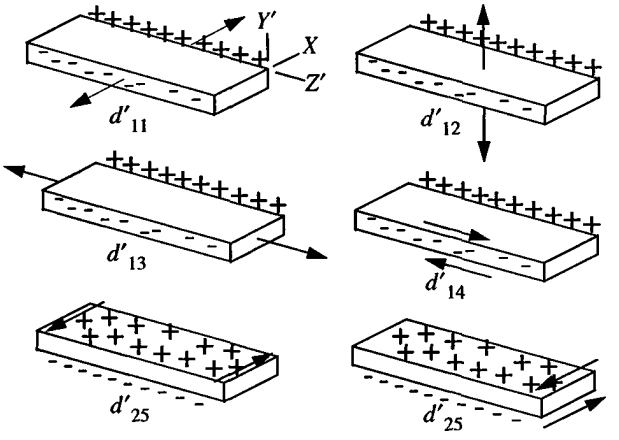


Fig. 4. Physical interpretation of piezoelectric strain constants. Arrows designate direction of applied stress.

II. B. Acceleration Charge Sensitivity

Acceleration charge sensitivity is a property of piezoelectric materials whereby charge is generated subject to mechanical input. It is the basis of operation of accelerometers. To enhance and simplify measurement of charge generation in this study, the crystal was connected to an amplifier but not to an oscillator circuit for which it was originally intended. With no applied electric field E, equation (8) for rotated axes becomes

$$D'_i = d'_{iq} T'_q. \tag{16}$$

For an AT-strip quartz resonator, the electrodes are placed on the X-Z' faces. Gauss' Law states that the total outward flux of the electric displacement over any closed surface is equal to the total free charge enclosed in the surface. Taking the closed surface S as the electrode area on one face of the plate at $Y' = 0$, the charge is given by

$$Q_c = \oint_S D'_2 dA. \tag{17}$$

Expanding equation (16) and using only non-zero piezoelectric strain constants for AT-quartz gives

$$D'_2 = d'_{25} T'_5 + d'_{26} T'_6. \tag{18}$$

For modes such as X-thickness shear that are essentially uniform across the electrode area A_e, equation (17) reduces to $Q_c = D'_2 A_e$. For flexural modes subject to vibration input, the stress and hence the electric displacement are in general a function of X and Z'. The effect of the conductive electrodes is to collect the charges across the plate regardless of nonuniformity of electric displacement. The total charge can be expressed as

$$Q_c = (d'_{25} T'_{5avg} + d'_{26} T'_{6avg}) A_e \tag{19}$$

where $T'_{5avg} = \iint T'_5 dA$ and $T'_{6avg} = \iint T'_6 dA$ are the average stress levels over the electroded area.

For the crystal used in this study, the electrode area is approximately 1.50 mm x 4.12 mm or

$$A_e = 6.18 \cdot 10^{-6} \text{ m}^2. \tag{20}$$

This includes only the area where positive and negative electrodes are opposite each other.

Substituting the values of the piezoelectric strain constants [1]

$$d'_{25} = -1.69 \cdot 10^{-12} (C/N)$$

$$d'_{26} = -3.42 \cdot 10^{-12} (C/N)$$

and the value of A_e above into equation (19) yields

$$Q_c = (-10.4) T'_{5avg} + (-21.1) T'_{6avg} \text{ pC/MPa} \tag{21}$$

where charge has units of pico Coulombs (pC) and stress has units of mega pascal (MPa) = 10^6 N/m^2.

III. MECHANICS OF CANTILEVER AT-STRIP

III. A. Stresses in the AT-Quartz Plate

To determine the stresses T'_5, T'_6 contributing to D'_2, the problem of the vibration of the AT-quartz plate may be simplified by noting that the Y-Z plane of quartz is a twofold symmetry plane [32], [8]. For the particular case of AT-cut strips with symmetric load in the Y-Z plane and symmetric geometry the response is symmetric with respect to the Y-Z plane; the plate then deflects as a beam as shown in Fig. 5(a). The non-zero stresses due to input motion of the foundation in the Y' direction are T'_3 (tensile/compressive stress in the Z' direction) and T'_4 (shear stress about the X axis).

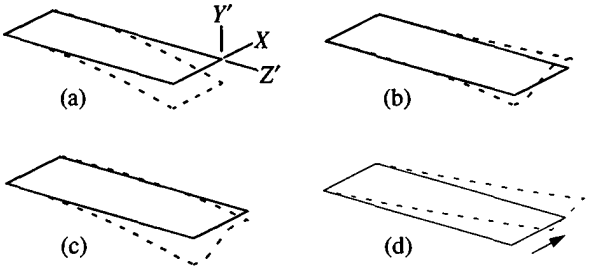


Fig. 5. Low frequency modes of vibration in cantilever plates. (a) First flexural mode. (b) First torsional mode. (c) Combined flexure and torsion. (d) Lateral flexure.

Considering only non-zero stresses as described above applied to equation (18) results in

$$D'_2 = 0. \tag{22}$$

Thus, theoretically, stresses in the cantilever AT-strip operating in the first flexural mode do not contribute to charge across the electrodes on the Y'-normal faces. Higher order flexural modes symmetric with respect to the Y-Z plane also do not generate charge.

Because the values of stress calculated in equations (33) and (34) are so low (many orders of magnitude below

the strength of quartz), one may expect that small values of the stresses T'_5, T'_6 may be coupling to the first flexural mode to cause the charge generation observed.

Exact calculation of the stresses in the AT-quartz plate for coupled mode vibration in the crystal used here is beyond the scope of this work. Difficulties arise because of anisotropic crystal properties and the effects of mounting structure deformation.

III. B. Resonant Frequencies and Mode Excitation

The purpose of this section is to obtain approximate resonant frequencies to confirm experimental results, to show that modes other than the first flexural mode are well above 2.8 kHz, and to discuss how piezoelectrically active modes may be excited in the problem at hand. While more exact techniques are possible for calculating resonant frequency, the techniques used here agree reasonably well with experimental results. Two difficulties with the calculation are the anisotropic properties of quartz and the deflection of the conductive epoxy supports within the crystal package. The finite element method is a good candidate for this application, since the program ANSYS, for example, can account for the anisotropic properties of quartz [32]. However, the material behavior of the conductive epoxy is difficult to model. Here, solutions for an isotropic plate with rigid support are used as an approximation.

The resonant frequency of a thin cantilever beam of rectangular cross-section with isotropic material properties and an infinitely stiff mounting structure is given by [33], [34]

$$f_i = \frac{(\beta_i l)^2}{2\pi l^2}\sqrt{\frac{yt^2}{12\rho}} \tag{23}$$

where i is the order of the mode, $(\beta_i l) = 1.875$ (dimensionless) for $i = 1$ (first mode), y is Young's modulus and ρ is density. For flexure of a long, thin beam, equation (23) is reasonably accurate for AT-quartz, given the use of the AT-strip rotated value of Young's modulus in the Z'-direction. For the case of the 17.76 MHz quartz AT-strip crystal used in this experiment, $y = 1/s'_{33} = 90.83 \cdot 10^3$ MPa [1], $\rho = 2650$ kg/m^2, $t = 1.0$ mm and $l = 6.10$ mm which yields

$$f_1 \approx 2.52 \text{ kHz } (l = 6.10 \text{ mm}). \tag{24}$$

This assumes the boundary condition is a perfect clamp at the very end of the crystal, and represents a lower limit on the resonant frequency. An upper limit on the resonant frequency can be obtained by treating the epoxy clamp (approximately 0.66 mm long) as rigid relative to the crystal elasticity. This effectively shortens the length and thus raises the resonant frequency. Using a shortened length $l = 6.10 - 0.66 = 5.44$ mm yields $f_1 \approx 3.17$ kHz. The next two modes ($i = 2, 3$) for $l = 6.10$mm occur at 15.8 kHz and 44.2 kHz. Neither mode should contribute to charge generation for deflection symmetric about the X-Y plane.

Other low-frequency vibrational modes are plate torsion and lateral bending shown in Fig. 5(b) and (d), both of which induce stresses that have significant piezoelectric coupling coefficients (d'_{26}, d'_{25}) for the AT-strip. It is important to establish that these modes resonate well above 2.8 kHz. The first torsional mode for the given plate parameters using handbook isotropic plate equations is [35], [1]

$$f_{t1} \approx 17.4 \text{ kHz}. \tag{25}$$

The first lateral mode occurs at

$$f_{lb1} \approx 47.6 \text{ kHz}. \tag{26}$$

which is close to one of the tap test frequencies reported in Section V. The frequencies for both torsion and lateral bending occur well above the lowest flexural resonant frequency. Approximate calculated resonant frequencies from equations (24), (25) and (26) are summarized in Table 2 for $l = 6.10$ mm.

For sinusoidal input, lateral flexure can be induced by providing support motion in the X-direction. The torsional mode can be induced in shock by rotational input support motion θ'_z. If the crystal is mounted at an angle θ'_z relative to the mechanical sine wave input, lateral bending and possibly torsion will couple to the first beam flexural mode. If the crystal has any mounting or dimensional asymmetry, the first torsional mode will couple to the first beam flexural mode as shown in Fig. 5(c). Any of these coupling phenomena may be responsible for charge generation associated with beam flexure since both torsional and lateral bending modes have piezoelectric coupling for the AT-strip.

Table 2. Approximate calculated resonant frequencies for $l = 6.10$ mm.

Vibration Mode	Frequency (kHz)
First flexural	2.52
First torsional	17.4
Lateral flexural	47.6

IV. EXPERIMENTAL SETUP

IV. A. Hardware

The experimental setup is shown in Fig. 6. It consists of the crystal in a surface-mount package mounted to the top of a sinusoidal vibration exciter. A reference accelerometer is placed beside the crystal to provide calibrated acceleration data for each data point.

In order to minimize electromagnetic interference (EMI), an aluminum extension (diameter 5 x height 15 cm) was used to position the crystal well above the exciter's coil, and the extension was grounded. A dielectric spacer (plastic) 3 mm thick was used as a platform for mounting the printed circuit board (PCB) made of FR-4 to which the crystal was soldered. The PCB and reference accelerometer were attached to the spacer using epoxy. A grounded copper sheet covered the entire platform including the crystal. Coax cable type RG-178B/U (29 pF/ft [36]) of length 20 cm connected the crystal to the amplifier, which was housed in a grounded enclosure. One technique used to troubleshoot EMI at the drive frequency was to use the crystal as an antenna, disconnected mechanically from the vibration exciter. The crystal was simply positioned close to, but not touching the platform. With proper shielding, interference at the drive frequency was indistinguishable from the broadband background noise level. To minimize 60 Hz interference further, two high-pass filters (HPF) were used. One consisted of a 20 M-ohm resistor across the crystal capacitance, which had the additional benefit of preventing DC drift.

The audio source frequency chosen was the HP8903B Audio Analyzer which allowed for digital control of the frequency with resolution of about 5 Hz as manually dialed in. Each data point was taken after 1 to 2 seconds of "on" time to make sure the mechanical exciter stabilized. At that point, the single mode function on the oscilloscope was selected to capture the continuous signal. The output from the reference accelerometer mounted to the exciter had a distortion of about 0.5% as measured by the audio analyzer. On the LeCroy 9400A oscilloscope, the time division was set at 2 ms per division, with the FFT set to 1250 points, flat top window and power spectrum. A crystal amplifier gain level of G_o = 100x was found to give good signal-to-noise ratio at resonance.

Crystal linearity was verified by taking data over 20 dB (one order of magnitude) of acceleration input from 0.2 g to 2.0 g rms at resonance. One sample crystal tested also continued to be linear up to +10 dB over test amplitude (up to 6 g rms acceleration).

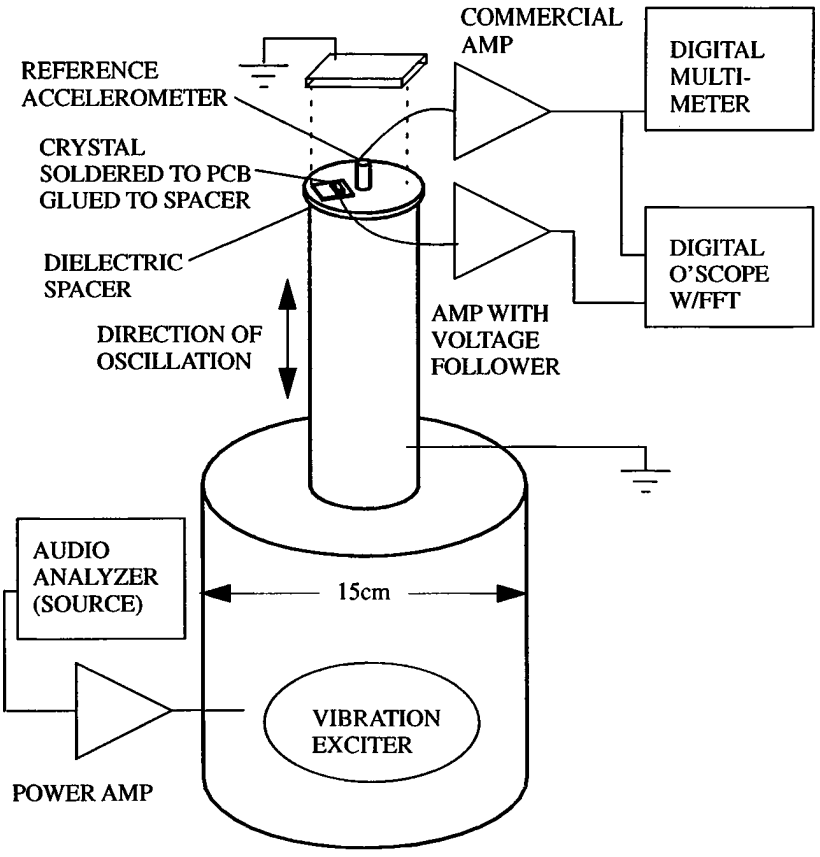


Fig. 6. Experimental setup.

IV. B. Amplification of Induced Charge

The detection of the charge induced in a crystal oscillator subject to mechanical input is a similar problem to that of a standard piezoelectric accelerometer. One main difference, however, is in the sensitivity of the generating device. A charge sensitivity of 0.6 pC/g (electrical charge per unit of acceleration) is typical for a low-sensitivity commercial accelerometer used for acceleration up to 4000 g [37]. For the crystal oscillator used in this study, peak sensitivity was found to range from 2 to 40 pC/g, and baseline sensitivity from 0.01 to 0.7 pC/g, where "baseline" refers to off-resonance in the limit as frequency approaches zero. Also, the input forcing function, a mechanical sine wave, was limited to a maximum of about 7 g peak because of equipment limitations. The reference accelerometer upper frequency calibration limit was 10 kHz.

The output from piezoelectric accelerometers can be amplified by either charge sensing or voltage sensing electronics. A charge converter senses charge and is essentially a voltage amplifier with capacitance feedback [38]. It is commonly used for accelerometers since the effect of variation in cable capacitance is minimized. The amplifier used in this experiment senses voltage for optimum sensitivity and is shown in Fig. 7. It is capable of generating a usable signal for crystal output as low as Q_c = 0.0005 pC rms based on a cable plus crystal plus amp input capacitance of C_t = 29 pF and a second-stage amplifier gain of G_o = 100, or $G_o(Q_c/C_t)$ = 2 mV rms output. A JFET-input operational amplifier, Motorola MC33182P, was used for both the first and second stages. It was selected because of its high input resistance and low input capacitance, which was important for the first stage. The crystal can be modeled as a capacitor C_o with a parallel current source i (ω) representing the crystal charge sensitivity that is a function of frequency.

Capacitance of the crystal, cable and amplifier input directly affect the output sensitivity. The crystal parallel plate capacitance is calculated as C_{o1} = 0.46 pF [1]. Measured capacitance of the combined crystal/case was C_o = 3.4 pF. The PCB (3.5 pF) and coaxial cable (19 pF) act as a parallel load capacitance C_x along with the amplifier input (3.0 pF typical) C_i resulting in a total capacitance $C_t = C_o + C_x + C_i$ = 29 pF. A shunt resistor R_c = 20 MΩ was used to eliminate DC drift and to create a HPF effect. These elements are shown in Fig. 8. The input resistance of the op amp, $10^{12}\Omega$, is neglected.

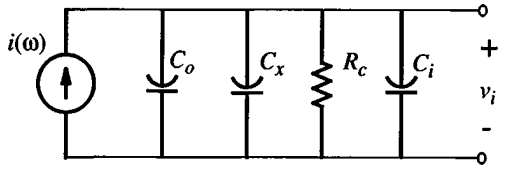


Fig. 8. Model of crystal (including charge sensitivity), cable and shunt resistance. v_i is the input to the first stage op amp.

The overall theoretical response of the amplifier considering two HPF's and one LPF is

$$\left|\frac{V_i C_t}{Q_c}\right| = \frac{1}{\sqrt{1+(f_{c1}/f)^2}} \frac{1}{\sqrt{1+(f_{c2}/f)^2}} \qquad (27)$$

$$\frac{G_o}{\sqrt{1+(f/f_t)^2(1+G_o)^2}} \; .$$

The two HPF's are separated by the first stage op amp so that the system response in the frequency domain is simply the product of the two individual responses [39]. A plot of equation (27) is shown in Fig. 9. Variation in response from 2.6 to 3.0 kHz is 0.03 dB.

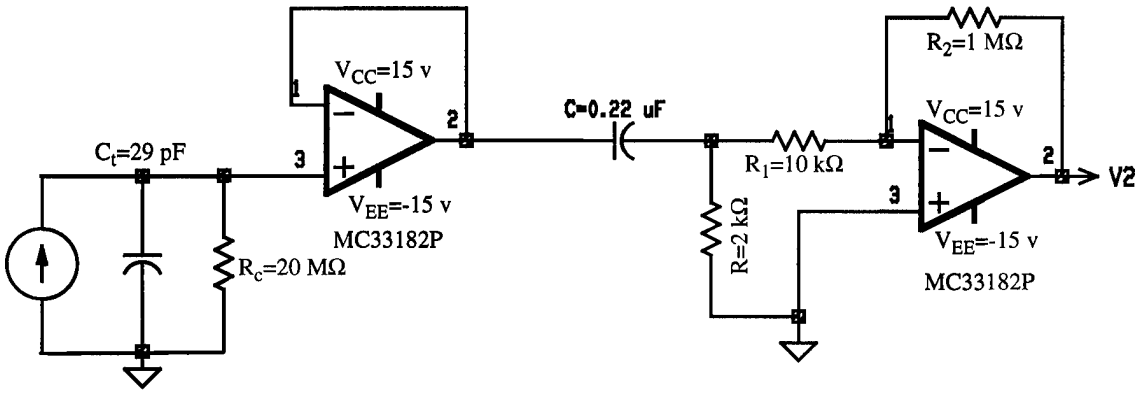


Fig. 7. Amplifier circuit schematic.

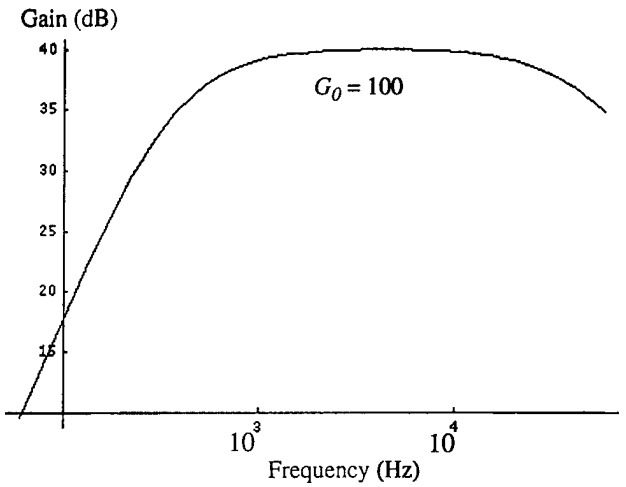


Fig. 9. Amplifier circuit theoretical frequency response.

V. EXPERIMENTAL RESULTS

Acceleration charge sensitivity data in the vicinity of resonance for six AT cut 17.76 MHz surface mount quartz crystal oscillators is summarized in Table 3. Data was taken for each crystal mounted in the horizontal position resulting in input acceleration perpendicular to the crystal plate (Y'-direction). Some crystals were also mounted in positions other than horizontal. Crystals C1, C2 and C3 were uncapped prior to testing to evaluate geometrical characteristics that might lead to asymmetry. The data with curve fitting is plotted in Fig. 10 for two crystals. Raw data and plots for all crystals are provided in reference [1].

Table 3. Resonance results.

Crystal #	Resonant Frequency (kHz)	Peak Charge Sensitivity (pC/g)	Damping Factor, ζ
C1 horiz	2.807	42.1	.00874
C2 horiz	2.805	3.73	.00307
C3 horiz	2.786	9.27	.00242
D1 horiz	2.806	6.51	.00236
D1 45°	2.797	4.54	.00223
D1 vert	2.803	2.58	n/a
D2 horiz	2.824	13.6	.00185
D2 -5° horiz	2.824	2.25	.00200
D2 +5°horiz	2.824	4.87	.00183
D3 horiz	2.699	11.68	.00271

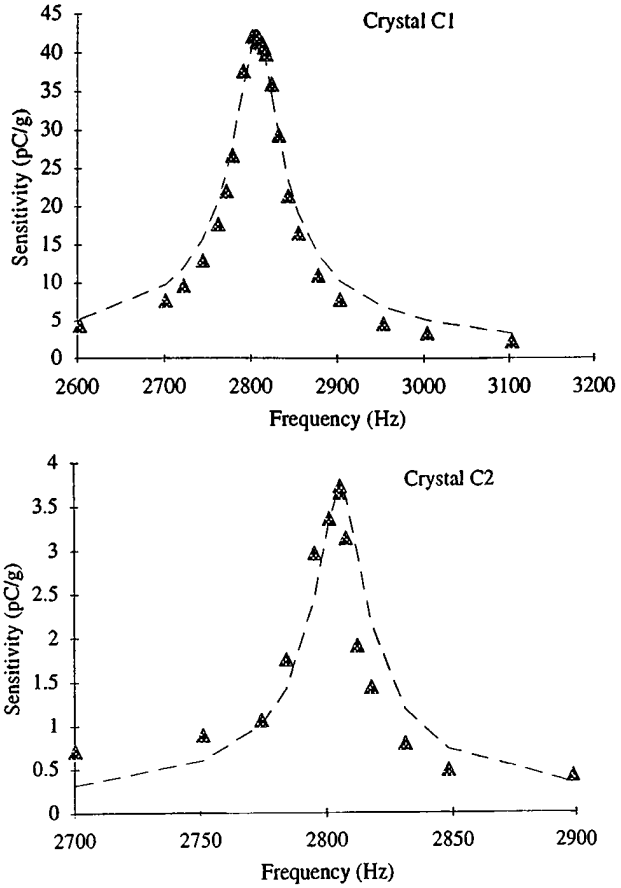


Fig. 10. Resonance data for two crystals.

It is clear that a resonance occurs for each crystal. Resonant frequency ranges from 2.699 to 2.824 kHz. These values fall within the range of the calculated value of the resonant frequency obtained from equation (23). This range was 2.516 to 3.165 kHz, accounting for different ways of treating the boundary conditions. Damping factor is seen to be on average about 0.0025. This is very low in terms of shock damping characteristics.

Crystal charge sensitivity is found from

$$\frac{Q_c}{g} = C_t \frac{(V_o/G_o)}{(V_o)_{ref}/\beta} \tag{28}$$

where V_o is output voltage from the crystal amplifier, C_t is total capacitance (crystal, PCB, cable and op amp input), G_o is second stage op amp gain, $(V_o)_{ref}$ designates reference accelerometer output, and β is the calibration factor of the reference accelerometer and amplifier. Baseline acceleration charge sensitivity, corresponding to theoretical response at frequencies well below resonance, is found from

$$\left(\frac{q}{g}\right)_{baseline} = \left(\frac{q}{g}\right)_{peak}(2\zeta) \,. \tag{30}$$

The signal-to-noise ratio was too low to obtain baseline charge sensitivity directly.

Curve fitting and calculation of damping factor are done by analogy with a single degree of freedom oscillator shown in Fig. 11.

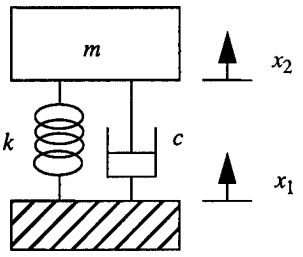


Fig. 11. Single degree of freedom mechanical oscillator.

The relationship between the translation x_1 of a moving support and the translation x_2 of the mass is given by the transmissibility equation [34]

$$TR = \frac{\sqrt{1 + (2\zeta r)^2}}{\sqrt{\left(1 - r^2\right)^2 + (2\zeta r)^2}} = \frac{x_2}{x_1} \tag{31}$$

where $\zeta = c/(2\sqrt{km})$, $r = \omega/\omega_n$ and $\omega_n = \sqrt{k/m}$. Damping factor is shown in reference [1] to be

$$\zeta \approx \overline{BW}_{3dB}/2 \tag{32}$$

where $\overline{BW}_{3dB} = BW_{3dB}/f_r = 1/Q$ is the normalized bandwidth and Q is the quality factor. The 3 dB point on either side of resonance was determined by linear interpolation between the two data points closest to 3 dB down from the peak sensitivity.

The data indicate a large variation in acceleration charge sensitivity between the various samples. Some of this variation may be due to asymmetric contamination since for C1, C2 and C3 the caps were removed. An experiment was conducted to evaluate the effect of the rotational position of the sample relative to the input forcing function. The crystal D2 was mounted at three different positions: horizontally, at -5° and at +5°. Even with this small angle, the charge sensitivity changed drastically, with the highest value at the horizontal position. The accuracy of soldering any one of the samples to the PCB is estimated to be +/-2°. Thus, the mounting angle may have had a significant effect on the variability reported.

The vertical orientation refers to acceleration in the X-direction. Sensitivity in this case is seen to have some form of resonance near 2.8 kHz, indicating that the first flexural mode is being excited. The shape of the resonance curve was not symmetric, so damping factor was not calculated.

The charge sensitivity results can be used to calculate the area-average stress due to an individual assumed mode. This is not the actual total stress, of course, since multiple modes may be present. Given a charge sensitivity of 10 pC/g (typical value at resonance) for an input acceleration of 2.83 g peak (2 g rms), and assuming for the moment that charge is due solely to stress T'_{5avg} ($T'_{6avg} = 0$) yields

$$T'_{5avg} = 2.72 \text{ MPa (394 psi) at resonance.} \tag{33}$$

At frequencies well below resonance, a baseline sensitivity of 0.1 pC/g at 2.83 g peak yields

$T'_{5avg} = 0.0272$ MPa (3.94 psi) baseline.

Similarly, if the charge were due solely to stress T'_{6avg} ($T'_{5avg} = 0$), then

$$T'_{6avg} = 1.34 \text{ MPa (195 psi) at resonance.} \tag{34}$$

Another possible cause of acceleration charge sensitivity, although believed to be insignificant based on rough calculations, is crystal X-axis alignment error during manufacturing. With rotation of the crystallographic axes about Y' and/or Z' by a fraction of a degree, additional terms in the piezoelectric stress coefficient matrix become small but non-zero. Any of the coefficients d'_{22}, d'_{23} or d'_{24} which are linked to stresses T'_2 (tension along Y'), T'_3 (tension along Z') and T'_4 (shear about X) could result in charge generated in the Y'-direction subject to the first flexural mode vibration.

Possible causes not investigated here are non-uniform (for example, wedge shaped) plate thickness, and nonlinear material behavior. The effect of non-uniform thickness in the form of a wedge for thickness-shear vibrations of a rectangular plate was considered by Dworsky [30].

A tap test was also conducted on one crystal. Oscilloscope traces are shown in reference [1] for the crystal and for the reference accelerometer (for the same event). The FFT of the crystal signal shows resonances at 2.8 kHz (+/-0.1 kHz) and 41.2 kHz. The latter value appears to correspond to the lateral flexural mode with a calculated resonant frequency from Section III. B. of approximately 47.6 kHz. The FFT of the simultaneous reference accelerometer signal demonstrates that these resonances are not due to some other characteristic.

VI. APPLICATIONS

Knowledge of frequency and damping factor is useful for analyzing the fragility of crystals subject to shock in portable communications products. Measurement of the lowest resonant frequency is accurate and can be used to validate analytical models. However, the crystal resonator is generally not a good choice as a shock accelerometer. For the AT-strip, for example, the first flexural mode of vibration that may contribute the most to stress has no piezoelectric coupling whereas other modes which may be excited in shock do have piezoelectric coupling.

The significance of low frequency resonance to crystal fragility may be explained using Fourier transform theory. The shock pulse on printed circuit boards in pagers closely resembles a 0.3-msec pulse-width half-cycle sine pulse. Its Fourier transform has magnitude

$$|F(\omega)| = \frac{A\pi}{2\omega_o}\left|\frac{\sin(\pi x_1)}{(\pi x_1)} + \frac{\sin(\pi x_2)}{(\pi x_2)}\right| \qquad (35)$$

where $\omega_o/(2\pi) = 1/0.3 = 3.33$ kHz, A is the pulse amplitude,

$$x_1 = \frac{(\omega - \omega_o)}{2\omega_o} \text{ and } x_2 = \frac{(\omega + \omega_o)}{2\omega_o}.$$

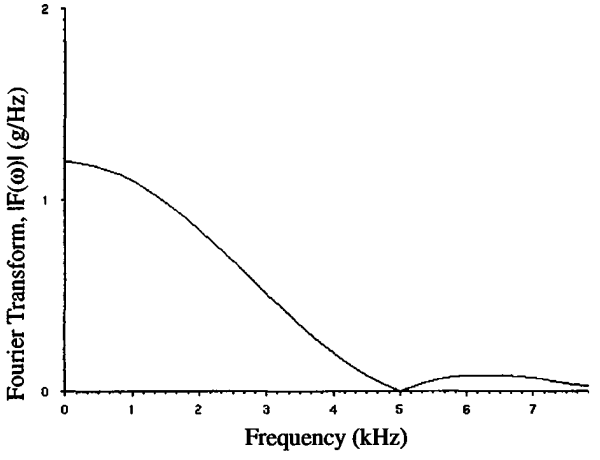


Fig. 12. Fourier transform magnitude for 0.3 ms half sine pulse.

Equation (35) is plotted in Fig. 12 for unit pulse amplitude. For a quartz crystal with a resonant frequency of 2.8 kHz and low damping factor subject to the input shock spectrum shown, the time-domain response will be amplified resulting in a tendency to fracture. If the resonant frequency was higher, for example at 5.0 kHz, the amplitude of the time-domain response would be much less. The time-domain transient response to a half-cycle sine pulse is given in reference [40]. Vibration theory is also presented in references [33], [34], [41] and [42].

VII. CONCLUSIONS

Use of an AT-quartz crystal resonator as a transducer has been demonstrated. Charge sensitivity ranging from approximately 4 to 40 pC/g at resonance was observed for input normal to the plane of the crystal, with resonant frequency ranging from about 2.70 to 2.82 kHz in six AT-strip crystals. The only vibrational mode with a resonance near this frequency is the first flexural mode. However, this mode theoretically does not generate charge. Mechanical coupling of the first beam resonant mode with piezoelectrically active modes such as plate torsion and lateral flexure is considered as a possible mechanism for charge generation. These modes could be coupled to the first flexural mode by asymmetric geometry of the strip or its mounting, or by excitation which is not purely in the Y' direction. Flexural resonant frequency and damping factor are shown to be important in design for shock in portable communications equipment. However, the crystal resonator is generally not a good choice as a shock accelerometer. Reduction of EMI was required, especially at 60 Hz and the drive frequency, using proper grounding, shielding and filtering.

VIII. RECOMMENDATIONS FOR FURTHER STUDY

1) The effect of vibration-induced charge on oscillator performance could be determined by computer simulation or by measurement of a single-shot event associated with a shock pulse. 2) The converse piezoelectric effect could be used to determine resonance characteristics. A circuit which is especially useful in detecting and measuring the resistance of weak modes is the bridge circuit [7]. 3) Exciting low-frequency modes by electrical input might allow the use of X-ray topography of these modes. 4) Other crystal packages and resonator geometries could be evaluated. 5) Crystal resonance characteristics can be used in the design of new geometries and mounting configurations in order to reduce susceptibility to shock. 6) Mode coupling as the cause of acceleration charge sensitivity could be quantified.

ACKNOWLEDGEMENTS

For many helpful discussions, the authors would like to thank Ken Burch, Weiping Zhang, Ted Lind, Bob Witte, Chuck Mooney and Evencio Fernandez, all of Motorola.

REFERENCES

[1] C. L. Anderson, *Acceleration Charge Sensitivity in AT-Quartz Resonators*. Masters Thesis, Florida Atlantic University, April, 1995.

[2] W. T. Woodland and R. S. Witte, "AT Strip Resonator Applications and Capabilities," *13th Annual Piezoelectric Devices Conf Proc*, EIA, 1991.

[3] J. P. Tomase and L. N. Dworsky, "Analysis and design of coupled mode AT rectangular resonators," *38th Annual Freq. Cont. Symp.*, 1984, pp. 141-149.

[4] "Instruction manual for Endevco piezoelectric accelerometers," Endevco, 1969.

[5] W. G. Cady, *Piezoelectricity*, Vol. 1 and 2. New York: Dover, 1964.

[6] H. F. Tiersten, *Linear Piezoelectric Plate Vibrations*. New York: Plenum Press, 1969.

[7] V. E. Bottom, *Introduction to Quartz Crystal Unit Design*. New York: Van Nostrand Reinhold, 1982.

[8] D. Salt, *Hy-Q Handbook of Quartz Crystal Devices*. UK: Van Nostrand Reinhold, 1987.

[9] ANSI/IEEE Standard 176-1987, IEEE Standard on Piezoelectricity.

[10] G. S. Kino, *Acoustic Waves: Devices, Imaging, and Analog Signal Processing*, Prentice-Hall, 1987.

[11] R. L. Filler, "The acceleration sensitivity of quartz crystal oscillators: a review," *IEEE Trans. Ult., Ferr., Freq. Contr.*, vol. 35, no.3, pp. 297-305, May 1988.

[12] J. R. Vig, *et. al.*, "Acceleration, vibration and shock effects - IEEE standards project P1193, *46th Annual Freq. Cont. Symp.*, IEEE, pp. 763-781, 1992.

[13] R. C. Smythe, "Acceleration effects in crystal filters," *IEEE Trans. Ult., Ferr., Freq. Contr.*, vol. 39, no.3, pp. 335-340, May 1992.

[14] M. M. Driscoll, "Quartz crystal resonator g sensitivity measurement methods and recent results," *IEEE Trans. Ult., Ferr., Freq. Contr.*, vol. 37, no.5, pp. 386-392, Sep., 1990

[15] J. A. Kosinski and A. Ballato, "Advances in acceleration sensitivity measurement and modeling," *46th Annual Freq. Cont. Symp.*, IEEE, pp. 838-847, 1992

[16] H. F. Tiersten, "Perturbation theory for linear electroelastic equations for small fields superposed on a bias," *J. Acoust. Soc. Am.*, vol. 64, no. 3, pp. 832-837, Sep., 1978.

[17] R. M. Langdon, "Resonator sensors - a review," *J. Phys. E: Sci. Instrum.*, vol. 18, pp. 103-115, 1985.

[18] E. P. Eernisse, R. P. Ward and R. B. Wiggins, "Survey of quartz bulk resonator sensor technologies," *IEEE Trans. Ult., Ferr., Freq. Contr.*, vol. 35, no.3, pp. 323-330, May, 1988.

[19] J. Soderkvist, "Electrical equivalent circuit for flexural vibrations in piezoelectric materials," *IEEE Trans. Ult., Ferr., Freq. Contr.*, vol. 37, no.6, pp. 577-586, Nov., 1990.

[20] J. Soderkvist, "An analysis of space-dependent electric fields used in exciting flexural vibrations of piezoelectric beams," *Meas. Sci. Techno.*, vol. 1, pp. 731-737, 1990.

[21] J. Soderkvist, "Dynamic behavior of a piezoelectric beam," *J. Acoust. Soc. Am.*, vol. 90, no. 2, pt.1, pp. 686-692, Aug., 1991.

[22] J. Soderkvist, "An equivalent circuit description of two coupled vibrations," *J. Acoust. Soc. Am.*, vol. 90, no. 2, pt.1, pp. 693-699, Aug., 1991.

[23] R. D. Mindlin, "Thickness-shear and flexural vibrations of crystal plates," *J. Appl. Phys.*, vol. 22, no. 3, pp. 316-323, March, 1951.

[24] R. F. Mitchell, "Generation and detection of sound by distributed piezoelectric sources," *Philips Research Reports Supplements*, Philips Research Laboratories, no. 3, 1972.

[25] P. C. Y. Lee and D. W. Haines, "Piezoelectric crystals and electro-elasticity," in *R. D. Mindlin and Applied Mechanics*, G. Herrmann, ed. New York: Pergamon Press, 1974.

[26] H. Bahadur and R. Parshad, "Acoustic vibrational modes in quartz crystals: their frequency, amplitude, and shape determination," *Physical Acoustics*, vol. 16, W. P. Mason and R. N. Thurston, ed., pp. 37-171, Academic Press, 1982.

[27] P. C. Y. Lee, S. Syngellakis and J. P. Hou, "A two-dimensional theory for high-frequency vibrations of piezoelectric crystal plates with or without electrodes," *J. Appl. Phys.*, vol. 61, no. 4, pp. 1249-1262, Feb. 15, 1987.

[28] P. C. Y. Lee and J. Wang, "Vibrations of AT cut quartz strips of narrow width and finite length," *J. Appl. Phys.*, vol. 75, no. 12, pp. 7681-7695, June 15, 1994.

[29] Y.-H. Pao and R. K. Kaul, "Waves and vibrations in isotropic and anisotropic plates," in *R. D. Mindlin and Applied Mechanics*, G. Herrmann, ed. New York: Pergamon Press, 1974.

[30] L. N. Dworsky, "Properties of AT quartz resonators on wedgy plates," *37th Annual Freq. Cont. Symp.*, IEEE, pp. 232-238, 1983.

[31] R. Bechmann, "Elastic and piezoelectric constants of alpha-quartz," *Physical Review*, vol. 110, no. 5, pp. 1060-1061, June 1, 1958.

[32] W. Zhang, personal communication, Dec., 1994.

[33] W. Weaver, Jr., S. P. Timoshenko and D. H. Young, *Vibration Problems in Engineering*, 5th ed. New York: John Wiley, 1990.

[34] F. S. Tse, I. E. Morse and R. T. Hinkle, *Mechanical Vibrations*, 2nd ed. Allyn and Bacon, 1978.

[35] W. F. Stokey, "Vibration of systems having distributed mass and elasticity," in *Handbook of Shock and Vibration*, C. M. Harris, Ed., 3rd ed. New York: McGraw-Hill, 1988.

[36] *Reference Data for Engineers: Radio, Electronics, Computer and Communications*, 7th ed. (Chapter 29). Howard W. Sams and Co., Indianapolis, Indiana, 1988.

[37] Data sheet, Endevco Model 25 Accelerometer. San Juan Capistrano, California: Endevco.

[38] Endevco Model 2775A Signal Conditioner. San Juan Capistrano, California: Endevco, 1992.

[39] M. E. Van Valkenburg, *Analog Filer Design*. New York: Holt, Rinehart and Winston, 1982.

[40] R. S. Ayre, "Transient response to step and pulse functions," in *Handbook of Shock and Vibration*, C. M. Harris, Ed., 3rd ed. New York: McGraw-Hill, 1988, Chapter 8.

[41] J. N. MacDuff and J. R. Curreri, *Vibration Control*, McGraw-Hill, 1958.

[42] L. S. Jacobsen and R. S. Ayre, *Engineering Vibrations*, McGraw-Hill, 1958

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SURFACE ACOUSTIC WAVE PROPAGATION ON PIEZOELECTRIC SUBSTRATES WITH THIN-FILM PECVD SILICON NITRIDE

F. S. HICKERNELL, H. D. KNUTH, and T. S. HICKERNELL
Motorola Government and Space Technology Group, Scottsdale, Arizona 85252
L. MANG
Motorola Phoenix Corporate Research Laboratories, Tempe, Arizona 85284

Abstract

The SAW characteristics of thin-film silicon nitride (Si_3N_4) on piezoelectric substrates of Y-Z and 64° Y-X lithium niobate, and 36° Y-X lithium tantalate have been measured in the frequency range from 30 MHz to above 1.0 GHz. Silicon nitride films 250, 500, and 1000 nm thick were deposited by plasma enhanced chemical vapor deposition (PECVD). The SAW velocity and propagation loss properties were measured using linear arrays of thin-film aluminum interdigital electrodes patterned on the upper film surface. The elastic constants of the films could be determined by fitting a theoretical velocity dispersion curve to the measured velocities for Rayleigh wave propagation. The SAW propagation loss values for the films on the different substrates were compared to the loss characteristics without the film layers. Temperature coefficient of frequency (TCF) measurements of the SAW velocity characteristics for the various film/substrate combinations were also made. The results of these measurements, comparative data on acoustic constants, and the performance enhancements for SAW devices which could be realized by the use of silicon nitride on piezoelectric substrates are discussed.

Introduction

Low-temperature plasma-enhanced chemical-vapor-deposited (PECVD) silicon-based dielectrics, such as silicon nitride (Si_3N_4), silicon dioxide (SiO_2), and silicon oxynitride have found wide application in microelectronic integrated circuits. The films are used for encapsulation, passivation, masking and patterning, and as dielectric platforms for thin-film resistors, capacitors, and inductors. Thin-film dielectrics can also be used to enhance the performance capabilities of surface acoustic wave (SAW) and bulk acoustic wave (BAW) devices. They can increase the coupling factor, improve temperature stability, and enhance wave propagation. Thin-film dielectrics can also be used as platforms for the fabrication of SAW and BAW devices on semiconductor substrates for integration with electronic circuitry. While the dielectric properties of films commonly used in semiconductor integrated circuit fabrication have been extensively investigated, there has been little data gathered on the acoustic properties of such films. Such data is essential in the design of SAW and BAW devices, such as resonators and filters, with improved performance characteristics.

Because silicon nitride film technology has matured in the semiconductor industry it seems reasonable to investigate the acoustic properties of Si_3N_4 for possible incorporation on SAW substrates. We have investigated the acoustic properties of thin-film PECVD silicon nitride on lithium niobate and lithium tantalate substrates using surface waves generated by a SAW interdigital transducer array capable of generating frequencies into the gigahertz region. Three different thicknesses of Si_3N_4 were deposited on the wafers. An aluminum linear interdigital transducer array was photo-lithographically patterned at the surface of the Si_3N_4 film. The transducer conversion loss, surface wave velocity, and propagation loss were calculated as a function of frequency from network analyzer insertion loss measurements and compared to those from substrates without the Si_3N_4 film. The velocity dispersion data were used to calculate the elastic constants of the films. PECVD Si_3N_4 forms amorphous layers, so its acoustic velocity properties can be characterized by its density and the compressional and shear elastic constants c_{11} and c_{44}.

Previous work has been reported on the SAW characteristics of PECVD films deposited on gallium arsenide [1-2], and Y-Z and 128° Y-X lithium niobate

0-7803-2500-1/95 $4.00 © 1995 IEEE

[3-5]. The work reported in this paper gives additional information on the SAW charactersitics of Y-Z lithium niobate with PECVD film layers and extends the SAW measurements to leaky wave propagation on 64° Y-X lithium niobate and 36° Y-X lithium tantalate.

Film Growth

PECVD silicon nitride films of three different thicknesses were grown on Y-Z lithium niobate, 64° Y-X lithium niobate and 36° Y-X lithium tantalate. Thicknesses of 250 nm, 500 nm, and 1000 nm were used. The PECVD process involved the reaction of two percent SiH_4 in a N_2 (135 sccm) carrier gas with NH_3 (2 sccm) in a N_2 (580 sccm) ambient. Deposition was performed on a large platten heated to 250 °C, at a chamber pressure of 900 milliTorr, with an RF power of 25 Watts to achieve a growth rate of 11 nanometers per minute. Because of the pyroelectric nature of $LiNbO_3$ and $LiTaO_3$ the wafers had to be brought to temperature from room temperature and cooled back to room temperature after deposition. The films had an index of refraction of approximately 1.98 as measured by ellipsometry. To facilitate the calculation of the two independent elastic constants, film densities were calculated from measurements of film masses and thicknesses. The mass change was determined by substrate mass measurement before and after Si_3N_4 deposition. Film thicknesses were determined by surface profilometry and by optical measurements and values given by the two methods agreed.

Measurement Technique

Several rows of thin-film aluminum interdigital transducers (IDTs) were patterned on each substrate with and without the silicon nitride films for the frequency and insertion loss measurements. The thickness of the aluminum film used was between 100 and 200 nm. The center-to-center spacing between transducers was 3.81 mm. Each transducer had 10.5 split-electrode pairs with a periodicity of 100 micrometers, and a 2.54 mm aperture. Each electrode within a transducer was 15 micrometers wide, and the space between electrodes was 10 micrometers. The transducer pattern facilitated the excitation of several higher harmonic waves and permitted delineation of an extended velocity and loss characteristic for several values of film-thickness to acoustic-wavelength ratio. For each sample the center frequencies of the fundamental and the higher harmonic responses for the observed Rayleigh and leaky wave modes were measured. An accurate determination of frequency for a given mode was used to calculate the phase velocity from the fixed wavelength periodicity of the interdigital transducer. Insertion loss was measured between selected transducer pairs of eight transducers in a single row, and the transducer conversion loss and propagation loss were determined as a function of frequency.

The calculated velocity dispersion data plotted as a function of film-thickness to wavelength ratio for Rayleigh-wave propagation could be fit with theoretical curves to determine the elastic moduli of the silicon nitride. Theoretical velocity dispersion characteristic for the Si_3N_4 layer on Y-Z $LiNbO_3$ substrate were calculated using software described and provided by Adler et al [6]. The material constants for the substrates used in these calculations were those given by Slobodnik, Conway, and Delmonico [7]. A relative dielectric constant of 7.5 was used for the Si_3N_4 film layers, and the density was calculated from the measurements of mass and thickness for the PECVD films. Elastic moduli were calculated by fitting the theoretical dispersion curves to the measured dispersion data. The mass-loading, electrical-shorting, and stored-energy effects of the aluminum transducer on SAW velocity dispersion were considered in the calculations.

Silicon Nitride on Gallium Arsenide

The acoustic properties of PECVD silicon nitride were previously evaluated on (001) (110) gallium arsenide along with other potentially important dielectric materials for use as platforms in the integration of acoustic wave devices on semiconductor substrates [2]. Gallium arsenide is a weak piezoelectric material with a low surface wave velocity. The presence of the silicon nitride on GaAs stiffened the surface and increased the surface wave velocity as a function of film-thickness to acoustic-wavelength ratio (t/λ). The two independent elastic constants were determined and the SAW propagation loss properties defined. On the GaAs there was a dependence of the elastic constants on film thickness and density. The observed elastic constants were in the range c_{11} = 170-190 GPa and c_{44} = 55 - 90 GPa. This places the Rayleigh wave velocity of the PECVD silicon nitride near 4400 m/s. The propagation loss decreased with increasing nitride thickness in the range of 200 nm to 1000 nm and was between 10 and 20 dB/cm at 1 GHz. The temperature

coefficient of frequency for a thickness to wavelength ratio of 0.2 was -36ppm/°C compared to -45ppm/°C for (001) (110) GaAs with no film. This basic acoustic information was used for comparisons made to the results of the measurements in this investigation.

Silicon Nitride on Y-Z Lithium Niobate

Y-Z lithium niobate has been used over the past several years for SAW device development because of its strong piezoelectric coupling factor (k^2 = 0.045) and low Rayleigh-wave propagation loss ($\alpha \cong 3.5$ dB/cm at 1.0 GHz). The presence of silicon nitride on the surface in an electrode configuration with the interdigital transducer at the film/substrate interface is calculated to increase the coupling factor by approximately fifty percent (50%) at a film-thickness to wavelength ratio of 0.1. The presence of the silicon nitride should also reduce the temperature coefficient of frequency thereby improving the overall SAW device properties of the substrate.

Figure 1 shows the Rayleigh-wave velocity dispersion characteristic for the three thicknesses (250 nm, 500 nm, and 1000 nm) of PECVD silicon nitride on Y-Z $LiNbO_3$ as calculated from frequency measurements between two adjacent IDT electrode patterns on the film surface. The SAW velocity rises from a value near 3440 m/s for zero film thickness to 3580 m/s for a film-thickness to acoustic-wavelength ratio (t/λ) of 0.15. Rayleigh waves were not observed beyond the 0.15 thickness to wavelength ratio indicating a coupling of the waves into bulk modes.

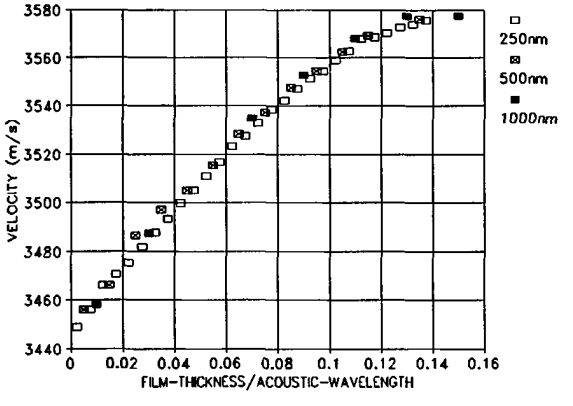


Fig. 1. Rayleigh wave velocity dispersion characteristic for the three thicknesses (250 nm, 500 nm, and 1000 nm) of PECVD silicon nitride on Y-Z $LiNbO_3$.

There is not a large deviation in velocity for the three film thicknesses indicating that the film density and two independent elastic constants are essentially the same. To determine the elastic constants which approximately fit this characteristic a velocity dispersion graph was developed as shown in Figure 2. The velocity dispersion for the 250 nm film is shown together with a theoretical plot of the metallized and unmetallized surface of Y-Z $LiNbO_3$ with a film of silicon nitride having a density of 2700 kg/m^3 and constants c_{11} = 160 GPa and c_{44} = 55 GPa. The fit is reasonable and the constants are close to those observed for silicon nitride films of similar thickness on GaAs. The indication is that the elastic constants of the PECVD deposited silicon nitride are not largely influenced by the substrate in at least this particular case.

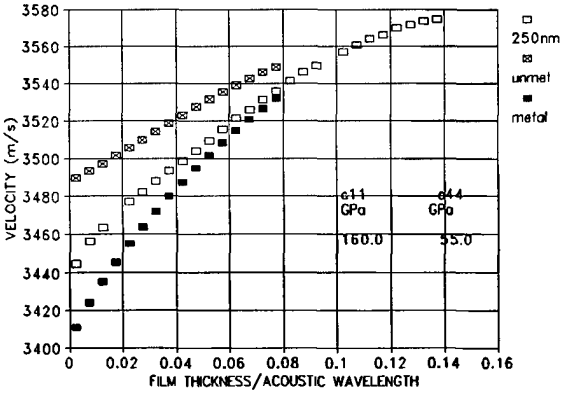


Fig. 2. The velocity dispersion for a 250 nm silicon nitride film together with a theoretical plot of the metallized and unmetallized surface of Y-Z $LiNbO_3$ for a film of silicon nitride having a density of 2700 kg/m^3 and constants c_{11} = 160 GPa and c_{44} = 55 GPa.

The measured Rayleigh-wave propagation loss is shown in Figure 3 for the three films over the frequency range from 35 MHz to over 1 GHz. The average loss trend is near a frequency squared dependence at the higher frequencies and in general the losses between different thicknesses tend to remain within a factor of 2 of one another at any given frequency. The up and down variations in loss may be due to the various combinations of longitudinal and shear motions which exist in the film substrate composite at particular wavelengths. Compared to a Y-Z $LiNbO_3$ substrate with no film using the same basic IDT pattern, the presence of the silicon nitride did not significantly affect the propagation loss. Laser probe measurements give propagation loss values for

Y-Z $LiNbO_3$ that are two to three times lower. A plot of the loss in dB per wavelength (dB/λ) at film-thickness to wavelength ratios in the .01 to 0.1 region gave values which were generally in the range of .003 to .005 dB/λ for all three film thicknesses. The SAW propagation loss for Y-Z and 128° Y-X lithium niobate reported by Hines, et al [5] had a similar range of values.

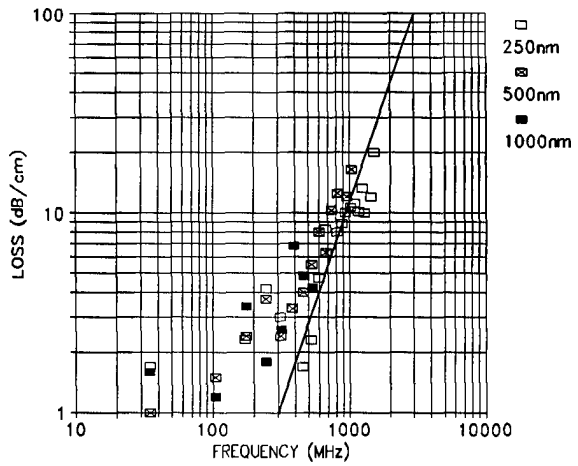


Fig. 3. Rayleigh-wave propagation loss in dB/cm for three film thicknesses of PECVD silicon nitride as a function of frequency with f^2 dependence (line) shown.

The temperature coefficient of frequency measured on the films showed a decrease in the TCF from a value near -90 ppm/°C with no film present down to approximately -70 ppm/°C at a film-thickness to acoustic-wavelength ratio of 0.1. This 25% reduction in TCF corresponds to the same percentage reduction observed with GaAs for a t/λ ratio of 0.2. Thus with a predicted 25% improvement in coupling factor, no substantial degradation in propagation loss, and an improvement in the temperature coefficient of frequency, the SAW properties of Y-Z $LiNbO_3$ have been improved by the presence of a silicon nitride film.

Silicon Nitride on 64° Y-X Lithium Niobate

The 64° Y-X lithium niobate substrate has been favored for wide bandwidth SAW devices because of its high coupling factor (k^2 = .11) in the leaky-wave SAW mode of propagation. The velocity is near 4600 m/s with a reported 0.036 dB/wavelength propagation loss. The TCF has been reported at -80 ppm/°C. The 64° cut also supports a Rayleigh wave with a velocity near 3700 m/s and a coupling factor of k^2 = 0.01. It is interesting to investigate the change in acoustic properties due to the presence of a silicon nitride film on the substrate for the two different propagation modes.

Figure 4 shows the velocity dispersion characteristic for leaky-wave and Rayleigh-wave propagation for two thicknesses of PECVD silicon nitride (500nm and 1000nm) as a function of film-thickness to acoustic-wavelength ratio. As would be anticipated the silicon nitride softens the surface for leaky-wave propagation and the velocity decreases after a small rise in frequency for t/λ ratios below 0.05. The silicon nitride stiffens the surface for Rayleigh-wave propagation. In both cases the velocity characteristic could be calculated out to t/λ ratios of 0.3 from the frequency measurements.

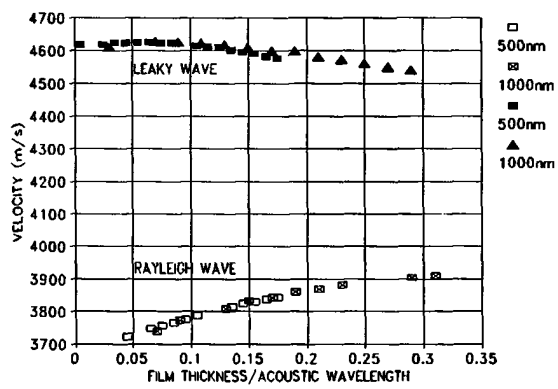


Fig. 4. Velocity dispersion characteristic for leaky-wave and Rayleigh-wave propagation for PECVD silicon nitride (500nm and 1000nm) as a function of film-thickness to acoustic-wavelength ratio.

The propagation loss characteristics in dB/cm are shown in Figure 5 for Rayleigh and leaky modes as a function of frequency. Shown are values for no film and a 1000nm film. The Rayleigh-wave mode shows a loss characteristic similar to Y-Z $LiNbO_3$ in frequency dependence with only a slightly higher loss with the 1000 nm film present on the surface. The leaky-wave has a very interesting characteristic which may be representative of the particular geometry of the transducers as well as whether a film is present or not. With no film present the loss rises slowly and levels in the 500 MHz region and shows a decrease at higher frequencies. The presence of the film gives a lower loss, particularly at the lower frequencies and then is very level from 500 MHz to 1.0 GHz. Looking at the loss characteristic in dB/λ as a function of film-thickness to wavelength ratio, the Rayleigh-wave is

increasing in a linear fashion while the leaky wave is decreasing, also in a linear fashion. For a film-thickness to acoustic-wavelength ratio of 0.2 the leaky-wave loss in dB/λ is less than 0.02 and the Rayleigh-wave loss is 0.006. Thus it appears that the presence of the silicon nitride film has decreased the leaky-wave loss and has not substantially affected the Rayleigh-wave loss.

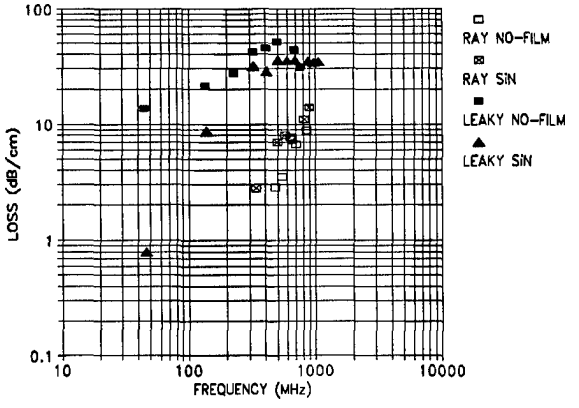


Fig. 5. Propagation loss characteristics in dB/cm for Rayleigh and leaky modes as a function of frequency for no film layer and a 1000nm PECVD silicon nitride film.

The temperature coefficients of frequency for Rayleigh and leaky wave propagation were observed to decrease in the presence of the silicon nitride film. The Rayleigh-wave TCF which was above -70 ppm/°C with no film present decreased to under -50 ppm/°C at a silicon nitride film-thickness to acoustic-wavelength ratio of 0.3. The leaky-wave mode had a projected TCF of near -55 ppm/°C with a no film (lower than values reported in the literature) and decreased to near -35 ppm/°C at the 0.3 ratio. Thus in both cases there was about a 20 ppm/°C drop due to the presence of the silicon nitride film.

The results of experimentation indicate that there would be an advantage to the use of silicon nitride in terms of temperature coefficient and propagation loss. In the leaky mode there is not a strong velocity change and the presence of the film reduces the propagation loss. The TCF is improved for both Rayleigh and leaky modes of propagation. The effect on coupling factor for different electrode configurations will need to be determined by measurement techniques different from those used in this investigation or by calculation using the measured elastic constants for the thin-film silicon nitride.

Silicon Nitride on 36° Y-X Lithium Tantalate

The 36° Y-X cut of lithium tantalate supports leaky-wave propagation with a free surface SAW velocity of 4212 m/s, a metalized velocity of 4112 m/s, a coupling factor of k^2 = 0.047 and a temperature coefficient of frequency in the -30 to -40 ppm/°C range. It has a very low leaky-wave propagation loss. The 36° Y-X cut has been used extensively for moderate bandwidth SAW devices requiring good temperature stability.

Silicon nitride films of 250 nm and 500 nm were deposited on the surface of 36° Y-X $LiTaO_3$. The leaky-wave velocity characteristic is shown in Figure 6 as a function of film-thickness to acoustic-wavelength ratio. In this case there is a distinct velocity separation between the two thicknesses by about 10 m/s whereas for the lithium niobate wafers the velocity characteristics were more closely grouped. The velocity characteristic is nearly a constant function of film-thickness to acoustic-wavelength ratio above 0.04. The 500 nm film appears to have slightly higher elastic constants. The silicon nitride velocity properties match well those of the 36° tantalate. There would be very little velocity dispersion for a SAW device on the 36° tantalate with a silicon nitride film.

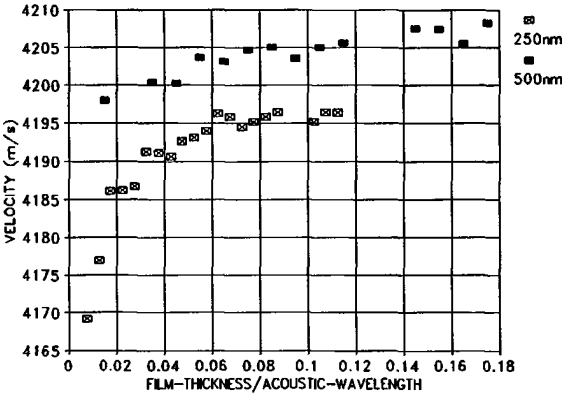


Fig. 6. Leaky-wave velocity characteristic, for 36° Y-X $LiTaO_3$ with two thicknesses of PECVD silicon nitride, as a function of film-thickness to acoustic wavelength ratio.

The propagation loss for the 250 nm film compared to no film is shown in Figure 7. The loss characteristic is a very slow changing function of frequency both with and without the presence of the silicon nitride film. The loss with the film appears only slightly higher at the lower frequencies. In fact well into the gigahertz region the loss is under 20 dB/cm with the nitride film.

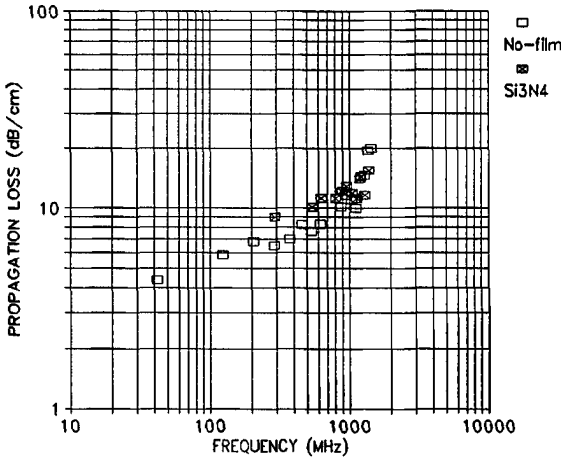


Fig.7. Leaky-wave propagation loss for 36° Y-X lithium tantalate with a 250 nm PECVD silicon nitride film compared to no film layer.

The propagation loss in dB/λ for 36° Y-X lithium tantalate under the particular transducer geometry had a value of .005 when measured at frequencies in the 400 to 700 MHz region. With the 250 nm film on the substrate this same value was measured in the 900 MHz to 1.3 GHz region. A graphic display of the measurements are shown in Figure 8.

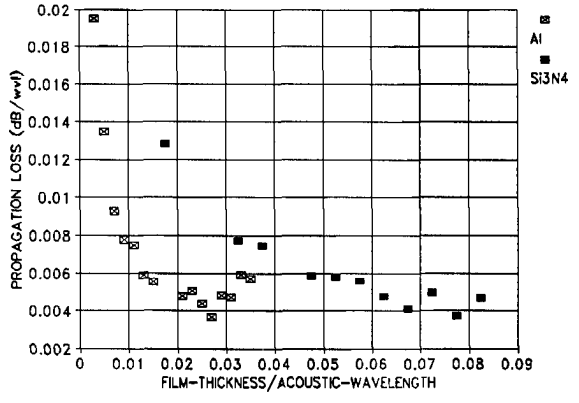


Fig. 8. Leaky-wave propagation loss in dB/λ for 36° Y-X tantalate with a 100 nm aluminum film and a 250 nm silicon nitride film as a function of film-thickness to acoustic-wavelength ratio.

The temperature coefficient of frequency for the 250 nm thick silicon nitride film is shown in Figure 9 and is interesting in the fact that beyond the film-thickness to acoustic-wavelength region of 0.04 the TCF remains very flat at between -30 ppm/°C and -32 ppm/°C. This corresponds well to the reported value of -32 ppm/°C for a metallized surface. There is a small drop in TCF for ratios near 0.02 before the TCF characteristic flattens. The 500 nm film showed similar TCF characteristics.

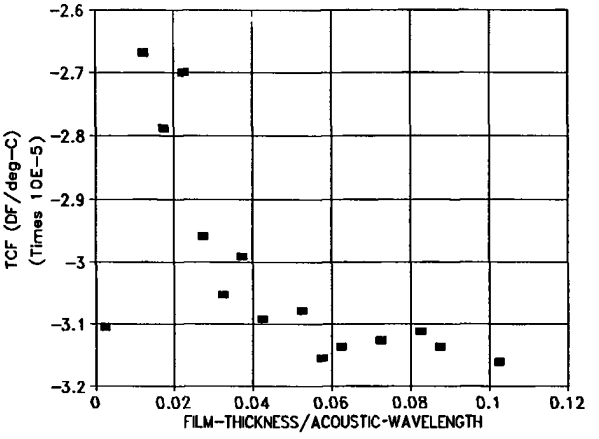


Fig. 9. The TCF characteristic for a 250 nm silicon nitride film on 36° Y-X lithium tantalate.

The silicon nitride film on 36° Y-X tantalate did not substantially alter the leaky wave acoustic propagation or TCF properties measured. A coupling factor advantage, if it is present, will have to be measured by other techniques.

Conclusions

The SAW properties of PECVD silicon nitride on Y-Z and 64° Y-X lithium niobate and on 36° Y-X lithium tantalate have been measured. The film thicknesses investigated were in the 250 nm to 1000 nm range. Velocity dispersion characteristics were calculated from measurements using a linear transducer array capable of generating surface waves in the frequency range from under 100 MHz to over 1.0 GHz. The velocity dispersion characteristics were small for leaky-wave propagation on the 64° niobate and the 36° tantalate. For Rayleigh-wave propagation the presence of the silicon nitride film did not substantially change the loss characteristic. For the leaky-wave propagation the film aided in lowering the propagation loss. The temperature coefficient of frequency was lowered by the presence of the silicon nitride for the niobate substrates and was unchanged for the tantalate substrate. Theoretical modeling indicates that the presence of a silicon nitride film on Y-Z lithium niobate would raise the coupling factor by 25 % at a t/λ ratio of 0.06 and the 128° Y-X lithium niobate

coupling factor would be increased increased by 40% at a t/λ value of 0.15 with an interdigital transducer at the film /substrate boundary.

Acknowledgment

The authors appreciate the technical assistance rendered by Ron Pennell in the film growth area, Claudia Jensen and Gloria Judge for transducer processing, Virginia Hernandez for wire bonding, and Bob Dablemont for assisting in the measurements.

References

[1] T. S. Hickernell, F. M Fliegel, and F. S. Hickernell, "The elastic properties of thin film silicon nitride," in *Proc. IEEE Ultrason. Symp.*, pp. 445-448, 1990.

[2] T. S. Hickernell and F. S. Hickernell, "Surface acoustic wave characterization of PECVD films on gallium arsenide," *IEEE Trans. Ultrason., Erroelec., and Freq. Cont.* vol. 42, pp. 410-415, 1995.

[3] K. R. Lee, D. C. Malocha, and K. B. Sundaram, "Investigation of PECVD silicon nitride films for surface wave devices," in *Proc. IEEE Ultrason. Symp.*, pp. 369-372, 1992.

[4] K. R. Lee, "Investigation of silicon nitride films by PECVD and applications to SAW devices," Ph.D. dissertation, University of Central Florida, Orlando FL 1992.

[5] J. H. Hines, D. C. Malocha, K. B. Sundaram, K. J. Casey, and K. R. Lee, "Deposition parameter studies and surface acoustic wave characterization of PECVD silicon nitride films on lithium niobate," *IEEE Trans. Ultrason., Ferroelec., and Freq. Cont.*, vol. 42, pp. 397-403, 1995.

[6] E. L. Adler, J. K. Slaboszewicz, G. W. Farnell, and C. K. Jen, "PC software for SAW propagation in anisotropic multilayers," *IEEE Trans. Ultrason., Ferroelect., and Freq. Cont.*, vol. 37, pp. 215-223, 1990.

[7] A. J. Slobodnik, E. D. Conway, and R. T. Delmonico, *Microwave Acoustics Handbook.* Air Force Cambridge Research Laboratories, 1973.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

LOW-LOSS SAW FILTERS USING THICKNESS DIFFERENCE TYPE OF IDT ON THE NSPUDT ORIENTATION SUBSTRATE

K.YAMANOUCHI, M.TAKEUCHI, H.ODAGAWA AND M.TANAKA

Research Institute of Electrical Communication,
Tohoku University, Sendai 980, Japan

ABSTRACT

The invention of the Natural Single-Phase Unidirectional Transducer (NSPUDT) in 1985 by P.V.Wright opened up a possible new class of low loss Surface Acoustic Wave (SAW) devices. The unidirectionality of the NSPUDT was obtained by using the crystal asymmetry of the substrate. Therefore, the reverse directivity transducers for low-loss SAW transversal filters using NSPUDT orientation of substrates are required. The reverse directivity of 2-finger/λ transducer is developed by using a different metallization on a transducer by P.V.Wright and grooves on quartz with the same meta-llization by C.S.Lam.

This paper reports a new method to forward and reverse the unidirectionality with the electrode thickness difference (ETD) type of UDT on the NPSPUDT orientation of substrates. The electrode widths of new IDT are about $\lambda/4$ for the narrow gap IDT and $\lambda/8$ for the $\lambda/8$ gap IDT with the short or open floating electrodes. The two transducers for forward and reverse directivities have almost the same mass loading effects. Therefore the center frequency shifts between two transducers due to IDT's electrode difference don't occur such as Wright's and Lam's methods. The calculation result of ETD-RDT (RDT: Reversal of Directivity IDT) for forward and reverse directivities about 10dB is obtained. The ETD-RDT is fabricated by the methods of Electro-Chemical Effects or Narrow Gap Anodic Oxidation Process. The experimental results show the large directivity and the low insertion loss filters.

1. INTRODUCTION

The frequency ranges of mobile communication systems are in GHz and in extending to 2~4 GHz. Moreover SAW devices require the frequency range around 10 GHz. Also, unidirectional SAW transducers are very important for high efficiency SAW devices, for example, low loss SAW filters and high efficiency convolvers, etc. Many efforts have been made to reduce the insertion loss of SAW filters, and UDT's filters are very important for digital communication systems because of their good phase characteristics.

It is well known that single phase unidirectional transducers (SPUDT), with high efficiency, operate due to the difference between the center of transduction and the center of reflection. In the development of SPUDTs some arrangements of new electrodes have proposed for the purpose of shifting the centers, for example, floating electrode[1], different width electrode[2] and different thickness or material electrode[3].

In 1985, Wright[4] reported the invention of the Natural Single-Phase Unidirectional Transducer (NSPUDT). In contrast to the shifting of reflection centers by conventional SPUDT, the NSPUDT is a regular $\lambda/4$ finger transducer which uses the crystal asymmetry to shift the reflection centers to attain a $\lambda/8$ phase offset. Since the unidirectional property is inherent to the substrate orientation, all reported low-loss SAW transversal filters using NSPUDT had to use a bidirectional transducer to receive the surface wave. It was suggested that by using a different metallization, the receiving transducer could be made unidirectional in the opposite direction. Recently, the experimental method to reverse the unidirectionality of the NSPUDT on quartz with buried electrodes by using a simple 2-finger/λ transducer have been demonstrated by Lam(5).

In this paper, a new method to forward and reverse the unidirectionality with the electrode thickness difference(ETD) on the NPSPUDT orientation of substrates is described.

0-7803-2500-1/95 $4.00 © 1995 IEEE

2. OPERATION MECHANISM AND ELECTRODE CONFIGURATION

Figure 1 shows the transduction(Tc) and reflection(Rc) centers on bi-directional substrate(BDT) and NSPUDT substrate. The Tc and Rc are same position on BDT. Therefore SAWs propagate to the both directions. On the other hand, the Tc has $\lambda/8$ phase shift to the Rc by the crystal asymmetry on NSPUDT substrate. The forward directivity is obtained, as shown in Fig.1, in the case of -90 ° reflection.

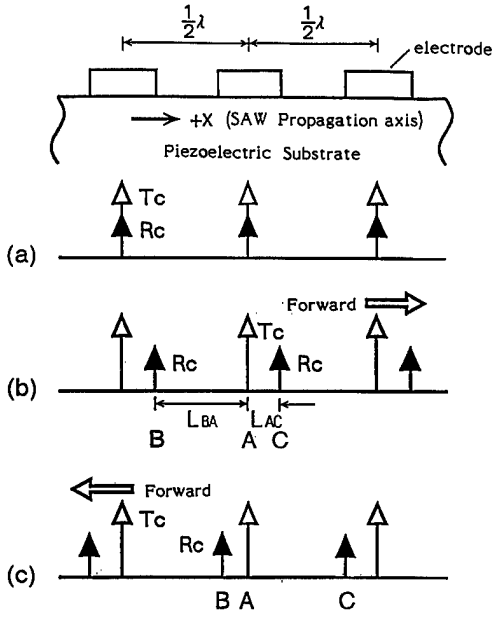


Fig.1 Operation Mechanism of NSPUDT
(a) Relation between Transduction and Reflection Centers on Bi-directional Transducer(BDT) Substrate, (b), (c) Relations between Transduction and Reflection Centers on NSPUDT Substrate

Figure 2(a) shows the Tc has $\lambda/4$ phase shift to the Rc on BDT substrate. In this case, the SAWs propagate to the both directions. On the other hand, the $\lambda/8$ phase shift between the Tc and Rc is given on NSPUDT substrate, as shown in Fig.2(b) and (c). The configuration of Electrode Thickness Difference (ETD) type transducers on NSPUDT substrate are shown in Fig.3 (a),(d) and (e). Figure 3(b) and (c) show bi-directional transducers on BDT substrates. The total reflections of (Rc2-Rc1) are given by the electrode

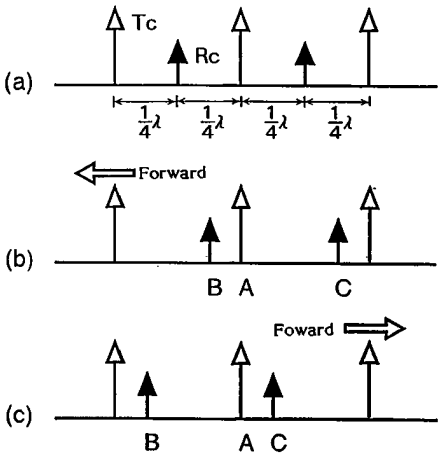


Fig.2 Operation Mechanism of Reverse Directivity Transducers on NSPUDT Substrates
(a) Relation between Transduction and Reflection Centers on BDT Substrate, (b), (c) Relation between Transduction(Tc) and Reflection(Rc) Centers on NSPUDT Substrate

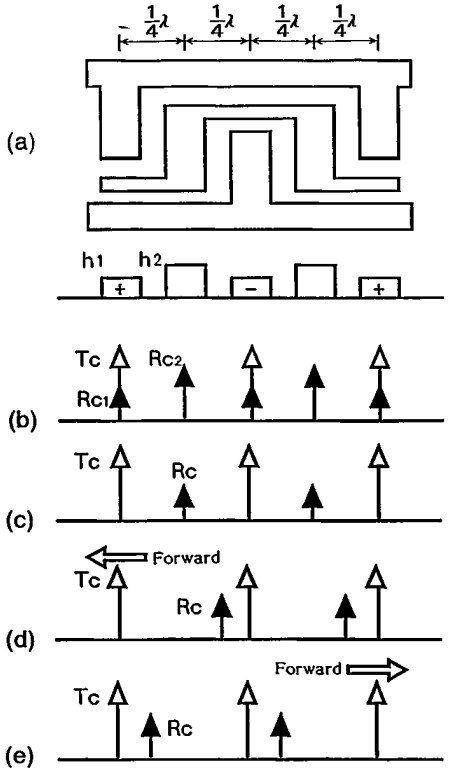


Fig.3 Configuration of Reverse Directivity Transducers of Thickness Difference Type on NSPUDT Substrate (ETD-RDT)
(a) Electrode Configuration
(b),(c) On BDT Substrate (Rc=Rc2-Rc1)
(d),(e) On NSPUDT Substrate

thickness difference between the transduction and reflection electrodes. The forward directivity in Fig.3(d) is given by thick Rc-electrode, if the reflection of the thicker electrode has larger than thiner one. On the other hand, the forward directivity in Fig.3(e) is given by thick Tc-electrodes. The electrode widths of new IDT are about $\lambda/4$ for the narrow gap IDT and $\lambda/8$ for the $\lambda/8$ gap IDT with the short or open floating electrodes.

Figure 4 shows the low insertion loss filter configuration with ETD and ETD-RDT on NSPUDT substrate. The two transducers for forward and reverse directivties have almost the same mass loading effects. Therefore the center frequency shifts between two transducers due to IDT's electrode difference don't occur such as in Wright's and Lam's methods.

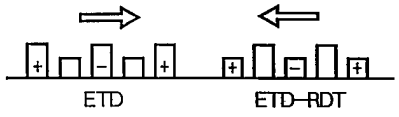


Fig.4 Example of Configuration of Sending and Receiving (ETD) Transducers

3. THEORETICAL AND EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, the NSPUDT operation with devices of a NSPUDT-BDT configuration is shown. To assess the directivity, the actual SAW chip has a BDT-NSPUDT-BDT structure as showen in Figure 5. The BDT has split finger (4-finger/λ) of 1 μ m feature size (i.e.λ=8 μ m). The NSPUDT with narrow gap electrodes is fabricated by the anodic oxidation techniques[6]. The electrode pattern widths of $\lambda/4$ are about 4 μ m. The propagation direction is 25 ° off the X-axis (toward the minus Z-axis) on 42 ° rotated Y-cut quartz. The fingers are slanted by 5.4 ° to compensate for the estimated power flow angle. They are perpendicular to the propagation direction and the bus-bars are parallel to the 30.4 ° direction.

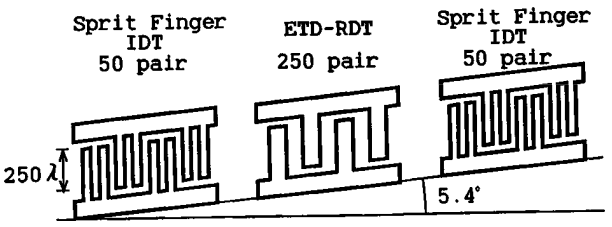


Fig.5 IDT Structure for Assessment of Directivity

The fingers are 750Åthick aluminum surface electrodes. The NSPUDT and BDT are both uniform and with no apodization. Figure 6 shows the frequency responses of the NSPUDT-BDT (forward direction) and the BDT-NSPUDT (reverse direction with the Rc thicker electrode configuration in the pass-band) when these sections were connected and measured in a 50 Ω system. The directivity is 10 dB. Figure 7 shows the frequency responses of the BDT-NSPUDT(forward direction) and the NSPUDT-BDT(forward direction with the Tc thicker electrode configuration). Figure 8 shows the low-loss filter response in the configuration of Fig.4. The thicker and thinner electrode thicknesses are 1600 Å(H/λ=0.02) and 800 Å(H/λ=0.01). When both transducers of the two NSPUDT configuration are roughly tuned to minimum insertion loss, 8.7dB insertion loss at 404 MHz was attained. Also, Fig.9 shows a filter response of 8.7dB insertion loss at 1GHz range when ideal matching circuits are connected with the experimental result.

As the unidirectional property of the NSPUDT is inherent to the crystal asymmetry, one possible way to reverse the effect is to reverse the sign of the reflectivity of the fingers. Wright briefly demonstrated that gold metallization on quartz would proved the reversal of unidirectionality. Lam also showed that grooves on quartz presented opposite reflectivity. In our new ETD structure on NSPUDT, the two transducers for forward and reverse directivties have almost the same mass loading effects. Therefore the center frequency shifts between two transducers due to IDT's electrode difference don't occur such as Wright's and Lam's methods.

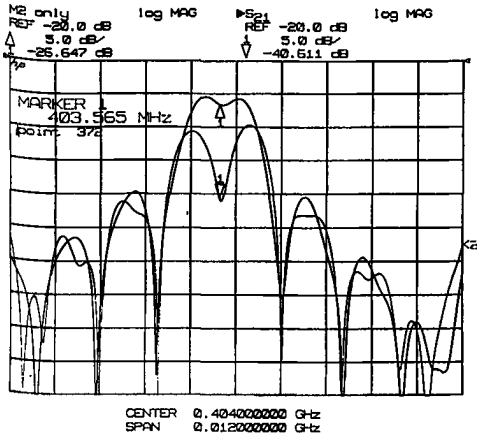


Fig.6 Experimental Results of ETD-RDT

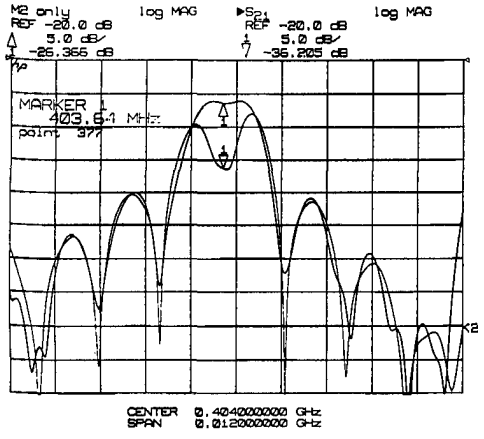


Fig.7 Experimental Results of ETD-NSPUDT

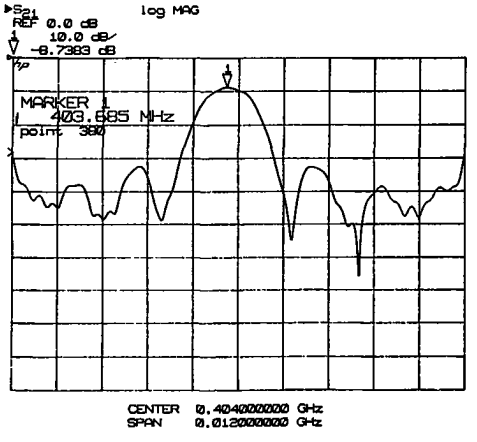


Fig.8 Frequency Responce of 400 MHz Filter

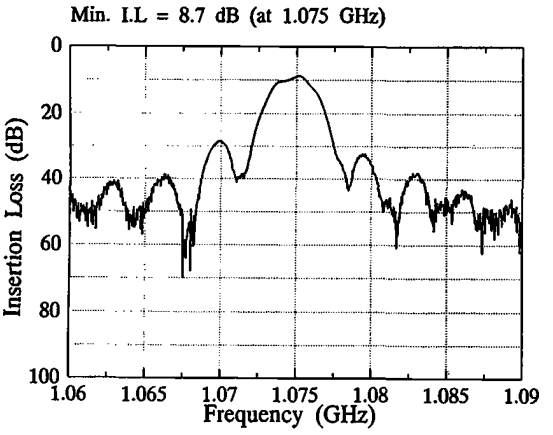


Fig.9 Frequency Responce of 1 GHz Filter

4. CONCLUSION

We have successfully demonstrated the fabrication of a 8.7dB, 404 MHz low-loss SAW transversal filter using two ETD types of transducers on the NSPUDT orientation of quartz with the same metallization and different thickness in one transducer. By the electrode thickness difference, we were able to forward and reverse the unidirectionality on the NSPUDT orientation of quartz. We are now investigating a higher frequency and lower insertion loss filters using anodic oxidation techniques.

REFERENCES

[1] K. Yamanouchi, H. Furuyashiki: "New Low-Loss SAW Filter Using Internal Floating Electrode Reflection Types of Single-Phase Unidirectional Transducer", Electron. Lett., 22nd, vol.20, No.24, pp.989-990, 1984

[2] K. Hanma and B. J. Hunsinger: "A Triple Transit Suppression Technique", IEEE Ultrason. Symp. Proc. pp.328-331, 1976.

[3] C. S. Hartmann, P. V. Wright, R. J. Kansy and E. M. Garber: "An Analysis of SAW Interdigital Transducers with Internal Reflections and the Application to the Design of Single-Phase Unidirectional Transducers", IEEE Ultrason. Symp. Proc. pp.40-45, 1982

[4] P.V.Wright, "The Natural Single-Phase Unidirectional Transducer: A New Low-Loss SAW Transducer",IEEE Ultrason. Symp. Proc. pp.58-63, 1985

[5] C.S. Lam and D. Gunes, "A Low-Loss SAW Filter Using Two-Finger Per Wavelength Electrodes on the NSPUDT Orientation of Quartz", IEEE Ultrason. Symp.Proc. pp.185-188, 1993

[6] K. Yamanouchi, T. Meguro and K. Matsumoto, "Surface Acoustic Wave Unidirectional Transducers Using Anodic Oxidation Technology and Low-loss Filters", Electron. Lett., vol.25, NO.15, pp.958-960, 1989

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

MODULATION TYPE DUAL-MODE OSCILLATOR INTENDED FOR MICRO-CHIP REALIZATION

Anatoly V. Kosykh, Sergei A. Zavyalov.

Omsk State Engineering University
11 Mira avenue, 644050, Omsk, Russia

ABSTRACT

Dual-mode oscillator is an oscillator based on simultaneous excitation of resonator two own modes. SC-cut resonators with B- and C- modes excitation are used as dual-mode most often. Each resonator mode has its own frequency vs. temperature dependence. When a resonator is being excited on two modes simultaneously, the mode frequency ratio is uniquely determined by temperature. This unique relation allows to use C- and B- mode frequency ratio as a resonator proper thermosensor. This thermosensor has full spatial coincidence with a vibrated resonator plate and, therefore, it is an ideal resonator temperature measuring device. Due to quartz nature such sensor is very precise and stable. From the above reasoning dual-mode resonator is an ideal base for thermocompensated (TCXO) and ovenized (OCXO) oscillators, in the cases when we need to measure the quartz plate proper temperature with maximum accuracy.

The concepts of dual-mode oscillators design have been discussed in this work. Considerable attention has been given to the constant B-mode excitation level keeping under external factors effects and resonator motional resistance variation. The micro-chip realizable dual-mode oscillator circuit is given. To demonstrate the validity of ideas the interim prototype of micro-chip dual-mode oscillator have been investigated. The main result of this work is that we show that dual-mode oscillators could be realized as a small-size micro device.

1. BASIC CONCEPT

What is dual-mode oscillator?

Dual-mode oscillator is an oscillator based on simultaneous excitation of resonator two own modes. There are known dual-mode oscillators based on SC-cut resonators of B- and C- modes excitation and of the 1-st and the 3-rd mechanical harmonic excitation. Each of these oscillators has own merits and demerits and the utility of one or other circuits is dictated by the function of the oscillator under concern. The advantages of B- and C- modes excitation for DTCXO purposes were carefully discussed in [1].

Why dual-mode excitation?

Each resonator mode has its own frequency vs. temperature dependence. When a resonator is being excited on two modes simultaneously, the mode frequency ratio is uniquely determined by temperature. This unique relation allows to use C- and B- mode frequency ratio as a resonator proper thermosensor. This thermosensor has full spatial coincidence with a vibrated resonator plate and, therefore, it is an ideal resonator temperature measuring device. Due to quartz nature such sensor is very precise and stable.

From the above reasoning dual-mode resonator is an ideal base for thermocompensated (TCXO) and ovenized (OCXO) oscillators, in the cases when we need to measure the quartz plate proper temperature with maximum accuracy.

Why dual-mode oscillators are not widely used?

In spite of its apparent advantages dual-mode oscillators are not widely used, which is caused by the three reasons:

A. Up to recent time the "ungaped" B-mode excitation in wide temperature range was a problem of great concern.
B. The worseing of output signal spectrum as compared to single-mode oscillators is occurred. There are components of B-mode frequency and B-and C-mode combination components in output C-mode spectrum.
C. The existing dual-mode oscillator circuit versions are of great complication. Actually, the oscillator

0-7803-2500-1/95 $4.00 © 1995 IEEE

involving up to 10 transistors and 50 resistors and capacitors [1] or to 5-6 inductance coils [2,3] is not very attractive for manufacturing.

What is the situation with dual-mode oscillators by now?

Due to intensive research investigations that were performed in the Omsk State Instrumentation Institute and Omsk Piezoengineering Company Ltd. [4,1] by now we have "ungaped" SC-cut resonators suitable for dual-mode excitation in oscillator circuit. Besides that we have some circuit ideas [1], that allow to create relatively reliable dual-mode oscillators. But those ideas have not yet been carefully investigated and their realization is not suitable for micro oscillators.

What is offered in this work?

The goal of the work submitted here is development of suitable for production dual-mode oscillators of small size intended for DTCXO or OCXO. The following criteria have been observed during oscillator development:

- minimum power consumption;
- minimum sensitivity to supply voltage variations;
- minimum of spurious spectrum components in output C-mode signal;
- the absence of large size elements unaffectable to minimization (e.g tuned inductance coils);
- the stable dual-mode oscillation with range adjustable B-mode level ability must be provided by the oscillator circuit;
- possibility of micro-chip realization.

Inasmuch as micro-chip orientation was accepted, transistor number in electronic circuit is not essential. Such line of attack opens a new approach to oscillator circuit design: the best circuit is not a one of transistor minimum but the simplest in micro-chip realization.

2. EXPERIMENTAL RESULTS

A modulation method of dual-mode excitation was used in this work as basic one. We started from the ideas submitted in [5,6]. When the modulation method is used the quartz resonator is excited on a single C-mode by usual oscillator circuit. The B-mode is used in this case as a phase or amplitude circuit included in separate controlling loop. When the base frequency (of C-mode) is 10 MHz the B-mode resonance frequency is near 10,9 MHz. Of course, it is possible to effect the oscillator by a signal of B-mode frequency, but this needs large power consumption (power consumption is proportional to signal frequency) and hampers output signal of B and C modes frequencies separation. The more rational way is using a lower frequency Fb-c which causes the Fb combination component under nonlinear interaction in oscillator with a oscillated C-mode. It is rational to use the Fb-c frequency signal as an output thermosignal. The equivalent diagram of this process is shown in Fig. 1.

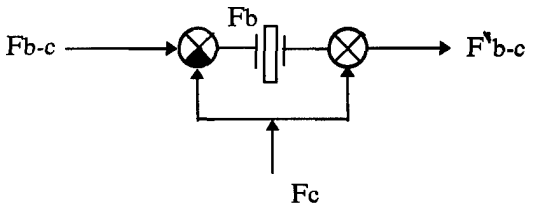


Fig.1. This Figure shows as B-mode may be used for Fb-c frequency filtering

The output signal Fb-c will differ from the input one by the amplitude and the phase. It is possible to realize such oscillator circuit configuration in which dual frequency conversion is performed automatically due to transistor nonlinearity (Fig. 2).

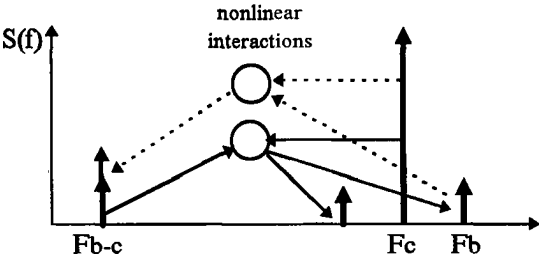


Fig.2.. The process of nonlinear interactions in modulated oscillator

In this case a spectrum component Fb resonating in quartz is indicated by the external frequency Fb-c. Interacting with C-mode oscillation this frequency produces a Fb-c mode component with different from initial signal phase. The structural diagram of modulation type dual-mode oscillator is shown in Fig.3.

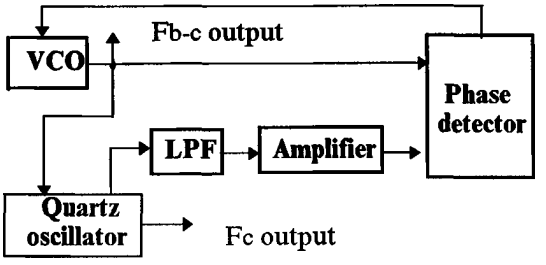


Fig.3. Modulation type dual-mode quartz oscillator block diagram

A RC-oscillator may be used as a VCO with the tuning range of 30-40 kHz. But far better parameters may be achieved by using a piezoceramic resonator on 900 kHz. Meanwhile it is better to use similar resonator instead of low pass filter for filtering the Fb-c signal.

Phase and amplitude characteristics of open loop modulation type system was investigated in circuit Fig.4

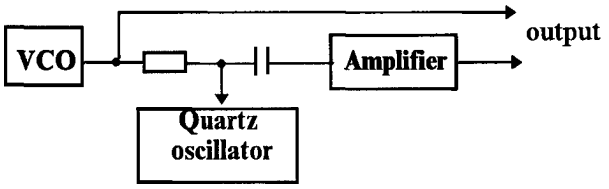


Fig. 4. Measuring system Block diagram of amplitude and phase vs. modulation frequency dependence

Figures 5 and 6 shows the phase (for different oscillator points of inclusion) and the amplitude characteristic respectively.

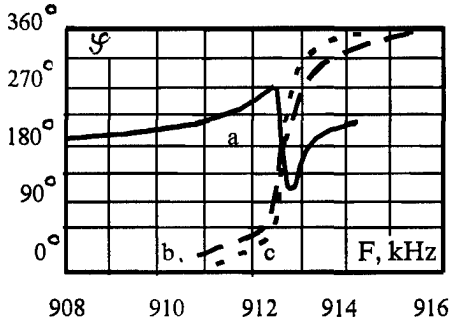


Fig. 5. Frequency shift vs. modulation frequency

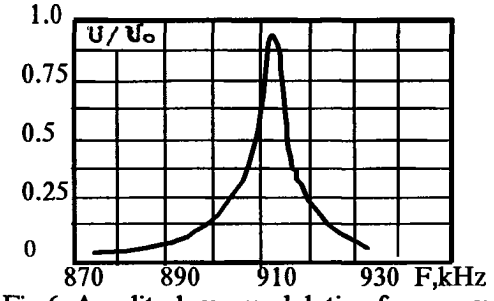
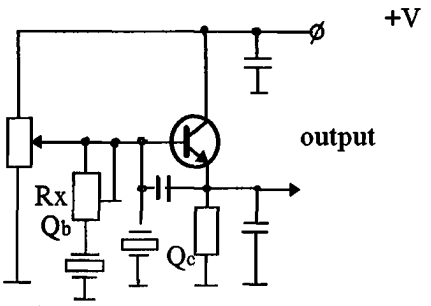


Fig.6. Amplitude vs. modulation frequency

It is seen from the figures that it is possible to provide a unambiguous phase characteristic of the system and it is possible to adjust the length of the linear part of this characteristic.

The problem of great concern for dual-mode oscillator is guaranteeing a stable B-mode excitation in wide temperature range. The matter is that B-mode motional resistance may change its volume several times when B-mode frequency is coincidence with other resonator anharmonic frequencies. The intermode energy transfer is being happened in this situation which cause the B-mode activity variation. To simulate this process we have investigated the oscillator circuit with two separate mono-frequency resonators on Fb and Fc frequencies with serial to resonator Fb connected resistor Rb (Fig. 7)



FQ_c=10.235 MHz, Rq=18 Ohm,
FQ_b=11.165 MHz, Rq=30 Ohm

Fig.7. Oscillator circuit for Rb motional resistance variation simulation

By varying Rx value we have obtained admissible values of Rb variation (Fig. 8). It should be mentioned that modulation method of dual-mode excitation provides a limitation less then other methods [1].

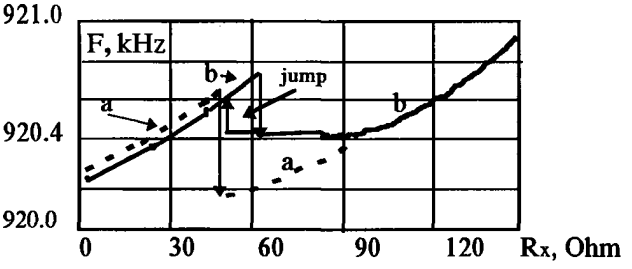


Fig.8. Admissible values of Rx variation
a- S_{VCO}=10 kHz/V b- S_{VCO}=1 kHz/V

The output signal spectrum component level is dependent on modulation signal value and on the regime of oscillator transistor. Minimal available discrete components level is limited by the oscillator and amplifier noise properties and by the necessity to have some safety margin. (When B-mode activity may undergo changes). Actually we have achieved B-mode level less than -60 dB relative to output C-mode signal. (Fig.9)

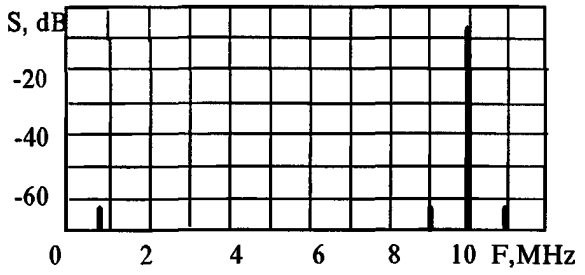


Fig.9. Output signal spectrum diagram

By using a piezoceramic band pass filter this level may be reduced to -90...-100 dB.

The circuit pictured in Fig. 3 may be easily realized as a analog ASIC matrix. Here the well known analog circuit technics for UHF FM receivers may be used. IC units packaged in MC3361 pack containing an oscillator, a limiting amplifier, a phase detector etc. were actually used in our experiments. These circuit solutions accompanied with the necessary units commutation can be transferred into ASIC matrix.

CONCLUSIONS

The above considerations permit to state the following conclusions:

1. There are TD-cut crystal resonators [7] (SC-cut Russian analog) that can be excited simultaneously at B- and C- modes in wide temperature range.
2. A modulation method of dual-mode excitation makes possible to lower the requirements to B-mode parameters stability.
3. It is advisable to use low-price miniature piezoceramic resonators and filters as filtering element.
4. B-mode motional resistance variations weakly effect the output thermo-sensitive frequency Fb-c
5. Modulation characteristics of a crystal oscillator are strongly dependent on the way (and the place) of external action.
6. The circuit of dual-mode modulation type oscillator is suitable for implementation in the technique of ASIC.

ACKNOWLEDGEMENT

The authors express their appreciation to Dr. I. Abramson for continuous letting dual-mode resonators of more and more increasing quality.

REFERENCES

1. Kosykh A.V., Abramson I.V., Bagaev V.P. Dual-Mode Crystal Oscillators With Resonators Excited on B and C Modes. Proceedings of 48-th Annual Frequency Control Symposium (A.F.C.S.), 1994, p. 578- 586.
2. Shodowski S.S. Resonator Self-Temperature-Sensing Using a Dual-Mode Crystal Oscillator. Proceedings of 43-rd A.F.C.S., 1989, p. 2-7.
3.Benjaminson A., Stalling S. C. A Microcomputer-compensated Crystal Oscillator Using a Dual-Mode Resonator. Proc. 43-rd A.F.C.S., 1989, p. 20-26.
4. Abramson I. Two-mode quartz resonator for digital temperature compensated quartz oscillators. Proc. 46-th A.F.C.S., 1992..
5. Шмалий Ю. С., Евдокименко Ю. И. Модуляционный метод адаптации опорных кварцевых генераторов. (Modulation Method of Adaptation of Reference Crystal Oscillators). В кн. Стабилизация частоты. Часть 1, ВИМИ, стр. 75-78, 1989 (USSR).
6. USSR Invention Certificate (Patent) 1640812 A2. Many-frequency Crystal Oscillator. Kolpacov F., Ivanov N., Khutornenko S., Mescheriakov O., Rack I. , 1991.
7. Invention certificate N 243977 (USSR) Oscillator Frequency Stabilization Device. Dikidji A., Dikidji L., Ivlev L. 1969.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DESIGN OF A 100KHZ TO 70MHZ, 2.7 TO 7 VOLT, CLOCKING OSCILLATOR CHIP

Hedley Rokos

C-MAC Quartz Crystals Ltd.
Edinburgh Way, Harlow CM20 2DE, England

The design of a high performance clocking oscillator chip is presented. The chip replaces an existing chip in an obsolete 2 μm process, and is required to be form compatible. Advantage has been taken of the replacement 1.2 μm process to extend the application range, both as regards its supply requirements, and the frequency range. A Colpitts oscillator with ALC is used to provide fast start-up, and a consistent oscillating condition. The sinusoidal output is AC coupled to an inverter chain, which is biased for optimum mark-space ratio. The start-up condition and division ratio are set via three level inputs, and the output is designed to avoid ringing. Die area is virtually unchanged at about 0.5 mm^2.

Background and Objectives

Clocking oscillators need to be reliable, fast starting, sufficiently precise, and low cost. For portable applications, operation at 3V nominal supply with low current drain is also required. Small size and flexibility of packaging are also becoming ever more important.

We have long used a single chip solution for many of these oscillators. The chip was manufactured in 2 μm CMOS, and was usable to above 25 MHz. The process was discontinued around two years ago, so a replacement is required. This is implemented in 1.2 μm CMOS, and is designed to operate over an extended voltage and frequency range.

Because crystals for these oscillators are not pre-tested, it is important that the oscillator will start with the full range of manufactured crystals: this both improves device yield, and minimises the occurrence of oscillators which only start intermittently.

The original chip was also used in a number of VCXO's, in spite of its oscillation condition not being ideally matched to such application. The replacement chip is therefore designed to operate closer to series resonance. A further issue for VCXO's could be parametric interaction with overtone inharmonics. Thus, the oscillator current should have as low a distortion as practical.

The original chip had rather fast transitions. This caused ringing at the ends of some tracks. Although the levels involved are not believed to have caused any practical problems, controlled transitions are generally considered to be desirable in any new design.

Oscillator design

The oscillator should have fast start-up, low steady-state supply requirements, and provide a sinusoidal oscillation current to the crystal. For compatibility with existing substrate lay-outs, it is desirable that the crystal can be connected between the supply, and a single oscillation-maintaining pin. These criteria are generally well met by a Colpitts circuit operating in current starvation. In the 1.2 μm CMOS process, however, we find that the high gm's needed for rapid start-up would require the use of excessively large devices if the output level is to be kept low enough for 2.7 Volt operation. This is a particular problem at high frequency. Accordingly, the output level is controlled by reducing the gm of the maintaining transistor (ALC). This has the additional advantage of minimising the steady-state current consumption.

The design of the ALC circuitry needs to take account both of the ALC loop itself, and of parasitic coupling of the gm controlling current into the level-sensor. This can be quite tricky for the simple arrangements which are possible on a small integrated circuit. Our design compromise generates the ALC signal from a small limiting current, which virtually guarantees stability of the ALC loop itself. The parasitic loop is stabilised by pole cancellation and by restricting the loop gain.

The negative resistance at start-up is selected according to the desired frequency range by wire-bonding. In addition to allowing start-up to be optimised, use of the appropriate gain allows overtone selection using a simple decoupled external resistor. The start-up control circuitry uses a band-gap style circuit to select one of four drive levels in the maintaining amplifier. As previously outlined, device size and speed restrictions mean that the oscillation maintaining transistor cannot be designed always to operate below threshold. Accordingly, the start-up control circuit is optimised so that gm of the MOSFET in the Colpitts circuit is substantially

0-7803-2500-1/95 $4.00 © 1995 IEEE

independent of the device's operating region, which allows the characteristics to be relatively stable with process and temperature variation. As a high negative resistance is required to cover the full range of crystals, the overhead to provide rapid start-up from noise is quite small. This avoids any requirement for provision and control of a start-up pulse.

The design presents a series capacitance of twenty pF to the crystal throughout the frequency range, and the oscillating voltage is around 1.5 Volt pk-pk. This provides an acceptable crystal current both for clocking, and for many VCXO applications, throughout the 100kHz to 70MHz frequency range.

[Note on limiting: as well as generating unwanted harmonics, it is found that applying voltage limiting around the crystal can result in quite variable operating conditions. This is because, under limiting conditions, the effective series capacitance presented by the maintaining amplifier is approximately equal to the effective parallel capacitance of the oscillator under start-up conditions. Where the design is optimised for rapid starting, this series capacitance can become quite small, and strongly dependant on process parameters other than the tightly controlled gate-oxide capacitance.]

Waveform Conversion

The sinusoidal output from the oscillator is AC coupled to an inverter chain, whose input is biased to generate a 50:50 mark/space ratio. This signal goes to a programmable divider (1 to 32), and thence to the output section.

The output drivers ensure that the two output transistors do not conduct simultaneously. Using a combination of progressive driver application and feedback, the output rise and fall times are also maintained at around two nano-seconds for loads between 5 and 50pF. This gives acceptable ringing on most boards, without excessively constraining the maximum frequency at which a good mark-space ratio can be provided.

The output can be set to "tristate" to allow widest product compatibility.

Ancillary Circuits

The ancillary circuits for selecting the operating conditions are required to occupy the minimum possible area. Accordingly, the nine selectable conditions of gain and frequency division are selected via three-level inputs, which are biased at a central level in the absence of wire bonds. Parasitic input capacitance is split evenly between the supply rails to ensure that operation is undisturbed during supply transients.

Conclusion

Use of a 1.2 μm process in the replacement for a 2 μm original allows improved control of the oscillation conditions, as well as extending the application range, while maintaining die size and form compatibility.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

CMOS CRYSTAL OSCILLATORS BASED ON CURRENT CONVEYORS

J. POPOVIĆ[1], A. PAVASOVIĆ[2], Z. ŽIVKOVIĆ-DŽUNJA[3] and D. VASILJEVIĆ[1]

[1]Faculty of Electrical Engineering, Un. of Belgrade, B. Revolucije 73, P.O.Box 816, 11001 Belgrade, Yugoslavia
[2]Lola Institute, Kneza Višeslava 70A, 11000 Belgrade, Yugoslavia
[3]Faculty of Technical Sciences, Fruškogorska 11, 21000 Novi Sad, Yugoslavia

Abstract

Analysis and design of RLC and crystal oscillators based on current conveyors, in 1.5μ-CMOS integrated technology, are described. Maximal operating frequencies are 350MHz and 50MHz, respectively.

Introduction

From the moment of invention of the operational amplifier (opamp), signals on which the processing was done in electronic circuits were usually voltages. The transistor is in essence a voltage controlled current source. A series resistor is needed at the output of the circuit in order to transform the current into a voltage. Thus, an additional time constant is introduced, possibly affecting the band-width of the whole circuit. This effect is not of significance for frequencies up to a few megahertz, and for a long period of time there was no incentive to replace the opamp by some other circuit. The voltage processing of signals was widely accepted. However, with a rapid move to higher frequencies, the use of opamps became a limiting factor.

Analog circuits operating in the current mode are suitable for realization in integrated technology. They have a significantly wider bandwidth than circuits operating in the voltage mode when transistors with the same characteristics are used. The current conveyor is the basic building block for current mode processing. It can be used for realization of negative impedance converters (NICs) with current or voltage controlled negative input resistance. Such NIC circuits have a broad bandwidth and a static characteristics whose shape can be easily modified to the optimal form for oscillator under design. Therefore, NIC circuits are suitable for design of the active part of the single pin oscillators with inductive or capacitive time network or with a quartz crystal unit [1].

In the first Section of this paper, the analysis of oscillators based on current conveyors is presented. The realizations of such RLC and quartz oscillators are presented in the second Section.

The Oscillator Analysis

The three main parts of the relaxation oscillator model are the active part with the parasitic elements, the bias circuit and the timing circuit.

The quartz crystal unit is used as a timing circuit in crystal oscillators, which analysis is known in literature [1]. The LC timing network can also be used in relaxation oscillators analyzed in this Section. Also, each crystal oscillator includes a parasitic RLC oscillator due to the crystal parallel capacitance.

The active part of the oscillator circuit determines the oscillator characteristic and conditions for timing network operation [2]. Its parasitic elements provide fast changes of relaxation oscillator voltage and current wave-shapes [2].

The bias circuit determines the position of the oscillator singular point (quiescent operating point) which is important for starting of oscillations [2].

The negative impedance converters based on current conveyers are convenient for realization of the active part of the relaxation oscillators in single-pin configuration.

The NIC is a two-port system which can have an "S" shaped voltage controlled ($i=\phi(v)$) or an "N" shaped current controlled ($v=\varphi(i)$) current-voltage characteristic.

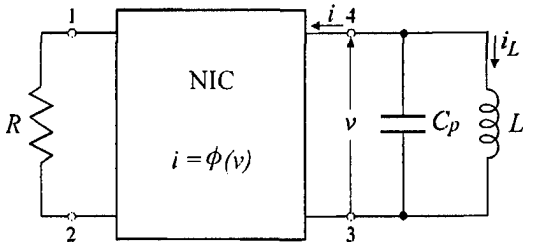


Figure 1. Oscillator with voltage controlled NIC.

The NIC implemented with a translinear current conveyor can have both types of current-voltage characteristics, depending on the location of the resistor

0-7803-2500-1/95 $4.00 © 1995 IEEE

R [1]. For the oscillator with voltage controlled NIC (Fig. 1), the dominant element in the timing network is inductance L, which allows step voltage changes, while the current is continuous function in time. Similarly, the oscillators with current controlled NICs (Fig. 2) have a capacitive timing network, which allows step current changes, while the voltage is continuous. The NIC parasitic elements which form part of the timing networks are, for convenience, drown outside the NIC two-port. Figures 1 and 2 are dual, and the analysis of the circuit is the same for both cases.

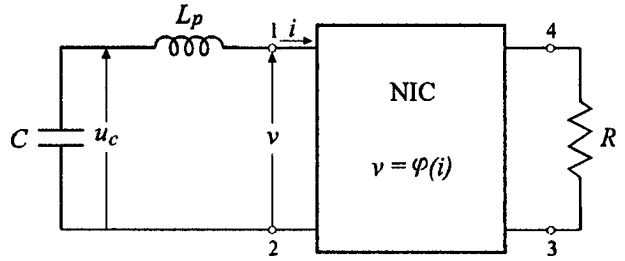


Figure 2. Oscillator with current controlled NIC.

Since the values of necessary capacitances are more practical than the values of inductances, the bulk of this work is concentrated on oscillators with capacitive timing network.

Conditions for the oscillations build-up

The current-voltage characteristic, $v=\varphi(i)$, for the current controlled NICs is shown on Figure 3. For small currents, the circuit images the resistance R from one port into a negative resistance - R_n, of approximately same absolute value, on the other port.

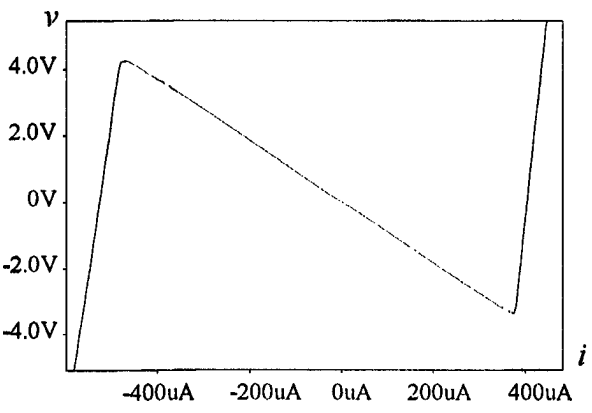


Figure 3. Current controlled NIC current-voltage characteristic.

The $v=\varphi(i)$ characteristic is almost ideally symmetrical for positive and negative control currents i. The characteristic can be easily adjusted to pass through the coordinate origin and there is no need to use additional biasing circuitry for placing the quiescent operating point, as is the case in some other oscillator implementations [1].

The RLC oscillator shown on Fig. 2 is implemented with a NIC having the above described current-voltage characteristic. A parasitic input inductance L_p, of arbitrarily small value, is taken out of the NIC and shown at the NIC input. The other parasitic elements, even if present, can be neglected, since they are lumped by the large capacitance C. A capacitance C is placed at the input of the circuit, and its current controls the operation of the oscillator. Using the notation from Fig. 2, the system of differential equations describing the time domain behavior of the given system is:

$$\frac{di}{dt} = -\frac{1}{L_p}\varphi(i) + \frac{1}{L_p}u_c \qquad (1)$$

$$\frac{du_c}{dt} = -\frac{1}{C}i. \qquad (2)$$

The characteristic equation of this system is:

$$\lambda^2 + \frac{1}{L_p}\varphi'(i)\lambda + \frac{1}{L_pC} = 0. \qquad (3)$$

The static characteristic of NIC can be approximated by linear segments of appropriate slope in the characteristic intervals of the control current values. The quiescent point of this system is at the coordinate origin:

$$\varphi'(i) = -R_n = \text{const.} \qquad (4)$$

The solutions of the characteristic equation of the system are:

$$\lambda_{1,2} = \frac{R_n}{2L_p} \pm j\sqrt{\frac{1}{L_pC} - \left(\frac{R_n}{2L_p}\right)^2}. \qquad (5)$$

Since the real part is always positive (R_n, $L_p>0$), the condition for oscillation build-up is always satisfied and does not depend on the values of the circuit parameters. The speed of the oscillation build-up depends on, and thus can be controlled by, the value of the imaged negative resistance R_n, and consequently by the value of the resistance R at the NIC input. By increasing R, the speed of the oscillations build-up is increased, but, as will be shown later, the other performances of the oscillator are degraded.

The imaginary part of the solutions (5) shows that, depending on the values of the circuit parameters, there can be two types of the oscillation build-up. If the condition:

$$R_n > \sqrt{\frac{4L_p}{C}} \qquad (6)$$

is fulfilled, the solutions (5) are real and different. In the phase plane (i,v), the quiescent point is moving on the strait line from the coordinate origin until the point where the current or voltage are limited by the circuit static characteristic. The limit cycle is thus almost immediately reached (Fig. 4), and the final amplitude of the output signal is achieved in the first period (Fig. 5).

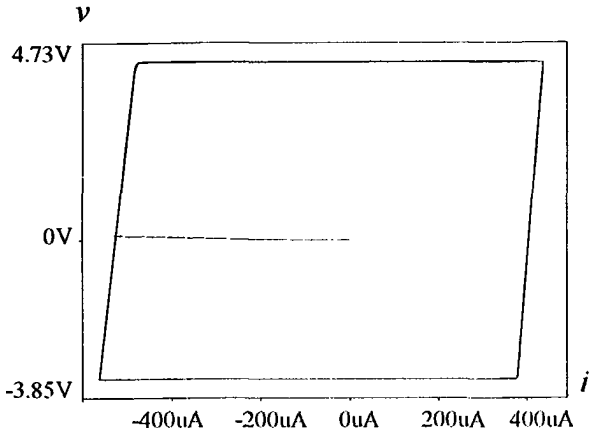


Figure 4. The phase portrait for the direct build-up of oscillations.

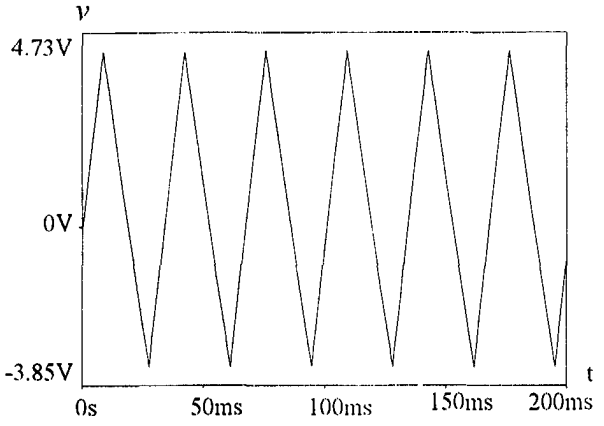


Figure 5. The direct build-up of oscillations.

The second type of oscillation build-up is present when the solutions (5) are complex. The quiescent point in that case leaves the coordinate origin on a spiral trajectory, and stays on it until the current and voltage values reach the values determined by the turning points of the characteristic $v=\varphi(i)$, when the limit cycle is reached.

In oscillators with voltage controlled NICs the oscillations are practically always of the second type (Fig. 6) because condition (6) becomes:

$$G_n > \sqrt{\frac{4C_p}{L}} \qquad (7)$$

which is hard to satisfy since the value of the parasitic capacitance C_p is small. In the time domain, the amplitude of the output voltage is exponentially increasing (Fig. 7).

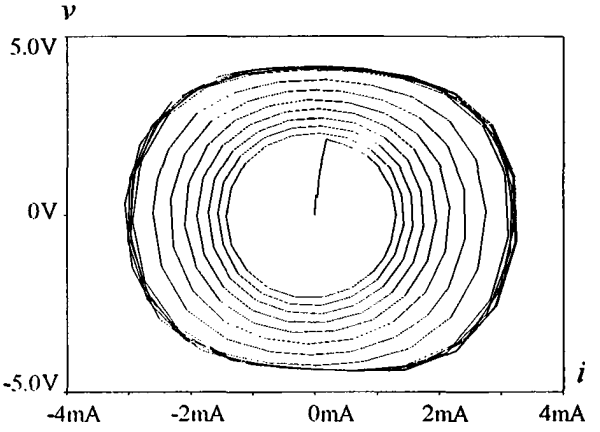


Figure 6. The phase portrait for the exponential build-up of oscillations.

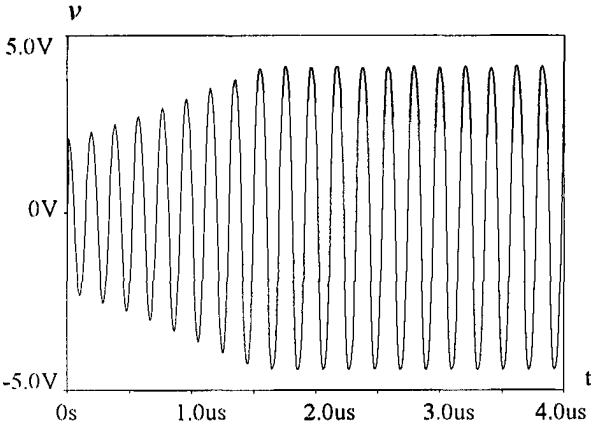


Figure 7. The exponential build-up of oscillations.

Circuit behavior analysis

The current controlled oscillator behavior can be explained by looking at the phase portrait in Fig. 4, and the corresponding static characteristic (Fig. 3). The negative slope of the static circuit characteristic in the neighborhood of the quiescent point allows the build-up of oscillations for any combination of the circuit parameter values. When the operating point reaches the

turning point of the static characteristic, there is a step change of the controlling current: from the value it had until then, to the value determined by the turning point. The operating point, after the jump, continues to move on the static characteristic with a large positive slope, until the other turning point is reached. Then there is a new step change of the operating point position, after which the operating point continues to move on the static characteristic. When the first turning point is reached for the second time, the system limit cycle is reached (Fig. 4).

The step change of the oscillator control current is, at first glance, impossible, since that current also passes through a small parasitic inductance L_p at the NIC input. However, from the relation between the current and the voltage of the inductance:

$$u_L = L_p \frac{di_L}{dt} \tag{8}$$

it can be concluded that the voltage drop on the inductance L_p is finite even if the current change is a step function, if the value of inductance L_p is arbitrarily small, i.e. if $L_p \to 0$. This condition is satisfied, since the inductance in question is parasitic. The step changes of the current can be seen on the time domain diagram on Fig. 8. For the voltage controlled circuits there is a step change of the control voltage on the parasitic capacitance C_p.

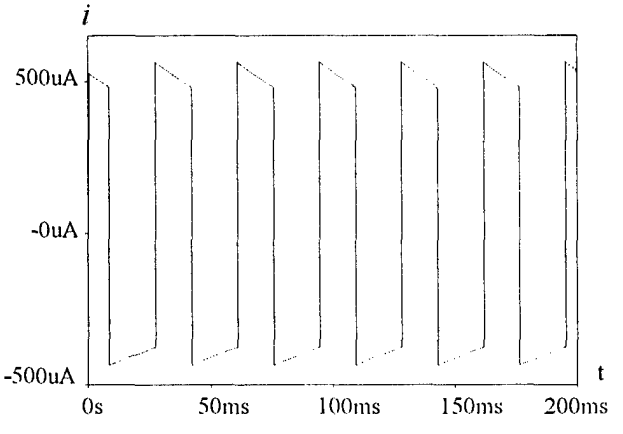


Figure 8. Step changes of capacitor current.

The determination of the period of oscillations

The static characteristic of the current controlled NIC can be approximated by three linear segments of appropriate slope. In order to make the analysis general, we assume that the characteristic is not ideally symmetrical (Fig. 9).

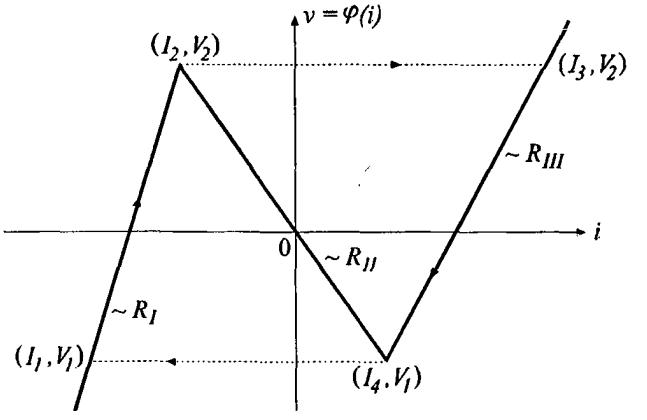


Figure 9. The current-voltage NIC characteristic; three segment approximation.

The period of oscillations is equal to the time needed for the quiescent point to go once through the limit cycle. The jumps of the quiescent point, described in the previous section, are almost instant and fall under the so called "fast" motions of the system (Fig. 9, doted lines). They can be practically neglected in the calculations of the period of oscillations. Thus, the predominant factor in determining the period of oscillations is the time that the quiescent point is in the "slow" motion, i.e. on the parts of the characteristic with a large positive slope (segments I and III):

$$T = \oint_c dt \approx \Delta t_I + \Delta t_{III}. \tag{9}$$

In the analysis of the slow changes, the parasitic elements can be simply assumed nonexistent ($L_p=0$) [2]. The system of equations (1)-(3) is then simplified:

$$u_c = v \tag{10}$$

$$i = -C\frac{dv}{dt} = -C\frac{d}{dt}(R_I i) \tag{11}$$

$$dt = -CR_I \frac{di}{i} \tag{12}$$

$$\Delta t_I = CR_I \ln\frac{I_1}{I_2} \tag{13}$$

$$T = CR_I \ln\frac{I_1}{I_2} + CR_{III} \ln\frac{I_3}{I_4}. \tag{14}$$

Equation (14) shows that the period depends on the shape of the NIC static characteristic (the slope of segments I and III, and the location of the turning points), and the capacitance C. It should be underlined that (14) is obtained under the condition that the system is in "slow" motion, thus when implementing oscillators operating at very high frequencies, we cannot expect the

period dependence on capacitance C to be as simple as (14) predicts.

For the oscillator whose time domain diagrams are shown on Figures 5 and 8, *($C=1\mu F$)*, (14) gives the period to be 33.73ms, while the simulated period is 33.56ms.

The current-voltage characteristic of the NIC with current conveyors can have one more narrow transitional segment with small positive resistance. In that case some additional parameters influencing the period of oscillations are included in (14), but the above analysis does not loose in its generality [2].

Oscillator Implementation

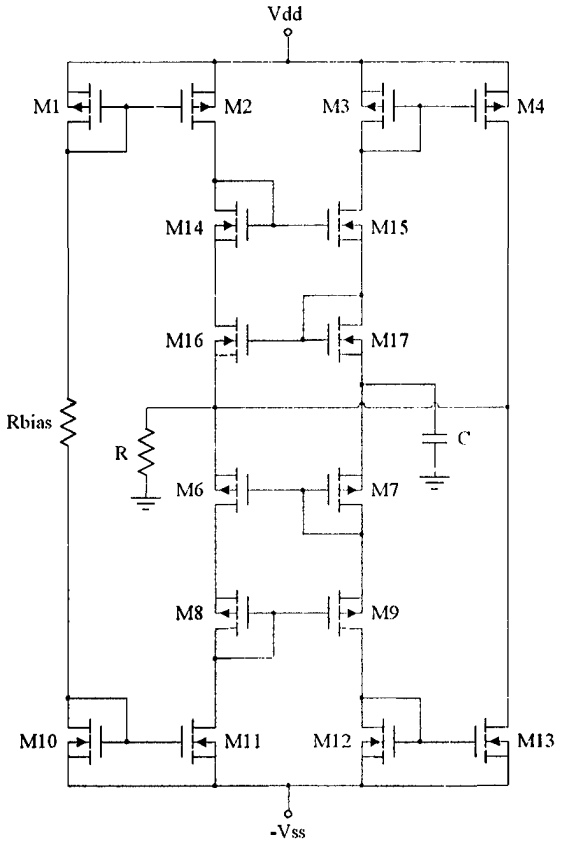


Figure 10. CMOS oscillator based on the current controlled Fabre-Normand NIC.

In the implementation of oscillators with current conveyors with discrete elements, it was shown that the characteristics of bipolar oscillators are much better than the characteristics of oscillators implemented with MOSFETs [1]. The reason of this poor performance of CMOS oscillators was that the transistors used had very wide channels, in order to achieve necessary transconductance. The input capacitance of such transistors is large and it reduces significantly the frequency range of the circuit. The maximum achieved frequency of bipolar oscillators was around 100MHz, while the frequencies of oscillation of CMOS oscillators were one order of magnitude smaller.

If the oscillator is implemented using integrated technology, the dimensions of the transistors can be chosen much more freely, in order to improve the performances of the oscillator.

For the implementation in the integrated technology, current controlled NICs are more attractive then voltage controlled NICs. The timing network in that case is composed of a capacitor, and thus the entire oscillator circuit can be integrated.

The Fabre-Normand current controlled NIC (Fig. 10) has the best characteristic for the implementation of oscillators [1]. It has a low power consumption and produces bipolar output voltage on isolated load. Negative resistance can be achieved in a wide range of values under current or voltage control. The cutoff frequency of the input impedance is high, up to a few hundreds of MHz. All oscillators described in this paper are based on this circuit.

Adjustment of the circuit characteristic

The speed of the build-up of oscillations, the amplitude and period of the output signal can be changed by changing the shape of the NIC static characteristic [2].

Simple analysis of the operation of the Fabre-Normand converter shows that the four "inside" transistors M6, M7, M16 and M17 are the transistors whose influence on the shape of the static characteristic of the circuit is predominant. The negative resistance part of the characteristic is at small currents, i.e. during the period in which all transistors are in saturation. At the moment when the input current becomes large enough to push these transistors out of the saturation region of operation, the static characteristic exhibits large positive resistance.

The other transistors in the circuit form current mirrors, and it is important that they should be matched and with large enough transconductance to be able to mirror the current.

By changing the geometry of the transistor pairs M6-M7 and M16-M17, the values of the currents at the turning points of the static characteristic can be changed, as well as the slopes of the positive resistance segments. The negative slope of the static characteristic depends directly on the resistance R which is mirrored by NIC, and the slope is always somewhat smaller then R due to non-ideality of current mirrors, Fig. 11.

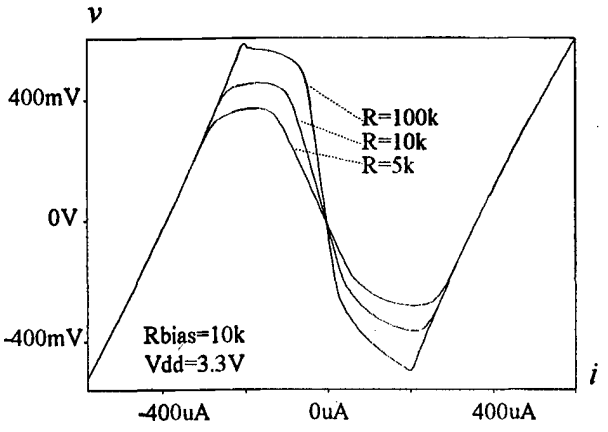


Figure 11. Static characteristics of the Fabre-Normand NIC.

The resistor R_{bias} determines the value of the bias current for the current mirrors. The change of this resistance does not significantly influence the static characteristic shape, but it changes the position of the quiescent point, which can thus be removed from the origin (Fig. 12).

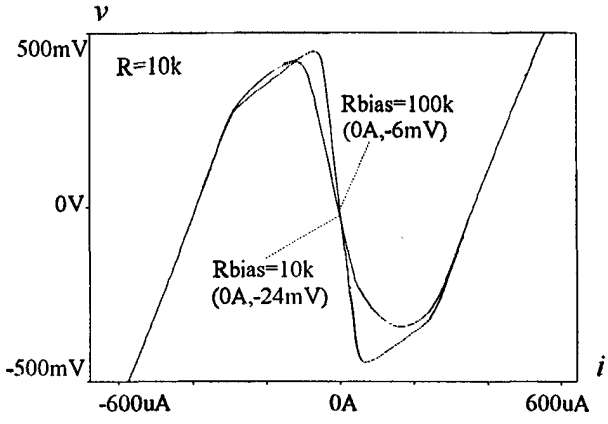


Figure 12. The influence of the resistor R_{bias} on the static characteristic.

The input time constant of the NIC is directly proportional to the resistance R. When R is increased, the NIC input impedance frequency range is narrowed, and the integrated implementation of the resistor is more difficult. For the build-up of oscillations, it is sufficient that a negative resistance section on the static characteristic exists, thus a small resistor can be chosen which just satisfies this condition. This will also decrease the speed of the build up of oscillations, but this is of secondary importance. The amplitude characteristics of the NIC input impedance are shown on Fig. 13. The cutoff frequencies for several values of the resistor R are noted on that figure.

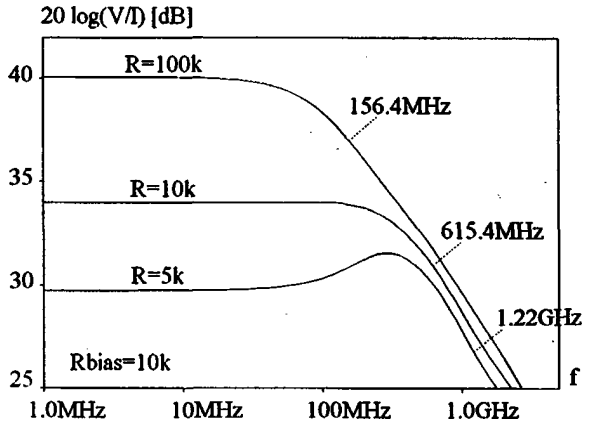


Figure 13. Amplitude characteristics of the input impedance of NIC.

A small power consumption of the circuit is also one of the targets when designing an oscillator in this class. The discrete realizations [1] and initial integrated realizations had a ±6V power supply. In order to decrease the power consumption, the voltage supply was decreased to ±3.3V, with further tendency of decreasing these values to stay compatible with the standard voltage supplies of today integrated circuits. This change led to the direct narrowing of the current range on the static characteristics with negative resistance, as well as to decreasing the value of voltage at the turning points. The current through R_{bias} was decreased proportionally, as well as the bias currents and transconductances of all transistors in the circuits, with the result that the negative slope of the static characteristic deviated even more from the value of the resistor R. The static characteristics of one oscillator realization with supply voltages of 3.3V and 6V are shown on Fig. 11 and Fig. 14, respectively.

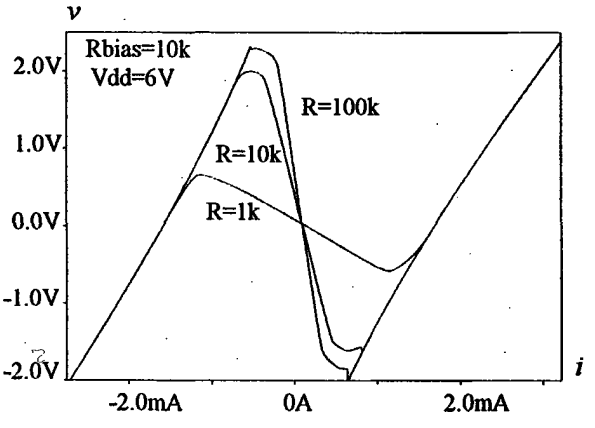


Figure 14. NIC static characteristics for V_{dd}=6V.

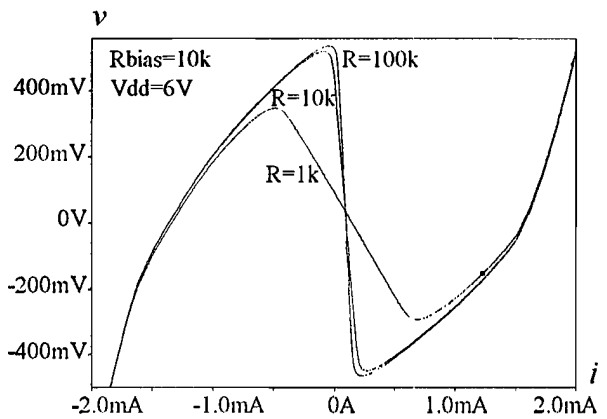


Figure 15. The influence of the diodes to NIC static characteristics from Fig. 14.

The power consumption of the oscillator can also be lowered by adding two antiparallel diodes in parallel with R, since the output voltage is thus limited to the threshold voltage of the diodes (Fig. 15). This modification does not work at smaller power supplies, where the output voltage is always smaller than the diode threshold voltage, and thus the diodes are never turned on. The parasitic capacitance of the diodes narrows the frequency range of the circuit, thus the use of diodes is not recommended in the case of RLC oscillators. However, because of the specific requirements when crystal oscillators are designed, the performances of the circuit can be significantly improved by the addition of the diodes.

RLC oscillators

An RLC oscillator with a Fabre-Normand current conveyor was designed, taking into account all the above mentioned effects of the values of circuit elements and transistor geometry on the oscillator characteristics. A 1.5μ-CMOS technology MOSIS compatible process was chosen for the implementation. The process in question is a p-well process, with two metal layers and two polysilicon layers [3]. Simulations of circuit behavior were performed using PSpice, using the second and fourth level transistor models [4].

The results are shown in Table 1. The maximal frequency of oscillations was achieved by narrowing the n-transistor channels to 10μm, and p-transistor channels to 30μm, while the "inside" transistors M6, M7, M16 and M17 were left with dimensions they had in the discrete implementation. The oscillator (osc. no. 1) with these parameters oscillated at 350MHz. The power consumption was around 18mW, and its output signal is shown on Fig. 16. Osc. no. 4, with similarly proportioned transistor geometries (i.e. different dimensions of inside transistors), has power consumption around 5mW, and the output signal amplitude of ±300mV.

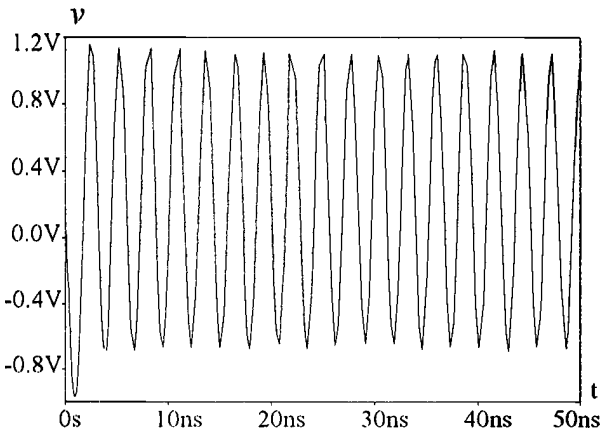


Figure 16. The time domain diagram of the RLC oscillator output voltage (f=350MHz).

The first four oscillator samples were designed with all transistors having the sources short-circuited to the substrate. Thus, these examples can be implemented only in dual-well technologies. In standard digital technologies, substrates of at least one type of transistors have to be at the same potential (at one of the power rails). In that case, the parasitic capacitance between source and substrate C_{bs}, as well as the dependence of the threshold voltage on the source-substrate voltage have to be taken in account. Due to this additional parasitic

osc. no.	V_{dd} [V]	R [kΩ]	R_{bias} [kΩ]	C [pF]	W_n * [μm]	W_p *[μm]	f [MHz]
1	6	5	20	0.15	10/27	30/61.5	350
2	6	5	20	0.19	27	61.5	295
3	3.3	10	20	0.5	27	61.5	125
4	3.3	10	20	0.2	27/135	61.5/307.5	165
5**	6	5	20	0.3	27	61.5	263

* *The length of all transistors is the same: $L_n=L_p=1.5\mu m$.*

** *All transistor substrates are tied to the appropriate power rail.*

Table 1. Examples of RLC oscillator implementations with Fabre-Normand converter in 1.5μ-CMOS technology.

capacitance, the oscillator operating frequency range is narrowed. This is shown on osc. no. 5, in which substrates of all transistors are short-circuited to the appropriate power rail. Its frequency of operation is 263MHz.

Crystal Oscillator Implementation

The implementation of crystal oscillators using current conveyors presents specific problems. Thus, the design of this type of oscillators is somewhat different then the design of RLC oscillators [1].

The quartz crystal can be modeled by a series resonant circuit oscillating at the resonant frequency, and a parallel capacitance due mainly to the parasitic capacitance of the crystal unit enclosure and leads. When this timing network is attached to the NIC, the parallel capacitance forms a parasitic RLC oscillator with the active part of the circuit. If the oscillating frequency of this parasitic oscillator is smaller than the frequency of the crystal unit, the parasitic oscillator starts oscillating first, the build up of oscillations at the series resonant frequency of the crystal unit does not happen, and the oscillator as a whole oscillates just at the parasitic frequency. This phenomenon can be eliminated by an adjustment of the static characteristics of the active part of the oscillator circuit, such that the result is in rising the parasitic frequency to the range above the crystal frequency.

The period of oscillations of the RLC oscillator based on current controlled NICs is directly proportional to the capacitance in the timing network and the positive slope of the NIC static characteristics. The parallel parasitic capacitance of the crystal unit depends on the kind of crystal unit enclosure, and is of the order of picofarads for crystals operating on frequencies up to 200MHz. If crystal parallel capacitance is assumed to be a given quantity, the only way to rise the corresponding parasitic RLC oscillator frequency of operation is by reducing the positive slope of the NIC static characteristic. The principal problem of the design of crystal oscillators using current conveyors is the reduction of the frequency range of the circuit thus that the limit frequency is between the crystal frequency and the parasitic RCL frequency. This problem can be solved by changing the geometry of transistors in the circuit or by adding two antiparalel diodes in parallel with the resistor R.

Parameters for three representative implementations of the crystal oscillators with current conveyors are presented in Table 2. The first two oscillators are one-pin oscillators with the crystal unit attached externally. The RCL parasitic frequency could not be raised above 40 and 70 MHz respectively for osc. 1 and osc. 2, although both described methods of rising this frequency were used. But, if the crystal unit is directly mounted to the integrated circuit, the parasitic capacitance is very small. The maximum operating frequency, if the parasitic capacitance is zero, is 200MHz (osc. no. 3).

The resistor R_{bias} in oscillator no. 2 was implemented using MOS transistor with the gate at the midpoint potential between the power rails. The time diagram of the output current for this oscillator is presented on Fig. 17. It should be noted that the Q factor of the crystal unit in all simulations was lowered in order to shorten the simulation time of the build-up of oscillations.

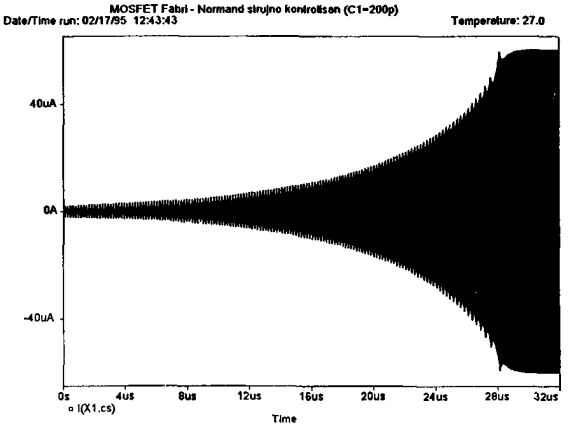


Figure 17. The time domain diagram of the output current for the crystal oscillator in the transient regime.

osc. no.	V_{dd} [V]	R [kΩ]	R_{bias} [kΩ]	C [pF]	W_n * [μm]	W_p *[μm]	f [MHz]
1	3.3	5	20	yes	27/135	61.5/307.5	30
2	3.3	2	trans.	yes	27/135	61.5/307.5	50
3**	6	5	20	no	10/20	30/60	200

* *The length of all transistors is the same:* $L_n=L_p=1.5\mu m$.

** *All transistor substrates are tied to the appropriate power rail.*

Table 2. Examples of crystal oscillator implementations with Fabre-Normand converter in 1.5μ-CMOS technology.

Conclusions

The CMOS current conveyer oscillators operate in a very wide range of frequencies, approaching the unit frequency of the active element in the circuit. The speed of the oscillation build-up, the amplitude and the frequency of oscillations, and the power dissipation of the oscillator can be easily adjusted by changing the shape of the static characteristic of the NIC. Of special interest are the oscillators with current controlled NIC which can be implemented in the integrated technology.

The presented analysis method is general and can be applied to other types of oscillators.

The RLC and crystal oscillators with current conveyors, due to their operation at high frequencies and low power consumption, can be used in a very wide range of electronic circuits. One of the possible uses of RLC oscillators is the replacement of monostable multivibrators in the systems with "watch-dog" timers, where reliable build-up of oscillations is required, and the frequency stability is of secondary importance. The crystal oscillators based on NICs can be used as integrated clock sources in many digital integrated circuits like processors.

References

[1] I. I. Ivanišević, D. M. Vasiljević, "The Quartz Crystal Oscillator Realization Using Current Conveyors", IEEE Trans. Circuits and Systems - I, vol. 40, pp. 530 - 533, August 1993

[2] A. A. Andronov, A. A. Vitt, S. E. Haikin, Theory of oscillations (on Russian), Moskva: Nauka 1981.

[3] ORBIT Semiconductors Inc., Data manual, 11/16/92.

[4] MicroSim Corporation, PSpice Circuit Analysis, Version 4.05, January 1991.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A SINGLE BRIDGE RESONATOR FOR LOW CONSUMPTION OCXO

Serge GALLIOU, Bernard DULMET and Marc MOUREY

Ecole Nationale Supérieure de Mécanique et des Microtechniques
Laboratoire de Chronométrie, Electronique et Piézoélectricité
26 Chemin de l'Epitaphe, 25030 BESANÇON CEDEX - FRANCE

Abstract

A Single Bridge Resonator (S.B.R.) concept has been previously proposed at Ref[8]. This quartz resonator is designed as an element of a 10 MHz ovenized oscillator, the power consumption of which is limited at 200 mW over the operating range [-20° C, +70° C] while its warm-up time is less than 3 minutes at +20° C. The results obtained from the first prototype revealed some weaknesses. This paper provides a brief review of its main features and gives also a presentation of a new device attempting to correct the defaults of the first one. Three essential changes have been decided in order to achieve a better temperature control and to decrease the thermal gradients and their effect upon frequency.
After the required coverage of these topics, we comment the experimental results obtained from the modified prototype. The emphasis is still essentially placed on the thermal behavior of the S.B.R. : warm-up time, power consumption, frequency stability over the operating temperature range.

1 - Introduction

The device described in what follows intends to combine the advantages of the Temperature Compensated Crystal Oscillator (TCXO) with those of the Oven Controlled Crystal Oscillator (OCXO). It must provide an output frequency immediately usable at turn-on with a low power consumption, like TCXO, and also an output frequency as stable as OCXO's especially versus the outside temperature. Such a device can be classified in the Tactical Miniature Crystal Oscillator family (TMXO) [1] to [9].
In fact the following discussion essentially deals with the quartz resonator which is the main non-conventional element of the tested devices. Indeed, all the conventional electronics excepted the quartz resonator is still submitted to the outside perturbations without any kind of protection or compensation. Aimed performances are a maximum power consumption of 200 mW with a warm-up time lower than 3 minutes at 25° C and operating temperatures within the [-20° C, +70° C] range.

2 - A review of the Single Bridge Resonator (SBR) concept [8] [10]

Basically, we intend to minimize the warm-up time under a limited power availability. This can be achieved through a drastic reduction of the heated volume. A thermal insulation of this volume is obviously necessary to ensure a low power consumption. In our case the minimum volume may be reduced to the quartz disk but the heating must operate without inducing any stress into the vibrating part of the resonator. That means first a resonator that can freely expand and second a proper inner temperature field. The former condition can be achieved with the monolithic structures shown in Fig.1 where the quartz disk is just tied to its environment by a single link (a bridge) made from the same matter. The latter condition would certainly be achieved if the temperature were homogeneous in the whole disk. Taking advantage that this resonator operates in a high-quality vacuum, the main thermal exchanges between the quartz lens and its enclosure are, at first sight, the conductive exchanges through its single bridge. As a consequence, imposing a constant temperature at the bridge will ensure isothermal conditions over the whole quartz lens and the bridge.
Finally, the single bridge is prolongated by a ring in such a way that a controlled heating can be performed over the bridge itself with a low power consumption according to the large thermal insulation provided by the ring-shaped part of this monolithic structure.

3 - Main changes operate on the prototype

The first S.B.R. prototype was based on the open ring solution presented in Fig. 1. Considering an open ring means that the thermal aspect was preferred to the mechanical one. The AT-cut was chosen rather than a less temperature-sensitive cut such as the SC-cut, precisely in order to emphasize the thermal behavior of the resonator and mainly because only one vibrating shear mode (the C mode, here the 5th overtone - 10 MHz) can be maintained without any chance of spurious B-mode driving. The electronics of the oscillator is then greatly simplified in comparison with an SC-cut

0-7803-2500-1/95 $4.00 © 1995 IEEE

oscillator that would certainly require an additional electric filtering to "kill" the corresponding overtone B mode : those filters increase the temperature sensitivity of the electronics.

Fig. 2 shows a photograph of the first analysed open ring S.B.R. prototype where it is visible that the free end of the ring is clamped into a metallic holder. Thus, this end is a "cold end" of which the temperature follows the outer temperature. At the other end of the ring, the "hot end", the quartz bridge is temperature-controlled by means of a surface-mounted resistor and a thermistor, the electronics of the control loop being located outside the quartz enclosure. For this device the thermistor was simply pasted on top of the resistor.

The last mounted S.B.R. is slightly modified by comparison with the former one. Some changes are shown in Fig. 3 :

- the metallic holder is replaced by a glass holder,
- the temperature probe is put into a little hole inside the bridge instead of being pasted on top of the heating resistor,
- the electrode surface is such increased that it almost covers the overall quartz surface.

In addition the azimuth of the bridge in the plane of the cut is optimized to suppress induced frequency shifts coming from the thermal gradients inside the bulk.
This changes have been made to improve first the disappointing results concerning the frequency change versus the ambient temperature : in the first design it reached 7.10^{-7} peak to peak within [-20° C, +70° C] ! Otherwise the warm-up and power consumption were expected ones.

4 - Discussion

The problem has been modelled with help of an electrical simulation (with the PSPICE software [11]) based on the electrical analogy of thermal problems and also with a finite element method (SYSTUS software [12]). The former method is easy to use and provides the opportunity to mix the thermal problem and the electrical control loop especially for the warm-up simulation. The latter modelling method is obviously more accurate (but also more cumbersome) regarding the temperature distribution and associates it to the induced stresses and strains map.
The analysis of the first prototype results revealed two main weaknesses :

- the first one was the ill-positioning of the temperature probe which measured a temperature too far from that of the quartz bridge. The probe temperature was correctly locked by the control loop but was not sufficient to lock the bridge temperature and also to lock the quartz temperature that remained too much dependent on the ambient temperature [8] ;
- the second one came from a parasitic flux still existing between the quartz lens and the enclosure: conduction was quite perfectly controlled but not radiation.

The thermal model of Fig. 4, already described in ref. [8], is convenient to understand the thermal behavior of the device and to calculate the orders of magnitude. The influence of the positioning of the probe, already discussed and simulated [8], is not detailed here. Nevertheless, it must be noticed that the practical choice which consists in making a hole (Ø 0.2 mm) in the thickness of the quartz bridge to measure the temperature of the heart of the bridge is very elaborated.

This simple model gives a good idea of the influence of the parasitic radiative fluxes. The shape factor [13] between the quartz lens and its enclosure is approximated by :

$$f_{QE} = \frac{\varepsilon_Q \, \varepsilon_E}{\varepsilon_E + \varepsilon_Q \frac{S_Q}{S_E}(1-\varepsilon_E)} \qquad (1)$$

where ε_Q and ε_E denote the quartz emissivity and the enclosure emissivity respectively, and S_Q, S_E their respective surfaces. The whole radiative flux is then divided into four fluxes $P_i^R = S_{Q_i} \, f_{QE} \, \sigma \left(T_{Q_i}^4 - T_E^4 \right)$, i = 1, ... 4, assuming that all the average temperatures T_{Q_i} of the four slices that model the quartz bulk are close together from the other. S_{Q_i} denotes the only corresponding non-metallized surface of a quartz slice : indeed, the deposited gold of the electrodes acts as a perfect reflector (i.e, its own emissitivity is zero). Moreover, the quartz material is considered here as a black body $(\varepsilon_Q = 1)$ because 98% of the emissive energy remains within the range $\left[\frac{\lambda_m}{2}, 8\,\lambda_m \right]$ with $\lambda_m \approx 8.2\,\mu m$ at about 80°C (353 K) according to the Wien's rule (below $\lambda = 4\,\mu m$ quartz is transparent !).

Figure 5 shows how significant is the effect of the electrode surfaces on the temperature distribution inside the quartz. The set of the more scattered temperatures simulates the first S.B.R. of which the electrode surfaces were the smaller. It leads to an average temperature deviation of about 4°C from -20°C to +70°C, that is to say ± 2°C around a turn-over point adjusted at the middle of the range, i.e 25°C. Considering the static frequency-temperature curve of the AT-cut this will give a relative frequency change of 8.10^{-8}. Moreover, an additional frequency shift have to be superimposed to this last value because of the temperature gradients (2.5° C at -20° C and 0.6° C at 70° C).

The other set of temperature values (the more gathered ones) is that of the new S.B.R. which is almost completely gold coated. The average temperature change falls down to about 1° C (that is ± 0.5° C around a turnover point or $\frac{\Delta f}{f} \approx 8.10^{-9}$) and gradients are then limited to 1.2° C à -20° C and 0.2° C at +70° C.
Another consequence of the electrode increasing is the power consumption gain. This is illustrated in figure 6 where the significance of radiative exchanges against conduction is pointed out. It can be mentionned that the simulation predicts a gain on the radiative fluxes of 2.2 for an increase of the electrodes surfaces by a 1.6 factor.

Performing a correct analysis of the frequency shift created by a non-homogeneous thermal bias requires some theoretical understanding. As a matter of fact, no FEM package is able yet to determine the global frequency shift at once, on the fly. Thus, different steps have to be distinguished. As a first principle, we consider that the class of problems of interest can be described by linearized equations for infinitesimally small fields superimposed on a bias [14].

$$\left.\begin{aligned} \tilde{K}_{L\gamma,L} &= \rho_0 \ddot{u}_\gamma \\ \tilde{K}_{L\gamma} &= G_{L\gamma Mv}.u_{v,M} \end{aligned}\right\} \quad (2)$$

where $\tilde{K}_{L\gamma}$ denotes the dynamic increment of the asymmetric Piola-Kirchhoff stress tensor, and G coefficients are called effective elastic coefficients because they depend on the bias. In this paper, we consider that the local variation of effective coefficients from an initial stress-free state chosen as a reference comes from adding up two effects:

$$\left.\begin{aligned} \Delta\left(G_{L\gamma M\varepsilon}\right) &= \Delta^1_{L\gamma M\varepsilon} + \Delta^2_{L\gamma M\varepsilon} \\ \Delta^1_{L\gamma M\varepsilon} &= \Delta\left({}_2C_{L\gamma M\varepsilon}\right) - \left[{}_3C^0_{L\gamma M\varepsilon AB}\alpha_{AB} \right. \\ &\quad \left. + \left({}_2C^0_{L\gamma MS}\delta_{\varepsilon R} + {}_2C^0_{LSM\varepsilon}\delta_{\gamma R}\right)\alpha_{RS}\right]\Delta\Theta \\ \Delta^2_{L\gamma M\varepsilon} &= {}_3C^0_{L\gamma M\varepsilon AB}\eta^1_{AB} + {}_2C^0_{L\gamma MS}v_{\varepsilon S} + {}_2C^0_{LSM\varepsilon}v_{\gamma S} \end{aligned}\right\} \quad (3)$$

The first term Δ^1 indicates the temperature induced (local) variation of the considered effective coefficients at zero stress. It depends on the thermal variation of so-called fundamental elastic constants $\Delta_2 C$ [15], on third order elastic constants ${}_3C$, on thermal expansion coefficients α, and on the temperature variation itself $\Delta\theta$, indeed. Thus, Δ^1 can be computed at every node of a proper purely thermal analysis made with help of finite element method.
All expansions in (2) have been made up to the first order in terms of temperature. Nevertheless, it has been shown at Ref [16] that the computation of Δ^1 can be extended up to the third order in terms of temperature :

$$\Delta^1_{L\gamma M\varepsilon} = \sum_{n=1}^{3} T^{(n)}G_{L\gamma M\varepsilon}\,\Delta\Theta^n \quad (4)$$

where $T^{(n)}G$ are the set of thermal coefficients defined at Ref [17].
The second term Δ^2 defines the part of the effective coefficient variation that comes from thermally-induced stresses. It does not depend on thermal expansion constants but involves the following quantities:

$$\left.\begin{aligned} \eta^1_{RS} &= E^1_{RS} - \alpha_{RS}\,\Delta\Theta \\ v^1_{\varepsilon S} &= w_{\varepsilon,S} - \delta_{\varepsilon R}\,\alpha_{RS}\,\Delta\Theta \end{aligned}\right\} \quad (5)$$

η^1 is the part of the static deformation E^1 which is associated to non-zero stresses. v^1 is the corresponding part of the gradient of total static displacement w.

FEM softwares usually provide with a means to compute quantities such as η^1 and v^1 in a so-called thermoelastic analysis performed after the purely thermal one. Of course, in the here-considered case, the existence of thermal gradients leads to a non-homogeneous variation of the effective constants governing the thickness-shear propagation. Then, a perturbation integral would be necessary to compute the global frequency shift of the resonant mode of interest for a given temperature map. Nevertheless, taking into account a relatively strong vibration trapping in the center region of the disk, we got reasonably accurate estimations averaging Δ^1 and Δ^2 cumulated effects over a few regularly spaced nodes in this restricted region. It should be observed that the whole procedure is greatly facilitated using the same mesh for the purely thermal analysis and the subsequent thermoelastic analysis.

We already mentionned in our previous paper [8] that solving the purely thermal problem in the framework of a simple FEM model that linearizes Stefan's law around the probe temperature leads to a somewhat overstimated frequency shift in compare with a fully non-linear analysis with respect to the modelling of radiative exchanges. We present at Fig. 7 newly obtained results showing the total frequency shift versus the bridge azimuth in the AT-cut plane. We considered the worst case where the temperature of the clamped origin of the ring and of the surrounding enclosure was imposed at -20° C, whereas the probe temperature was imposed at +80° C. Because of the fair amount of hard disk storage required for both purely thermal and thermoelastic analyses for a given value of azimuth, we performed the computation with an azimuth step of 30 degrees. The

continuous curve has been obtained by a least square fit of Fourier symmetric coefficients accounting for the axial symmetry around X axis. Thus, it appears that a positioning of the bridge in ± 45° direction in the AT cut plane ensures a minimal sensitivity.

The exchange of the metallic holder for a glass holder has been notified above. It is obvious that the second one provides a better thermal insulation, butthis is not significant here, in comparison with the large thermal insulation given by the open ring. Nevertheless, this glass holder will be the solution for the insulation in the closed ring case.

5 - Experimental results

Practical results are roughly those expected by the simulation. Fig.8 shows the power consumption of the heater (without taking into account the proper consumption of the oscillator and its output amplifier) versus the ambient temperature. The continuous line which fits the points of measurement shows that non-linearities of thermal exchanges (induced by radiation) are significant.

Fig.9 is an example of warm-up behavior at room temperature. The control loop reaches here a linear working area less than 30 secondes after turn-on, with a maximum power of 230 mW in saturation (the main supply voltage has been chosen at 24 Volts and the heating resistor value at 2.2 kΩ for being able to load a conventional operational amplifier). In the same time the frequency drops towards its final value with a few overshoots as shown in Fig.10 a and 10 b : these dynamic effects are difficult to avoid with an AT cut. Finally, the frequency is within $\pm 5.10^{-9}$ in 200 seconds (a little more than the aimed 180 seconds). When the ambient temperature comes from -20° C to +70° C, the maximum frequency deviation is 3.10^{-8}, which is a good performance (see Fig. 11). This value can be considered as a limit when all the electronics is submitted to the temperature change.

The elements of the equivalent electrical circuit of the new S.B.R. have also been measured and compared to those of the first S.B.R. (see Table 1). It appears that the increase of the electrode surface has no appreciable effect on the motional parameter values (R_1, C_1, L_1).

	C_0 (pF)	R_1 (Ω)	C_1 $(10^{-3}\ pF)$	L_1 (H)	Q (10^6)
First S.B.R.	5.3	113	0.17	1.5	0.84
New S.B.R.	7.4	114.8	0.16	1.6	0.88

Table 1

6 - Near future perspectives

Although significant improvements (roughly by a 10 factor) have been attained with help of the increase of electrodes surfaces, the 45° azimuth orientation of the bridge, and the new probe positioning, we think that ultimate performances of this class of devices have still not been reached. For instance, the closed ring design would easily allow better mechanical performances whereas an even better thermal response would be obtained with a proper integration of the oscillator and thermal control electronics on a single chip directly positioned on the bridge itself.

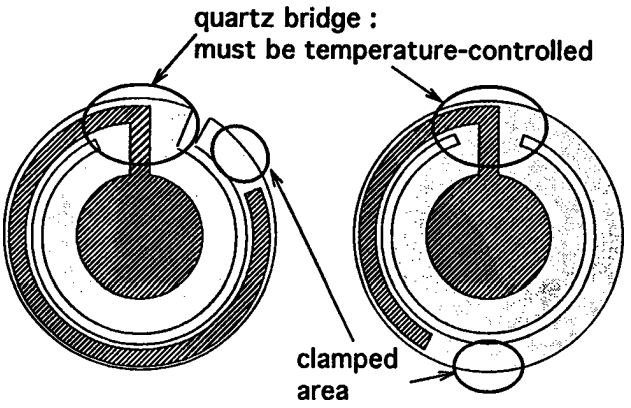


Fig. 1 : The open ring SBR - The closed ring SBR

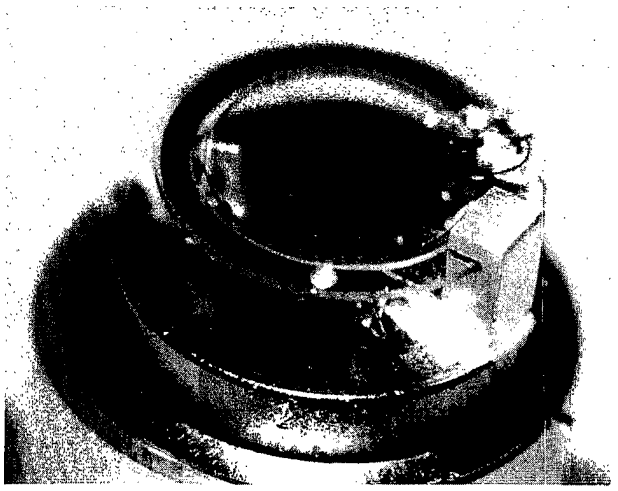


Fig. 2 : The first SBR prototype



Fig. 3 : The new SBR

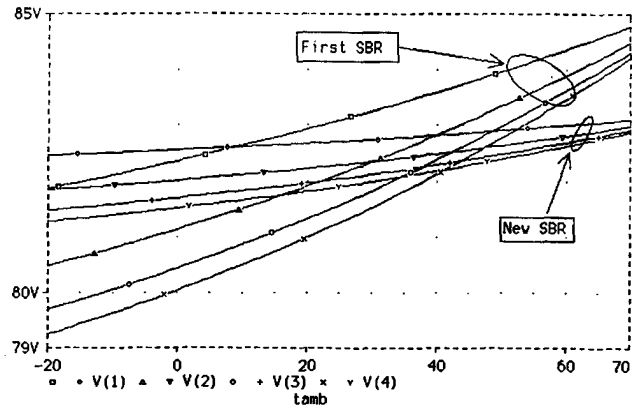


Fig. 5 : Temperatures inside the quartz resonator are less scattered in the new SBR than in the previous prototype.

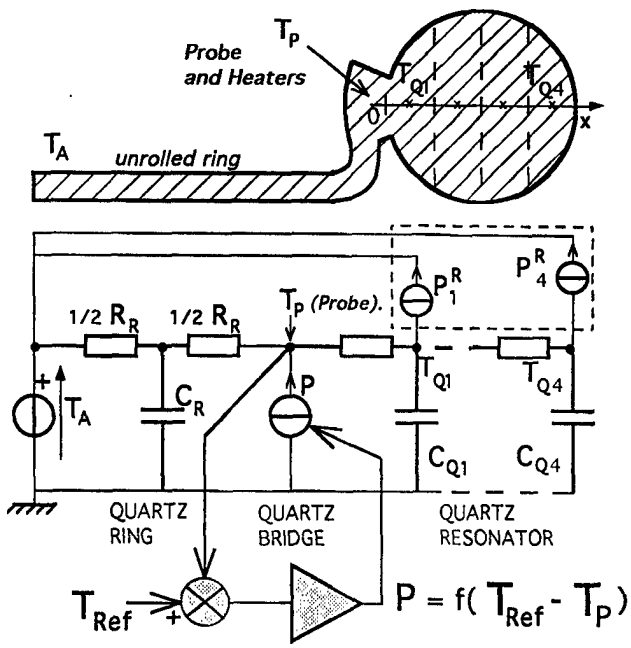


Fig. 4 : The equivalent circuit.
There the contact between the probe and the bridge is ideal. In the first prototype it was necessary to insert a resistor between them (the equivalent thermal resistance of the surface-mounted resistor).

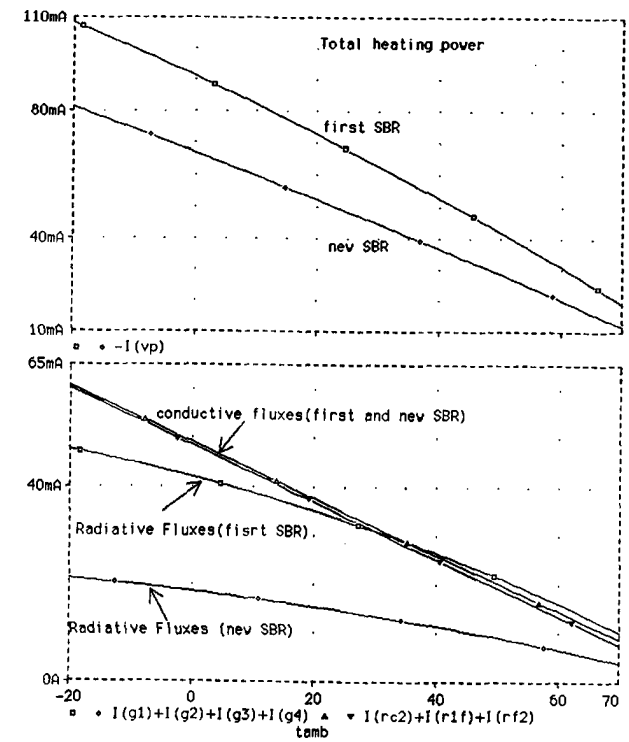


Fig. 6 : Main fluxes and heating power consumption. Lower values are those of the new prototype compared with those of the first prototype.

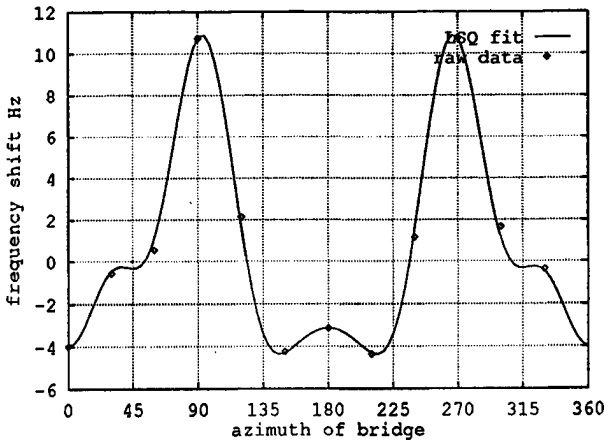


Fig. 7 : Thermally-induced frequency shift versus azimuth of bridge (external temperature : -20° C).

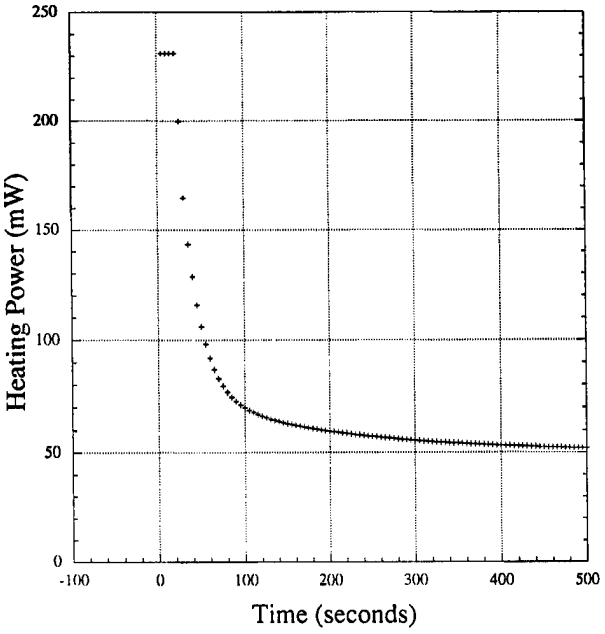


Fig. 9 : Heater power consumption during the transient period of the warm-up (the ambient temperature is the room temperature).

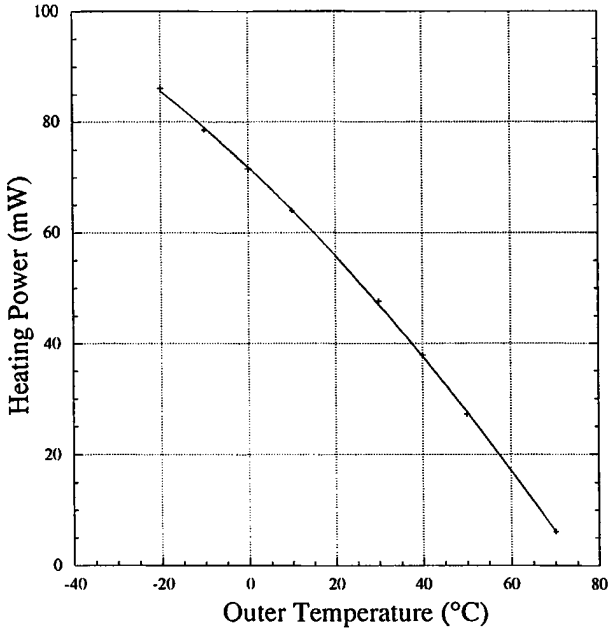


Fig. 8 : Heater power consumption versus the ambient temperature.
The heater is a 2.2 kΩ surface-mounted resistor which loads an operational amplifier supplied with 24 Volts.

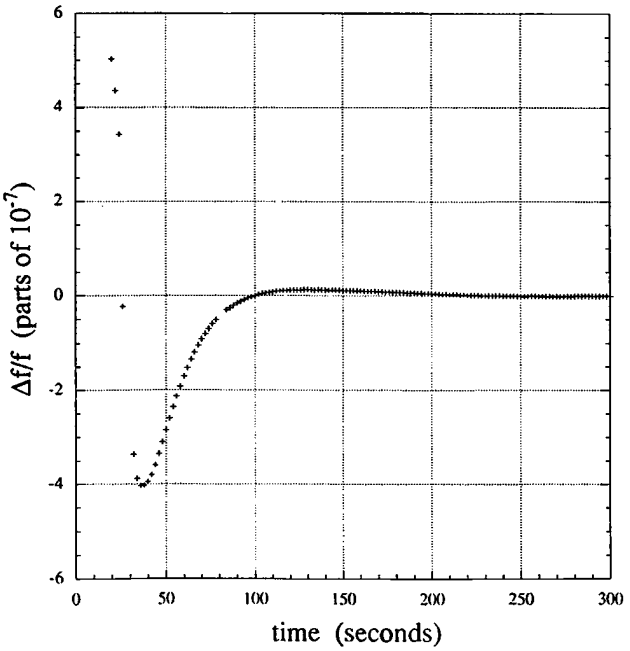


Fig. 10a : The output frequency behavior corresponding to the warm-up of Fig. 9.

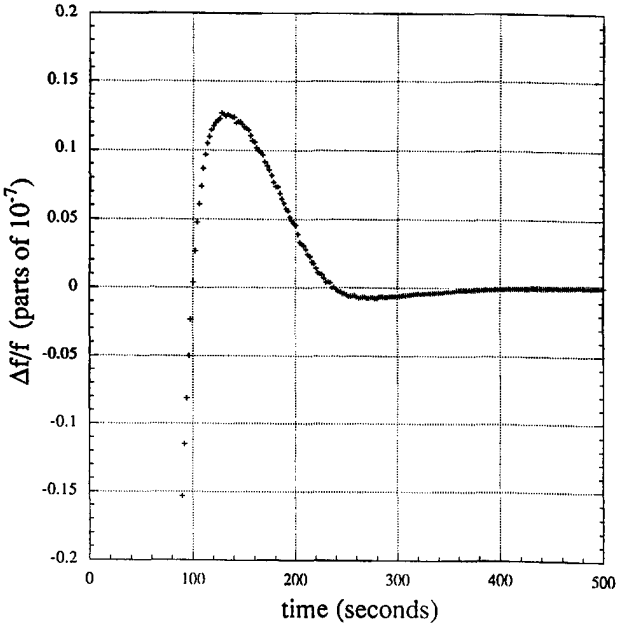


Fig. 10b : A zoom of Fig. 10a.

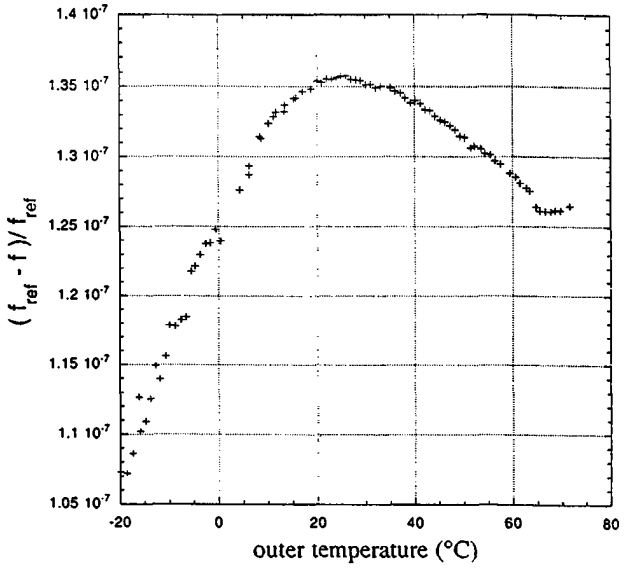


Fig. 11 : The output frequency change versus the ambient temperature.

REFERENCES

[1] H.W. JACKSON, "Update on the tactical miniature crystal oscillator program", Proc. of the 36th Annual Frequency Control Symposium, pp 492-498, 1982.

[2] S. GALLIOU, "Chauffage rapide avec résonateurs BVA tactiques", Contrat DRET n° 85/180, pp 31-43; 1987.

[3] D.A. SYMONDS and D. ALLEN, "An update on the TMXO", Proc. of the 44th Annual Symposium on Frequency Control, pp 615-621, 1990.

[4] B. LONG and G. WEAVER, "Quartz crystal oscillators with direct resonator heating", Proc. of the 45th Annual Symposium on Frequency Control, pp 384-392, 1991.

[5] Y.L. VOROKHOVSKY and B.G. DRAKHLIS, "High-stability quartz oscillators on internally-heated quartz resonators with AT and SC cuts", Proc. of the 45th Annual Symposium on Frequency Control, pp 447-450, 1991.

[6] J.P. AUBRY, G. WAGNER and V. CANDELIER, "Miniaturization and stability : a new generation of ovenised oscillators". Proc. of the 6th European Frequency and Time Forum, pp 173-177, 1992.

[7] I.V. ABRAMSON and A.N. DIKIDSHI, "Improvement of characteristics of quartz resonator thermostate with direct heating piezoelement", Proc. of the IEEE Frequency Control Symposium, pp 499-504, 1992.

[8] S. GALLIOU, M. MOUREY, B. DULMET, "A single bridge resonator", Proc. of the 8th European Frequency and Time Forum, Vol. II, pp 792-801, 1994.

[9] I.V. ABRAMSON, "Recent advances in composite heated quartz resonators intended for high stability mobile OCXOs", Omsk Piezoengineering Co. Ltd. - Omsk, Russia.

[10] French Patent n° 92 03175, BESSON R., GALLIOU S., MOUREY M., "Résonateur pour oscillateur thermostaté à faible consommation et chauffage rapide". Patent pending Europe : n° 93 400638.8 ; USA : n° 08/033,092 ; Japan : n° 5-57467.

[11] PSPICE circuit analysis - MicroSim Corporation. 20 Fairbanks, Irvine, California 92718 - USA.

[12] SYSTUS - Framasoft / CSI. Tour Fiat Cedex 16, 92084 PARIS LA DEFENSE - FRANCE.

[13] M.N. ÖZISIK, "Radiative transfer and interactions with conduction and convection", John Wiley and Sons, New-York, pp 158-166, 1973.

[14] J. C BAUMHAUER, H.F. TIERSTEN, "Non linear electroelastic equations for small fields superimposed on a bias", J.Acoustical Soc. Am. 54, 1017, 1973.

[15] B. K SINHA, H.F. TIERSTEN, "First temperature derivatives of the fundamental elastic constants of quartz", D. Appl. Phys. 50 (4), April 1979, pp 2732-2739.

[16] M.R. HAMROUR, B. DULMET, S. GALLIOU, "Finite elements modelling of infrared-sensitive resonators used as thermal sensors", EFTF 1995, Besançon-France.

[17] B. DULMET, R. BOURQUIN, "Influence of the definition of material constants in the study of crystal plates thermal behavior", Proc. of 3rd EFTF, pp 305-312 (1989).

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

FAST START-UP CRYSTAL OSCILLATOR CIRCUITS

Yasuo Tsuzuki, Takehiko Adachi and Ji Wen Zhang
Yokohama National University, Yokohama 240, Japan

Abstract

This paper presents a reduction method of the start-up time after the switch-on of the D.C. power supply of a crystal oscillator circuit. It is shown that a fast start-up crystal oscillator can easily be obtained by changing only the circuit connection of the crystal resonator. The new scheme can be applied to the usual Colpitts crystal oscillator circuits, to the cascode crystal oscillator circuits, and also to the overtone crystal oscillator circuits.

Introduction

The crystal oscillator plays an important part in portable telephone-sets. In such a telephone-set, switch-on and switch-off of the power supply are repeated frequently for the power saving of the batteries. As the quality factor Q of the crystal resonator is very high, the switch-on behavior of the crystal oscillator is the slowest in the electronic circuits of the telephone-set. Therefore, the reduction method of the start-up time is desired from the telephone-circuits designers.

We have previously reported on an analysis method of the start-up characteristics of the crystal oscillators [1]. In the following section, the start-up behavior obtained by this analysis method is outlined. Next, the principle of the reduction method of the start-up time is discussed, and some application examples are shown. Finally, the effect of the new scheme is verified by the computer analysis program SPICE [2].

Start-Up Behavior of Crystal Oscillator Circuit

The circuit configuration of the usual Colpitts crystal oscillator is shown in Figure 1.

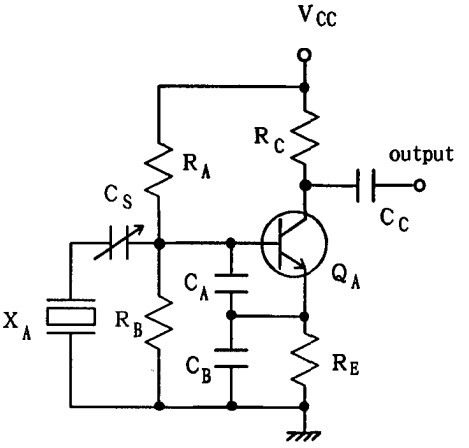


Figure 1. Colpitts crystal oscillator circuit

According to the results of the computer analysis on this circuit, the start-up behavior from the switch-on to the steady oscillation state can be divided into the three time intervals from the operating condition of the transistor. The three intervals are as follows. (1) The first interval is from the switch-on time to the time when the transistor bias reaches its operating point, and this interval is very short. (2) The following second interval is

0-7803-2500-1/95 $4.00 © 1995 IEEE

the period while the oscillation amplitude is so small that the characteristics of the transistor can be treated by using the small-signal equivalent circuit, and this interval is the longest. (3) The third interval is the period when the oscillation amplitude becomes so large that the characteristics of the transistor must be treated by using a large-signal model such as the Gummel-Poon model.

The oscillation amplitude of a crystal oscillator can be expressed by the high-frequency current amplitude of the crystal resonator which is proportional to the vibration amplitude of the crystal resonator. Corresponding to the three intervals mentioned above, the high-frequency current amplitude of the crystal resonator varies in the following manner. In the first interval, the resonator current keeps constant value. In the second interval, the resonator current increases exponentially with time. In the third interval, the resonator current approaches gradually to the level of the steady oscillation state. As an example of this characteristic, that of a 10 MHz oscillator is shown in Figure 2.

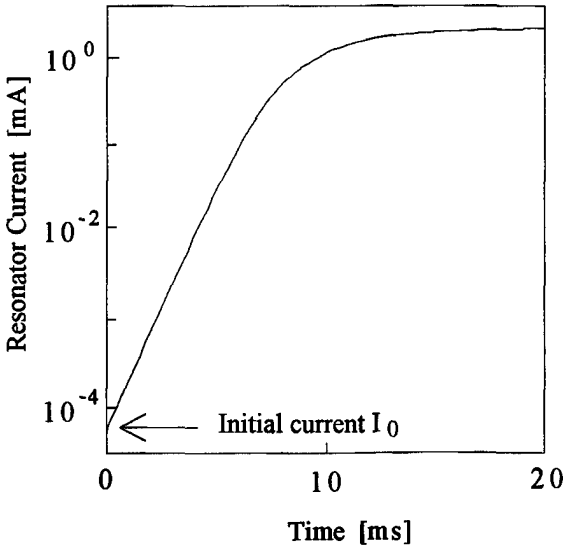


Figure 2. Start-up characteristic of resonator current in a 10 MHz crystal oscillator

The initial resonator current I_0 in the first interval can be calculated approximately from the following equation.

$$I_0 = \sqrt{C_1 / L_1} \times V_t \quad (1)$$

where, C_1 and L_1 are the equivalent motional capacitance and the equivalent motional inductance of the crystal resonator respectively, and V_t is the amplitude of the step-like voltage that appears across the resonator by the switch-on of the power supply. In the case of the usual Colpitts oscillator circuit, V_t is provided through the base-collector capacitance C_c of the transistor, therefore, the value of V_t is very small. If we can make the value of V_t higher, the initial resonator current I_0 becomes larger and the start-up time becomes shorter.

Principle of Reduction of Start-Up Time

In the case of the usual Colpitts oscillator circuit, one side of the resonator pins is grounded as shown in Figure 1. If we change the grounded pin connection to the D.C. power line as shown in Figure 3, the amplitude of V_t becomes almost equal to the power line voltage V_{cc}.

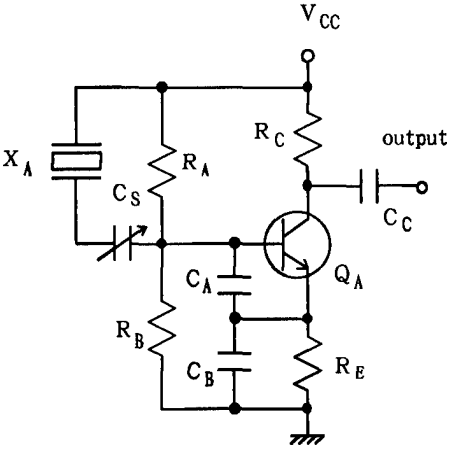


Figure 3. Fast start-up Colpitts crystal oscillator circuit

In the second interval, where the resonator current I increases exponentially with time, the relation between

the time t after the switch-on and the resonator current I can be expressed as the following equation.

$$I = I_0 \exp\left(\frac{R_n - R_1}{2L_1} t\right) \quad (2)$$

where, R_n is the negative resistance of the transistor circuit in the small-signal operation, and R_1 is the equivalent motional resistance of the crystal resonator.

Let I_1 is the value of the initial resonator current of the circuit in Figure 3. Then, the reduction time t_R can be obtained from the equation (2) by letting $I = I_1$ and $t = t_R$. The result is as follows.

$$t_R = \frac{2L_1}{R_n - R_1} \ln\left(\frac{I_1}{I_0}\right) \quad (3)$$

In the case of the circuit shown in Figure 3, the similar reduction effect can be obtained by changing the grounded connection of the capacitor C_B instead of that of the resonator.

Applications of New Scheme

In recent portable telephone-sets, the cascode type crystal oscillator circuit, as shown in Figure 4, is widely used. This circuit has better frequency stability against the variation of the load impedance. However, as the effect of the switch-on of the D.C. power supply is transferred through the two transistors to the crystal resonator, the initial resonator current is extremely low, and the start-up time is much longer compared with the usual Colpitts circuit. In the case of this circuit, the initial resonator current can be made extremely larger by changing the grounded pin connection of the resonator or the capacitor C_B to the power line.

Another example is the application of the new scheme to the overtone crystal oscillator circuits. In the usual

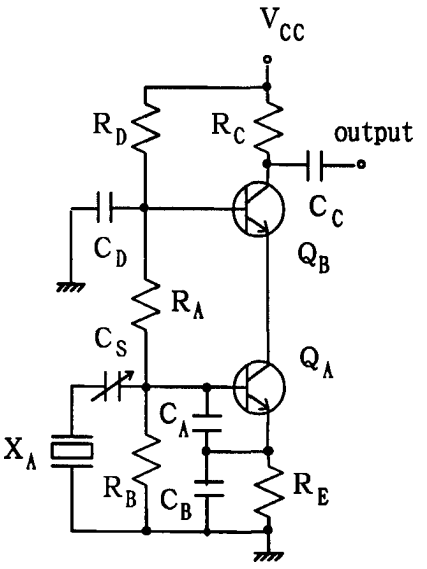


Figure 4. Cascode crystal oscillator circuit

Colpitts oscillator circuit of Figure 1 and the cascode oscillator circuit of Figure 4, the crystal resonators are used to vibrate in their fundamental-mode. In the very-high-frequency range, the overtone-mode oscillator circuits are in wide use. An example of the overtone-mode oscillator circuits is shown in Figure 5. In the case of this circuit also, the initial resonator current can be made larger by changing the grounded pin connection of the resonator to the power line.

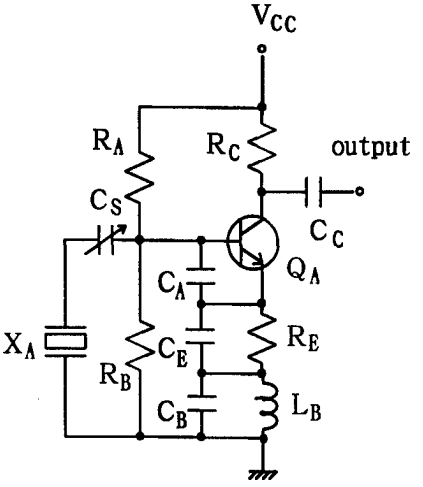


Figure 5. Overtone crystal oscillator circuit

Verification by Computer Analysis

The effect of the new scheme has been examined on a 10 MHz cascode crystal oscillator circuit by means of the computer analysis program SPICE. The circuit parameters of Figure 4 are listed in Table 1, and the equivalent circuit parameters of the crystal resonator used are listed in Table 2.

Table 1 Circuit parameters in Figure 4

R_A	R_B	R_C	R_D	R_E	C_A	C_B	C_D
68	62	1.8	47	5.1	180	220	10^4

R: kΩ, C: pF, V_{CC}=5[V], Q_A & Q_B: 2SC1359

Table 2 Equivalent circuit parameters of crystal resonator

f_S [MHz]	L_1 [mH]	C_1 [fF]	R_1 [Ω]	C_0 [pF]
10.017	18.7	13.5	11.9	2.75

We analyzed the start-up characteristics of the two circuit configurations, the usual connection and the fast start-up connection of the resonator. We evaluated the start-up time as the time interval from the switch-on time of the D.C. power supply to the time when the resonator current amplitude reaches to the 90% value of the steady oscillation state. The initial resonator current and the start-up times of the two circuits are shown in Table 3.

Table 3 Initial resonator currents and start-up times of two circuits

circuit	initial current	start-up time
usual	0.16 [nA]	33.6 [ms]
fast start-up	3.89 [μA]	16.0 [ms]
ratio	2.4×10^4	48 %

It is shown in Table 3 that the initial resonator current becomes extremely larger by changing the resonator connection, and the start-up time can be reduced by approximately 1/2. In Table 3, the difference of the start-up-time between the two circuits is 17.6 ms. The reduction time t_R calculated by equation (3) is 17.3 ms, and good agreement has been obtained.

It has been also proved by the computer analysis that the new scheme has the similar effect on an overtone cascode oscillator circuit.

Conclusion

It is concluded that a fast start-up crystal oscillator circuit can easily be obtained by changing only the circuit connection of the crystal resonator. The new scheme can be applied to the usual Colpitts crystal oscillator circuits, to the cascode crystal oscillator circuits, and also to the overtone crystal oscillator circuits.

References

[1] J. Lu and Y. Tsuzuki, "Analysis of Start-up Characteristics of Crystal Oscillators", Proceedings of the 45th Annual Symposium on Frequency Control, pp. 360-363, May 1991.

[2] T. Adachi, M. Hirose and Y. Tsuzuki, "Computer Analysis of Colpitts Crystal Oscillator", Proceedings of the 39th Annual Symposium on Frequency Control, pp. 178-182, May 1985.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

UNIVERSAL, COMPUTER FACILITATED, STEADY STATE OSCILLATOR, CLOSED LOOP ANALYSIS THEORY

ANALYSIS AND DESIGN OF CRYSTAL OSCILLATOR ALC SYSTEMS

Benjamin Parzen
consulting engineer
3634 Seventh Avenue
San Diego, CA 92103

Abstract

This paper constitutes a partial revision of Chapter 14 in Ref 3, treating ALC oscillators, to include the applicable titled theory, developed 10 years later than Chpt 14, and material on the design of ALC systems. It takes advantage of the additional insight, provided by the theory, and the interim advances in the state of the art.

1. Introduction

When the very best oscillator performance is required, it is necessary to provide the oscillator with a properly designed ALC system (Automatic Level Control System), sometimes erroneously called an AGC system, for the following reasons:

1. The system maintains the crystal current, Ix, constant and thus avoids the frequency instabilities produced by Ix drifts or shifts.
2. It permits linear operation of the frequency generating active device, which results in the highest signal to noise ratio and, consequently the lowest phase noise.
3. It provides a high starting loop gain for dependable oscillator starting.

The disadvantages of the ALC system, when compared to self limiting systems is that it is more complicated and costly and therefore perhaps a bit less reliable. It also appears to be much more difficult to quantitatively understand and design. The difficulty is primarily due to the interactions between the RF and the baseband (DC and AF) operations.

The literature, too, pays little attention to the ALC system and all its references can be characterized as being hazy, qualitative, and generally inadequate. One of the better references, Chap 14 of Ref 3, also needs update, clarification, and expansion. The material in this paper has been prepared to satisfy that need. It serves as an addition to Chap 14 and not to supplant Chap 14.

The theory, titled above, developed much later than Chap 14, sheds new light upon this subject and this paper therefore treats the ALC system using the theory as a basis.

The nomenclature and symbols are those in Ref 3 complemented by Ref 1. The reader's particular attention is directed to the List of Symbols beginning on Page 441 of Ref 3.

2. Steady State Oscillator ALC System Fundamentals

The oscillator carrier amplitude is indirectly controlled by the real parts of basic Eq 14a or 16a, of Ref 1,

$$Rt = \Sigma R = dR \qquad (1)$$

which requires that the value of a parameter, Pa, of a physical component be set so that the equation is satisfied. The system performing these operations is called the ALC system and is represented by dR (see Ref 1). In every case the direct purpose of the variation in Pa is to produce a corresponding variation in the oscillator negative resistance, RN, [1][2] so that

$$RN = f_{RNPa}(Pa) \qquad (2)$$

It is also very important that the variation in RN be accomplished without an accompanying variation in frequency.

0-7803-2500-1/95 $4.00 © 1995 IEEE

f_{RNPa} (Pa) in Eq 2 is read as the value of RN at location Pa in the RNPa domain.

3. Linear Oscillator

3.1 General (See Fig 1b)

The linear oscillator, is defined as one in which the output of the frequency generator active device, Q1, FGAD, is directly proportional to its input. Also, the input signal amplitude is small (less than 20 mV) so that small signal parameters can be used to describe Q1.

In the linear oscillator, Eq 1 is satisfied over a wide range of oscillator amplitudes. The desired amplitude, V_f, is set by the ALC system consisting of an output amplitude, V_f, sensor controlling a DC voltage generator, detector, that produces V_O, proportional to V_f, after filtering by low pass filter, LPF,. V_O is subtracted from a DC reference voltage V_{REF} to create V_E so that

$$V_E = V_{REF} - V_O \tag{3}$$

V_E, is amplified by K_3 to V_C so that

$$V_C = K_3 (V_{REF} - K_1K_2K_{LPF}V_f) \tag{3a}$$

Note that in many ALC systems, K_3 is not necessary, i.e. $K_3 = 1$.

V_C in turn controls Pa. Only 1 value of V_C, V_Co, can cause Eq 1 to be satisfied. V_Co is unique in that its value is dependent only upon the RF circuitry. Any variation of V_Co with V_f is attributable to the oscillator non linearity and can be reduced by improving the linearity.

Since, as stated above, V_C controls Pa, then

$$Pa = f_{PaVC}(V_C) \tag{4}$$

and combining Eqs 2 and 4,

$$RN = f_{RNVC}(V_C) \tag{5}$$

V_Co, can be determined by means of

3.1.a. Direct measurement of V_Co in a standard prototype oscillator.

or

3.1.b. Computation of V_Co by using BPT [1][2] or equivalent computer program. It is important that the component values, entered into the computer, be close to those in the real oscillator at the operating frequency.

Normally, Method b is easier and much faster.

3.2 Starting Loop Gain, ALo

ALo should, preferably, be at least 2 to ensure dependable starting.

The starting loop gain, ALo, is controlled by the ALC system to be larger than 1, since at starting,

V_f --> 0 and V_C --> V_{REF} * K3

Note that for $V_C > VsatK3$, $V_C \approx VsatK3$

where VsatK3 is the saturation output voltage of amplifier K3.

Let RNo = the steady state RN (6)

at $V_C = V_Co$ (6a)

and RNs = the starting RN (7)

at $V_C = V_Cs$ (7a)

then ALo = RNs / Rno (8)

$= f_{RNVC}(V_Cs) / f_{RNVC}(V_Co)$ (9)
from Eqs 5 to 8.

Also, an approximate value of ALo can be experimentally obtained by adding a high resistance, Ra, in series with the crystal, at $V_C = V_Cs$, and gradually decreasing Ra until the oscillator just barely starts. Then, ALo $\approx (R_x + Ra) / R_x$ (R_x is the crystal resistance).

3.3 ALC LOOP BLOCK DIAGRAM

Fig 1a shows the block diagram of a typical ALC system. Fig 1b shows the same diagram redrawn to demonstrate that the block diagram of the regulator type of control system is also applicable to the ALC system. The tremendous amount of theory developed to analyze and design the regulator control system is therefore completely usable for the ALC system.

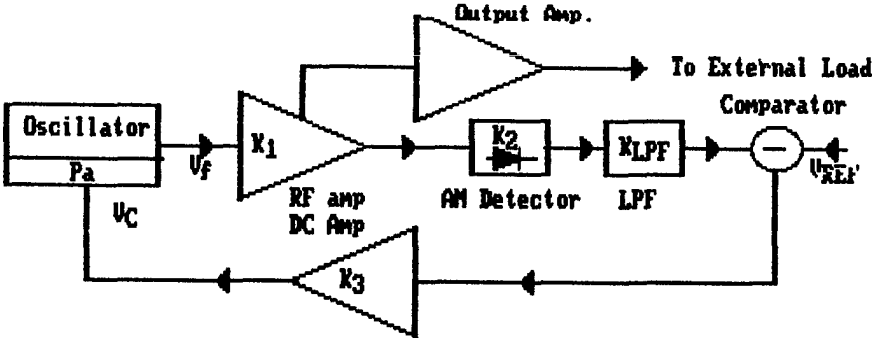


Fig 1a BLOCK DIAGRAM OF BASIC ALC SYSTEM

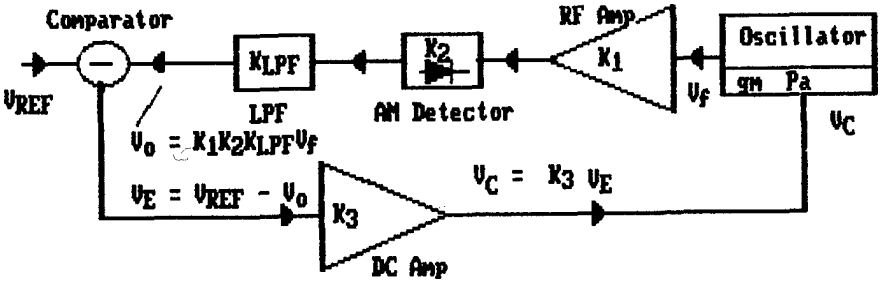
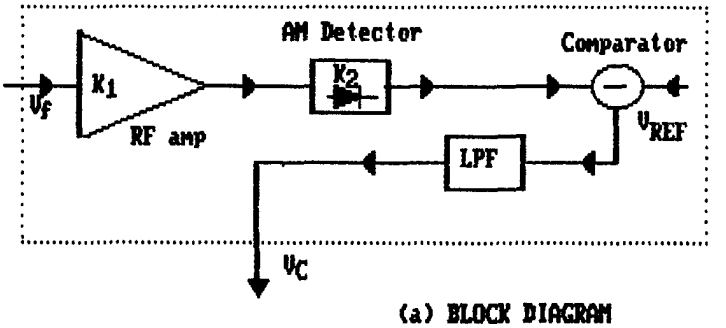
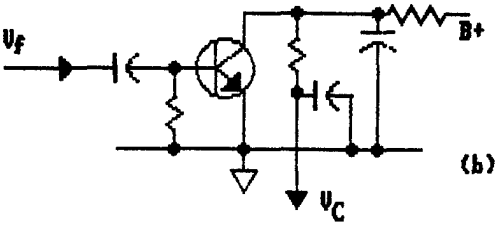


Fig. 1b BLOCK DIAGRAM OF ALC SYSTEM
Redrawn to show similarity to REGULATOR SYSTEM



(a) BLOCK DIAGRAM



(b) TYPICAL SCHEMATIC

Fig 2 DCVREF

The loop configuration is strongly dependent upon the type of Pa which must have the capabilities to

a. Provide an ALo of at least 2.

b. Include adequate rejection of the undesired B mode and overtones responses of the crystal.

c. Perform the necessary amplitude adjustments without simultaneously adjusting the frequency.

The selection of Pa is dictated by the oscillator circuitry and falls into 2 major categories 1. For Pierce family oscillators. 2. For oscillators in which the crystal current is also the equivalent emitter current of the FGAD.

3.3.1 Pa for the Pierce Oscillator Family For these oscillators, Pa can be the effective transconductance, gme, of the frequency generating active device, FGAD, provided that concurrent frequency variation is minimized by using a high Beta and f_T FGAD. The gme variation can be produced by 1 of 2 methods.

3.3.1.a Method 1 Voltage V_C is applied to the base of the FGAD. The small signal gme is varied by varying the DC base voltage which in turn changes the DC collector current, I_C, and

$$gm\ (MHO) \approx 38 * I_C\ (AMP) \qquad (9)$$

This is the simplest, least cost, and most common method. Its main disadvantage is that a large loop gain, ALo, is difficult to realize when amplifier K3 is absent ($K_3 = 1$).

3.3.1.b Method 2 Voltage V_C is applied to the base of a JFET, Q2, operating as a voltage controlled variable RF resistance, Rv, between the drain and source, controlled by V_C, DC isolated and in parallel with a high value resistor to shunt the DC emitter current, is connected in series with the emitter of Q1. Then

$$gme = gmo / (1 + Rv * gmo) \quad \{gm \text{ in Mhos}\} \quad 10)$$

3.3.2 Pa for oscillators in which the crystal current is also the emitter RF current of the FGAD.

Sect 3.3c requires that Pa should adjust RN without disturbing the frequency. A study described in Sect 8 of Ref 1 discloses that this type of oscillator is without a Pa as satisfactory as gme in the Pierce oscillator family. The best that can be achieved is typical of that shown in Fig 5 of Ref 1 and then, only when the oscillator is carefully tuned. Accordingly ALC systems are not recommended for this type of oscillator when the best frequency performance is mandated.

3.3.3 Loop Performance

Performance relationships have been developed in Sect 14.2.2 of Ref 3 and are repeated here for convenience.

$$V_C = K_3 (V_{REF} - K_1K_2K_{LPF}V_f) \qquad (3a)$$

or

$$V_f = (1 / K_1K_2K_{LPF}) (V_{REF} - V_C / K_3) \qquad (14.2a)$$

$$\text{let} \quad V_f = r_{os} Ix \qquad (14.3)$$

then

$$Ix = (1 / r_{os}K_1K_2K_{LPF}) (V_{REF} - V_C / K_3) \qquad (14.6)$$

from which

$$\delta Ix = (1 / r_{os}K_1K_2K_{LPF}) (-\delta V_C / K_3 \qquad (14.7)$$

and

$$\delta Ix / Ix = -\delta V_C / (K_3V_{REF} - V_C) \qquad (14.8)$$

Eq 14.8 describes the fractional change of Ix caused by the change in V_C due to component changes over time and environmental stresses.

It shows that δ Ix / Ix is principally determined by δV_C, K_3, and V_{REF}. This does not imply that the stability of the other components is not important. Their effects show up when computing absolute changes as in Eq 14.7.

For a description of the components and their requirements see Sect 14.3 of Ref 3.

3.3.4 Combination detector, comparator, and V_{REF} source, DCVREF (See Fig 2)

This device consists of a single transistor, the input to which is V_f. The transistor is biased so that its output, V_C, is $(V_{REF} - K_1K_2V_f)K_{LPF}$

$$\equiv V_{REF} - K_1K_2V_fK_{LPF}, \text{ since } V_{REF} \text{ is pure DC, the effect of filtering it is nil.}$$

so that

$$V_C = V_{REF} - K_1K_2V_fK_{LPF} \qquad (11)$$

All the equations in Sect 3.3.3 are still valid. However, the values of V_{REF}, and K_2 are greater in Fig 2 than in Fig 1a. In particular, typical values of the most important parameter, V_{REF}, are 6 in Fig 1a and 36 in Fig 2.

The DCVREF is very attractive because of its econ-

omy but the reader is cautioned about the use of reference voltages that are ill defined and not explicitly identifiable as stated in Sect 14.3.6 of Ref 3.

3.4 Phase noise performance of ALC systems.

Usually the ALC oscillator can be expected to have better noise performance than the equivalent self limiting oscillator. However, the reader is cautioned that the noise sidebands close to the carrier are partially dependent upon the baseband characteristics such as the crystal time constant, the low pass filter, LPF, characteristics and K3 gain. If the LPF is improperly designed large spikes close to the carrier may be present in the noise vs. Fourier frequency characteristic. Unfortunately, not much more on this important subject can be said, at this time, because of the necessity of preserving confidentiality.

4. Design of ALC Systems

4.1 Major input specifications

4.1.1 $V_{C}o$ - Determined by direct measurement or computation, as described in Sect 3.1a and 3.1.b.

4.1.2 $(\delta V_{C}o)$ max This is the maximum expected change in $V_{C}o$. This is principally caused by changes in crystal resistance and in transistor characteristics. Nothing much can be done about the crystal resistance beside using a better crystal or improving the crystal environment. However, the transistor can be and should be stabilized as required.

4.1.3 $(\delta Ix / Ix)$ max. This is the maximum $(\delta Ix / Ix)$ corresponding to the $(\delta V_{C}o)$ max to be designed into the ALC system.

4.2 Calculation of $(\delta Ix / Ix)$ max

This is done with Eq 14.8. If $(\delta Ix / Ix)$ max is satisfactory for the avilable V_{REF} and $K_3 = 1$, then make $K_3 = 1$, i.e. no K3. Otherwise make K3, the calculated K_3 for the available V_{REF}. The design is now complete.

5. Conclusions

The analysis of ALC systems, based upon the titled theory and upon Chpt 14 of Ref 3 has been developed. The analysis clearly identifies the baseband simultaneously active modes of operation, shown in the following descending order of precedence:

1 .The principal RF mode, in accordance with the titled theory, and in which the contribution of the other modes evidence themselves as dR for the DC mode and the effect of the ALC loop characteristics upon the close in phase noise as the AF mode.

2. The DC mode, which sets the value of dR and the consequent magnitude of the carrier amplitude.

3. The AF mode, which contributes to the stability of the ALC loop and to the carrier and noise content of the V_C signal. As a side effect, it modifies the close in phase noise. This is transmitted to the RF mode by modulating the value of dR.

It should be noted that, while the treatment is slanted towards the crystal oscillator, all the material is applicable to all oscillators.

The paper also includes material on ALC system design. This material follows directly from the ALC system analysis.

References

[1] B. Parzen, "Universal, Computer Facilitated, Steady State Oscillator Analysis Theory and Some UHF and Microwave Applications", Proc. 45th Annual Frequency Control Symposium, pp. 368-383. May 1991.

[2] B. Parzen, "Universal, Computer Facilitated, Steady State Oscillator Analysis Theory, Some Applications to Precision Crystal Oscillators", Proc. 46th Annual Frequency Control Symposium, pp. , May 1992.

[3] B. Parzen, Design of Crystal and Other Harmonic Oscillators. New York: Wiley, 1983.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DTXCO THERMAL CONVERTER REALIZED BY MEANS OF A METAL FILM TEMPERATURE SENSOR DEPOSITED ON THE CRYSTAL PLATE

V. P. Bagayev, B. P. Ionov, S. A. Zavyalov,
Omsk State Engineering University,
11, Mira av. , 644050, Omsk, RUSSIA

Abstract

Temperature compensated crystal oscillator parameters are dependent to great extent on the accuracy of crystal temperature measurements. The most prominent results were attained for DTCXO with resonator - temperature sensitive element configuration based on dual-mode excitation principle. In this case parameter high stability is attained and the temperature lag between temperature sensor and a measurement object is practically avoided.

The investigation of an alternative crystal temperature measurement procedure shows that high DTCXO performances can be attained without dual-mode excitation. A thin metal film deposited along the crystal plate contour is used as a primary thermosensor.

To increase the stability and the accuracy of temperature measurement a thermal converter had been developed using the principle of digital-analog system of automatic frequency control.

Using by way of example a simple DTCXO with AT-cut crystal and piece-wise compensation function theoretic and experimental investigation were carried out.

Introduction

Much attention is being given to the problems of DTCXO parameters perfection. At least three rather independent problems could be marked out for DTCXO:

1 - crystal temperature measurement; 2 - thermosensor signal functional conversion according to f-T curve of the crystal; 3- oscillator output frequency controlling or synthesizing. In the present paper the problems are considered of crystal temperature measurement and pre-sensor signal conversion into a form convenient to be used by the following DTCXO circuit. The best results for DTCXO were obtained when dual mode excitation of TD-or SC- cut crystal resonators were used [1,2]. In this case practically perfect geometric coincidence is attained of a thermosensor and an object of measurement, which makes possible to obtain DTCXO frequency temperature instabilities in the range of 0.1...0.3 ppm and less. However after rather long investigation of the dual mode operation a number of problems became evident: 1 - it is difficult to realize a resonator free of thermosensitive B-mode anomalies in a broad temperature range; 2 - an oscillator circuit becomes rather complicated; 3 - spurious components with B- mode frequency are present in output spectrum when a DTCXO is composed according to the traditional circuit without output frequency synthesis.

There was an attempt to replace the dual mode resonator excitation by an alternative solution [3]. A thermosensor was realized as a metal film deposited on the piezoelement surface. Such configuration thermosensor-resonator shows minimal temperature lag. Resistance to frequency conversion was performed by a special circuit designed as a lock system. The results of this solution investigation are stated in the paper.

DTCXO.Temperature meagurement problems.

Let us examine a typical DTCXO configuration (fig.1). Depending on f-T curve of the crystal resonator used temperature mesurement error causes an output frequency drift (in fig.2 the dependence is shown for AT-cut). In static temperature conditions temperature difference between the sensor and the resonator is practically absent. In dynamic conditions it becomes evident depending on the speed of the temperature variation and on the construction of the resonator - thermosensor system. Several types of constructions had been investigated by these authors earlier [4] from the view point of temperature - dynamic properties. The best results were obtained with dual mode resonators. Further investigation was carried out of configurations where a metal film or a filmed thermoresistor deposited directly on a piezoelement surface was used as a thermosensor [5].

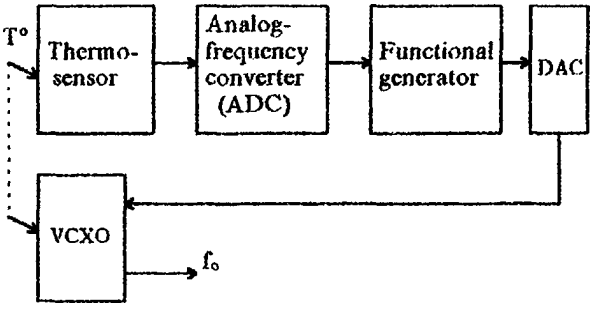


Fig.1 The simplest DTCXO block diagram

0-7803-2500-1/95 $4.00 © 1995 IEEE

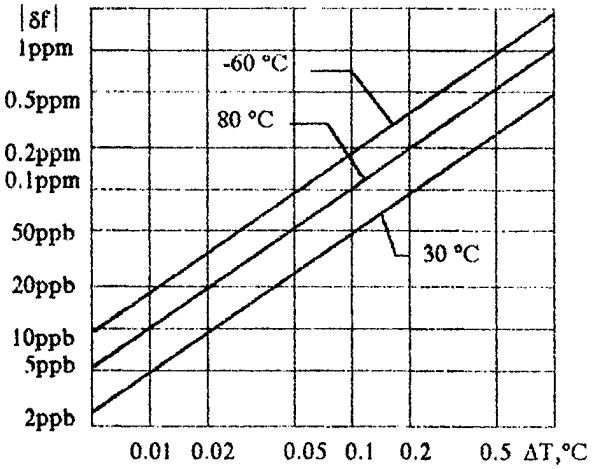


Fig.2 DTCXO output frequency variation versus temperature measurment error for different temperature points (AT-cut crystal used).

The results of investigations carried out accoding to the thermal time constant method [4] showed that such configurations possessed practically the same dynamic temperature characteristics as those of dual mode versions. The temperature lag with time constant of 70...75 s does not exceed 1 s. This leads to a dynamic temperature output frequency drift of less than 10 ppb when temperature variation speed is 1 K/min and maximum f-T curve slope is 0.5 ppm/K..

A schematic of crystal plate construction with a film thermosensor is given in fig.3. Similar film configuration is used in direct oven controlled crystal oscillators. In this case of thermocompensation maximal ohmic resistance as well as resistance temperature coefficient are necessary. Curves of resistance variations in the temperature range are given in fig.4.

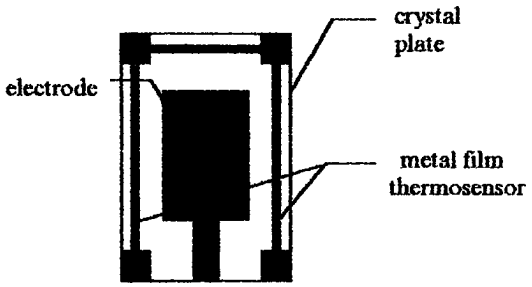


Fig.3 Piezoelement construction with a metal film thermosensor

An alternate solution of the primary sensor is a thermoresistive layer deposited on crystal plate surface and used as a temperature sensor in OCXO. An obstacle for such sensor utilization in DTCXO is high non-linearity of resistance- temperature curve.

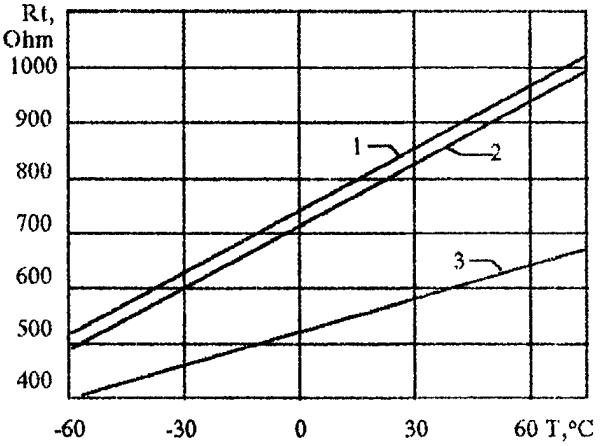


Fig.4 Experimental R-T curves; double-branch (1,2) and one-branch(3) metal film thermosensor

Problem of temperature sensor signal conversion

It should be noted here that there is no such problem for the case of DTCXO with frequency thermosensor (e.g., in B-mode). For analog temperature sensors the traditional way of signal conversion is by means of an analog-to-digit converter. At first glance high performance of modern A-D converters suffice all the requirements. However, there arise some problems.

1.With high levels of DTCXO's frequency stability the effects of A-D converter reference voltage variations are still remain valid in dynamic temperature conditions.

2. When A-D converters are used short-term stability is essentially deteriorated at the moments of digits switching. So, with the accuracy of temperature measurements of 0.1° (11 bit A-D converter) a maximal frequency increment will be roughly 0.05 ppm. This is a high level for DTCXO temperature stability, but when temperature fluctuations occur such spectrum deviations are rather large.

3. A low value of intrinsic conversion transconductance of a metal film sensor and nonlinearity of a thermoresistive sensor require to introduce networks of amplification and conversion. This leads to additional errors of temperature measurement.

Resistans - frequency converter based on AFC system.

Automatic frequency control (AFC) system used for converting resistance into frequency or a period enables to eliminate some errors. Let us consider the converter given in fig.5. A film thermosensor R_t is included into a branch of a bridge in series with precision resistor R_o possessing a minimum temperature resistance coefficient. The second branch of the bridge is realized as a controlable CMOS switch. In this way a pulse-time

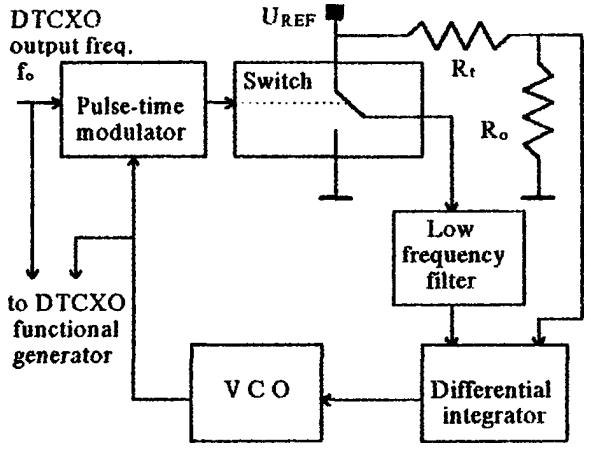


Fig.5 AFC thermoconverter block diagram

modulator can control the period-to-pulse duration ratio of the switch which ensures the equality of voltage constant components at the input of differential integrator. In other words the AFC system is constantly maintaining the bridge balance and so the reference voltage fluctuations effect on the converter parameters is practically eliminated. Intrisic temperature and long-term frequency instabilities of the VCO also do not effect the converter parameters because VCO frequency in the AFC loop is in fact compared with the output frequensy of the DTCXO (frequencies f_o and f_t ratio is being analyzed) and is supported at the needed level independently of the rest external conditions.

The most evident effect resulting with additional conversion errors is due to bias drift and input current difference drift in an operational amplifier used in the differential integrator circuit. Converter dynamic properties defined by such parameters as settling time and short-term stability of a thermodependent frequency can be optimized by proper selection of the integrator time constant.

The action of the system is performed as follows. An error signal taken from a bridge diagonal is fed at the differential integrator. Integrator output voltage variation causes the variation of VCO frequency. VCO output signal and a reference signal from DTCXO output are applied to the pulse- time modulator which forms a pulse sequence controlling the CMOS switch. VCO frequency variation leads to the variation of voltage constant component at the output of the switch and a low frequency filter so that the error signal becomes less and, as a result, tends to zero.

The form of the resistance - frequency or resistance - period conversion characteristic depends on the way of the pulse - time modulator composition. Six different versions can be distinguished (fig.6). For fig.6a and 6b pulse consequence period is constant and makes up N periods of τ_o of DTCXO output signal. Positive (6a) and negative (6b) pulses durations are variable and make up K periods of τ_t of VCO signal. For fig.6c and 6d pulse

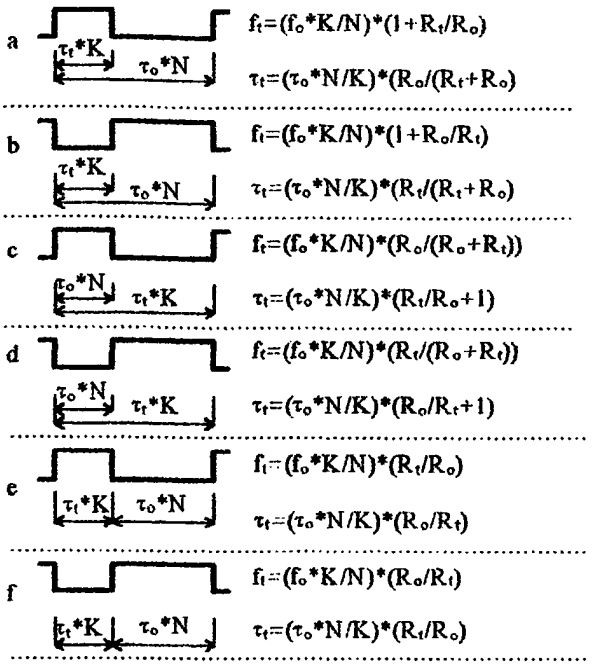


Fig.6 Versions of thermoconverter characteristic versus pulse-time modulator output signal

consequence period is variable and makes up K periods of τ_t . Pulse duration is constant and makes up N periods of τ_o. For fig.6e and 6f consequence period is composed of two parts: a constant one - $N*\tau_o$ (negative (e) and positive (f)) and a variable - $K*\tau_t$ (positive (e) or negative (f)).

The functional relations gained are of a linear and hyperbolic nature which is suitable for converter realization by means of a metal film thermosensor with its linear R - T curve and by way of partial compensation of nonlinear thermistor sensor R - T curve. The choice of one or another composition version for pulse-time modulator is decided by the level of circuit simplicity and compatibility with the rest part of DTCXO circuit. It is desirable also to have the possibility to use both metal film and thermistor temperature sensors.

Conversion parameters are to great extent dependent on resistor R_o nominal value. The choice of the resistor should be a compromise: decreasing of R_o/R_t ratio on one side causes the increase of primary conversion resistance-voltage transcondactance; on the other side the current across the thermosensor is growing. The primary conversion transconductance increase contributes in reducing effects of additional errors induced by the differential integrator. The growth of the current through the sensor leads to its self-heating and to the increase of power consumption which becomes actual when a low - ohmic metal thermosensor is used. Limiting the current by the value of 0.3...0.5 mA the value of R_o will be of 10 ...15 kOhm. When $R_o \gg R_t$ practically all the versions of pulse-time modulator composition are suitable for DTCXO with respect of linearity.

An experimental investigation was carried out of a converter circuit with a pulse-time modulator according to to the version 6b, having coefficients $N=2^{15}$, $K=5$ and $R_o=15kOhm$. Experimental conversion characteristics for temperature range of -60...+80 °C are shown in fig.7. Experimental investigation of short- term stability and the settling time versus inegrator parameters are shown in fig.8 and fig.9. The comparision shows that AFC thermoconverter output charasteristics are near to those of dual-mode oscillator thermosensitive output.

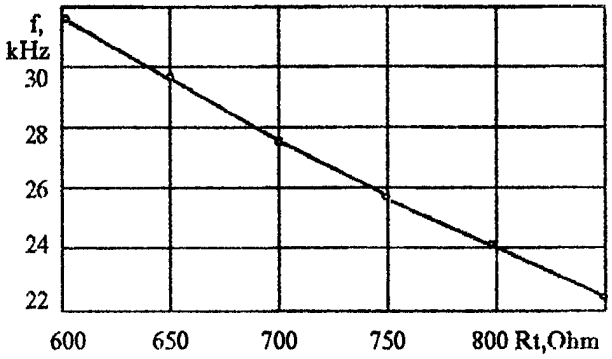


Fig.7 Experimental frequency versus resistance characteristic of a thermoconverter

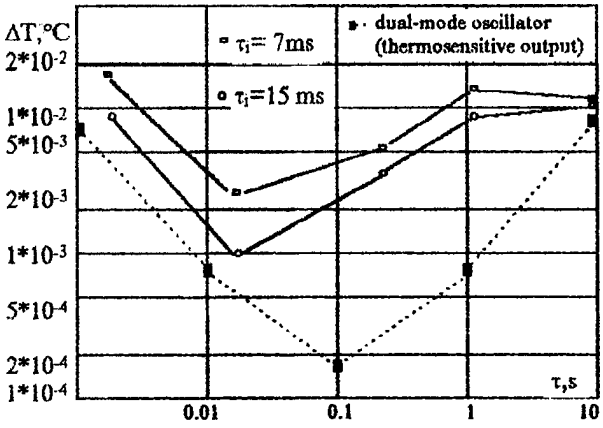


Fig.8 Thermoconverter and dual mode oscillator (B-mode) output signal short-term stability.

DTCXO Experimental results.

A specialized CMOS chip intended for application in DTCXOs with stability level 0.3...1ppm has been developed which combines main digital units of both temperature-frequency converter and subsequent part of

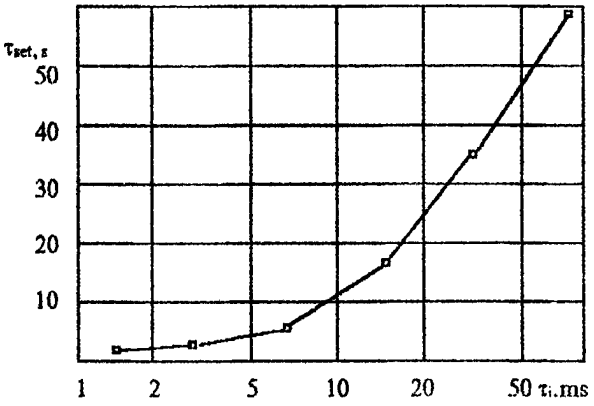


Fig.9 Thermoconverter frequency settling time versus integrator time constant (for precision < 0.01°C)

a thermocompensation system. DTCXO block-diagram is given in fig.10.

The device has been composed according to the conventional series scheme: a temperature sensor and a thermoconverter - a functional generator - a DAC - a voltage controlled crystal oscillator. In the VCXO an AT-cut crystal was used excited on the 1ST oberton. For the sake of simplicity a pulse 12-bit DAC was used in the digital part of the DTCXO alongside with piecewise linear forming of the compensation function [6]. This enabled to essentially lessen the capacity of the ROM used (only 64 12 -bit words for a temperature range of --60 ... +80 °C). A special device is provided for synchronisation of interaction between the cotrolable switch of the thermoconverter and the counter ,the pulse fed ROM and the DAC with period about 3 ms. The results of DTCXO temperature testing are given in fig.11, and experimental short-term frequency stability investigation — in fig.12.

Conclusion

Investigations carried out have shown that a metal film deposited on a crystal plate together with electrodes, for one fabrication cycle can be used in DTCXO as a thermosensor. A converter based on an AFC system and adapted for film thermosensors implementation in DTCXOs permits to obtain oscillator parameters of precision and temperature dynamics near to those obtained with dual-mode excitation crystals of double-rotated cuts. Utilization of specialized chip does not practically complicate the DTCXO circuit.

Further oscillator parameters perfection is possible through the use of more precise micro power operational amplifiers , by in-fabrication increasing resistance of the film thermosensor and using one mode SC-,TD- cuts oscillators . This makes thermoconverter utilization promising also for high-stability TCXOs with sythesized output frequency.

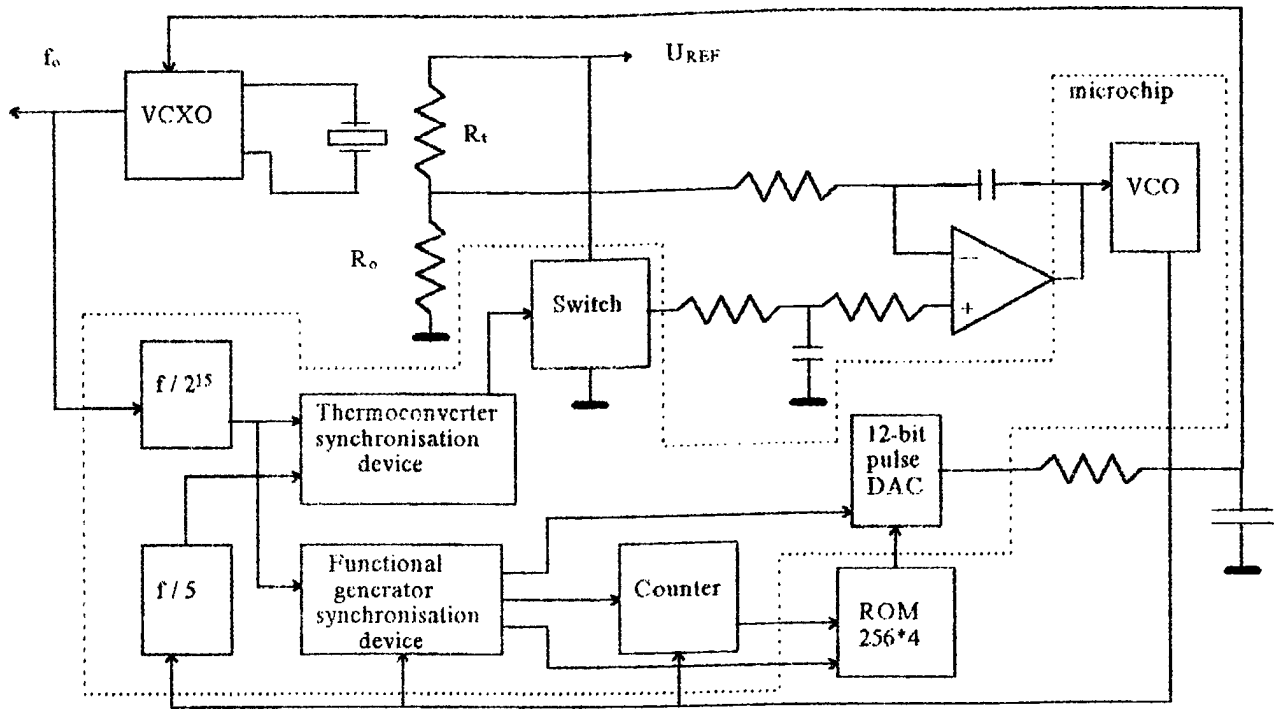


Fig.10 DTCXO block diagram

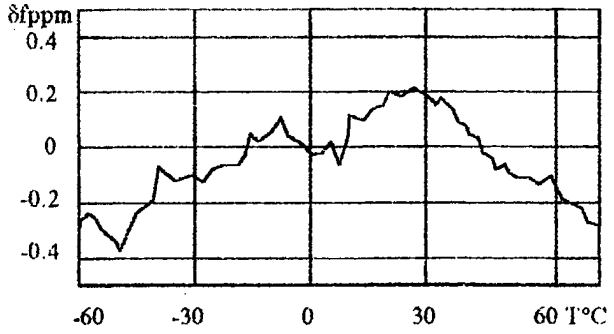


Fig.11 Experimental DTCXO f-T curve

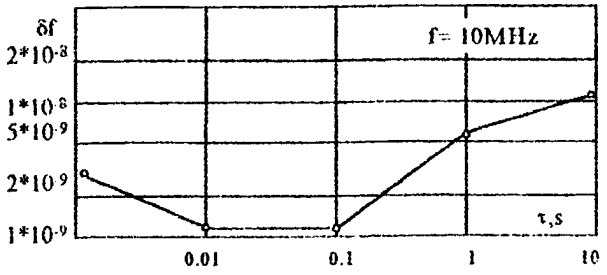


Fig.12 DTCXO output frequency short-term stability

References

[1] R.L.Filler, J.R.Vig. Resonator for the microcomputer compensated crystal oscillator (MCXO). Proc. 43 A.F.C.S. , 1989.

[2] I.V.Abramson. Two-mode quartz resonator for digital temperature compensated quartz oscillators. Proc. 46 A.F.C.S. , 1992.

[3] Patent RU 2017087 C1/ Багаев В.П., Завьялов С.А. Датчик температуры с частотным выходом. Опубл. Б.И. N° 14 , 1994 .

[4] A.Kosykh, B.Ionov, A.Vasiliev. Temperature dynamic model and temperature dynamic compensation of crystal oscillators. Proc. 46 A.F.C.S., 1992.

[5] Inventantion sertificate (Patent) N°1105946 (USSR) Способ изготовления тонкопленочных термо-резисторов. / Н.И.Алексеева, А.Н.Дикиджи, М.И.Ярославский. Опубл. Б.И. N°28, 1984.

[6] A.Kosykh, V.Bagayev, B.Ionov, A.Lepetaev, S.Zavyalov, A.Vasiliev. The new method of statistic piecewise-linear interpolation and its application to DTCXO creation. Proc.47 A.F.C.S., 1992.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THE MODULATIONAL METHOD OF THE PRECISION QUARTZ-CRYSTAL OSCILLATORS AND STANDARDS FREQUENCY STABILIZATION

Yu.S.Shmaly

Kharkiv Military University & "Sichron" Centre,
"Sichron" Centre, 4 Skripnika Street, Kharkiv,
310057,UKRAINE

ABSTRACT

A new trend in self-contained (without the use of quantum discriminators) frequency stabilization of oven control crystal oscillators (OCXO) and frequency standards is proposed and discussed in this paper. The method developing the trend is called a modulational method and is based on the use of metrological properties of natural volume oscillations of a quartz resonator (double-frequency and multifrequency oscillators are not used in this case).

The concept is given of dynamic modulation characteristics of a quartz-crystal oscillator and basic relationships are found for their calculation in the field of unharmonic resonances of a quartz-crystal resonator. The construction principles of frequency stabilization systems are formulated substantiating mathematically the essence of the method. Basic ratios of modulating signals are determined, the solution of which provides a slight influence of the modulation signal on the Allan variance and spectral density of a quartz-crystal oscillator.

The results of the practical use of the method are considered with the view of constructing a quartz-crystal oscillator with the adaptation to the environment temperature and a quartz-crystal frequency standard with the long-term frequency instability compensation.

INTRODUCTION

An essential shortcoming of precision quartz-crystal self-excited oscillators (XO) is the high-level of operational instabilities dependent on the temperature, stress,degradation processes of aging, etc.[1,2] and exceeding the Allan variance level by some orders of magnitude. Modern methodology of this shortcoming elimination is concentrated in two directions, traditional one consisting in technology and circuit design refinement of basic oscillator structures and quartz units constructions[1-9]; and non-traditional one consisting in creation of the principally new structures of multifrequency oscillators and resonators on the basis of the use of a nonmonofrequency properties of the quartz units as a cavity resonator[3,10]. Thanks to the progress in the first direction the frequency instability of a self-excited XO diminishes on the average by an order of magnitude[11]. The second trend is at the development stage and hasn't yielded any substantial results yet.

The typical curve of the relative frequency instability variation of OCXO under the influence of various internal and external destabilizing factors is given in Fig.1[3]. Taking into account the

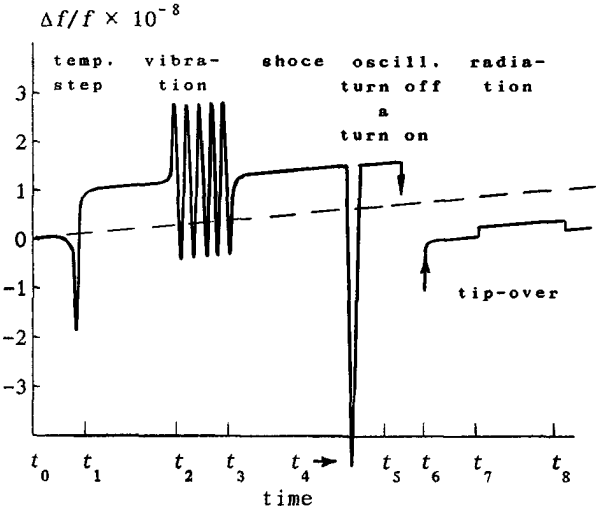


Fig. 1. Types of frequency instabilities .

volume character the quartz resonator vibrations it becomes obvious that the low (Fig.1) is qualitatively characteristic of the whole spectrum of volume acoustic oscillations of the quartz resonator generally defined by A, B, C modes, their harmonics and unharmonics.

If we determine first the laws of variation of n oscillation frequencies of the quartz resonator under the effect of n-1 destabilizing factors and excite all the n frequencies during the OCXO operation, then by independent formation and consequent solution of n-1 equations system in n-1 unknowns it is possible in principle to determine all unknown frequency instabilities and to carry out their compensation.

This approach which is used, for example, in[3,10], to constructing TCXO on the basic of the SC-cut resonator, yields very modest results because of a very important restriction consisting in thermodynamic interaction of the fundamental and additional oscillations of the quartz resonator observed with two- and multimode excitement of the

0-7803-2500-1/95 $4.00 © 1995 IEEE

oscillator. The decrease of the excetement level of additional excillations, necessary for decreasing the thermodynamic interdependence results in the situation which is more dangerous - operation mode instability of the multifrequency oscillator.

A combined trend in self-contained (without the use of quantum discriminators and oscillators) frequency stabilization has been proposed and analyzed in this paper in which the advantages of the first and the second trend are integtated on the basic of the use of internal natural redundancy of the crystal resonator frequency spectrum (unharmonic resonances and additional modes) due to nonuniformity of dynamic modulational (DM) characteristics of the crystal oscillator by introducing a modulating signal in the circuit. Precision self-excited crystal oscillators (XO) realized within the scope of the method which is developing the trend and is called modulational, turn out to be free of the shortcomings of the second trend having smaller operational frequency instabilities in comparison with XO implemented within the scope of the first trend.

THE ESSENCE OF THE MODULATIONAL METHOD

Let us consider a quartz resonator which is characterized, for example, by three modes h_{n11}, h_{n12}, h_{n13} in the operational frequency range, where n is the number of the mechanical harmonic, 11, 12, 13 $\in$ pq, pq $\in$ 1,∞ is the oscillation index (Fig.2a). The amplitude-frequency characteristic $K_{QR}(f)$ of the quartz resonator in this range has the form shown in Fig.2b. During the excitation of the quartz resonator in the XO circuit due to the greater activity of the h_{n11} mode the oscillation with the f_o frequency in the inter-resonance gap is set and with the spectral density of the phase fluctuations power $S_\varphi(f)$, which has nonuniformities in the vicinity of the h_{n12} and h_{n13} modes. The energy distribution in nonuniformity zones is subject to the Rayleigh-Rice law[12] (Fig.2c). Under the FM-AM of the XO signal with a rather small index of FM $\varepsilon_{FM} \ll 1$ and the depth of AM $\varepsilon_{AM} \ll 1$ two lateral components appear in the oscillations spectrum S_{FM-AM} the amplitudes of which vary with the variation of the modulation frequency in the vicinity of the h_{n12} and h_{n13} modes according to the laws (Fig.2d)[13,14].

The amplitude variation of the upper component of the FM-AM spectrum results in nonuniformities in the amplitude-frequency and plase-frequency modulation characteristics of FM $K_\omega(F)$, $\varphi_\omega(F)$ and AM $K_A(F)$, $\varphi_A(F)$, where $F=f-f_o$[14], (Fig.3), which is displayed in practice by variations of the FM and AM depths as well as by phase shifts between the FM, AM envelopes and the modulating signal in the narrow modulation frequencies band F', F'', $F''' \in F$, equal to the passband of the equivalent circuit of the quartz resonator of the h_{npq} mode.

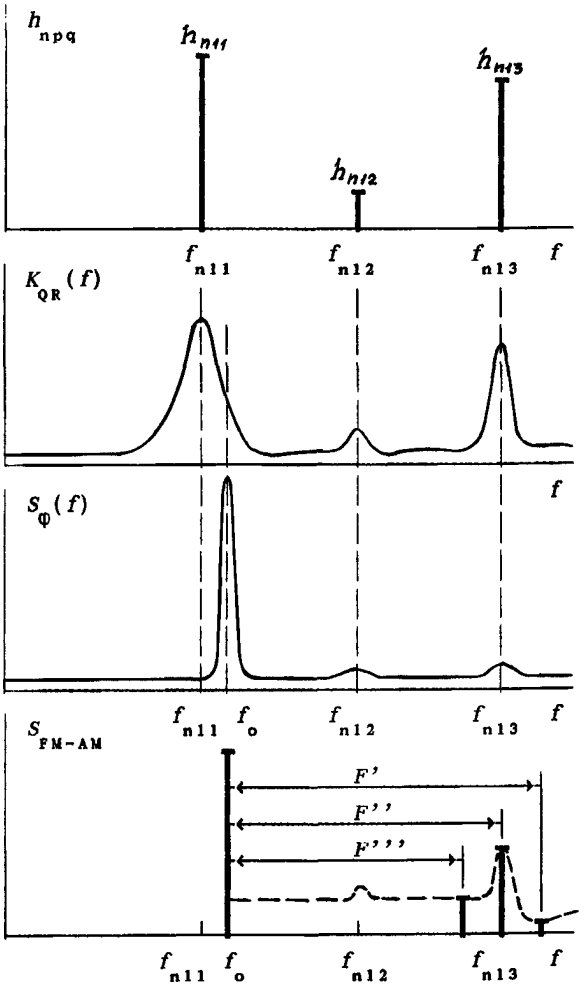


Fig. 2. The XO oscillations modes h_{npq} (a), amplitude-frequency characteristics of the XO (b), spectral power density of the XO phase fluctuations $S_\varphi(f)$ (c) and the XO oscillations spectrum with FM-AM S_{FM-AM} (d) in the narrow frequency band $F=f-f_o$ of the tuning away from the carrier f_o.

Under the influence of the θ component of the destabilizing factors vector $\vec{\theta}$ (temperature, aging, stress, etc.) the f_{npq} frequencies of quartz resonators of single and double rotated thickness shear vibrations change according to the law[15,16].

$$f_{npq}^2(\theta)=\frac{n\delta_1^2(\theta)}{4\pi h(\theta)\varrho(\theta)}\left\{1+\frac{h^{0,5}}{n\pi R^{0,5}}\left[\delta_2(\theta)(2p-1)+\delta_3(\theta)(2q-1)\right]\right\} \quad ; \qquad (1)$$

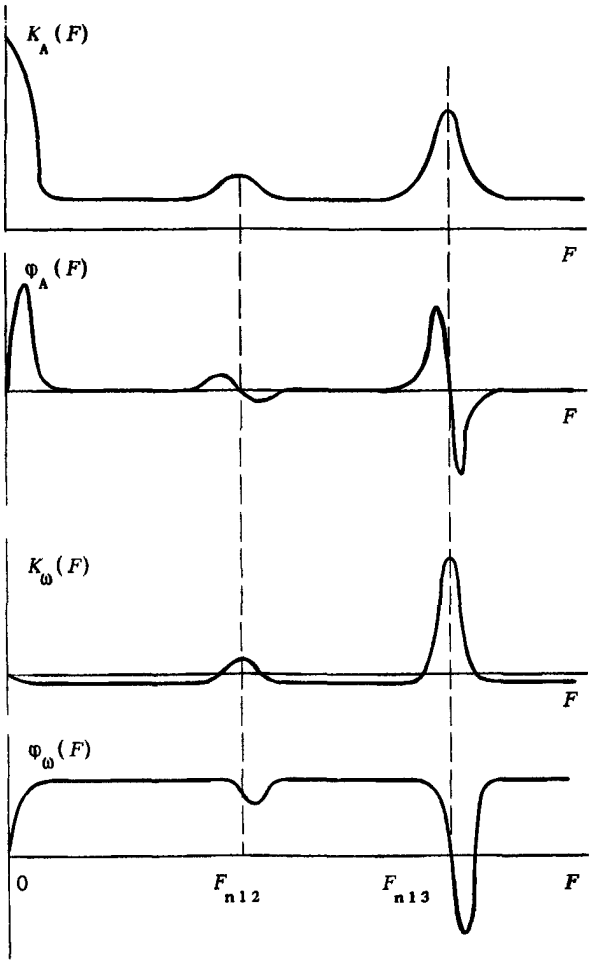


Fig. 3. The XO DM characteristics in the narrow frequency band of turning away from the carrier f_o .

where $h(\theta)$, $\varrho(\theta)$ is the thickness of the disc piezoelectric-crystal plate and the quartz density; $\delta_1(\theta)$, $\delta_2(\theta)$, $\delta_3(\theta)$ are the constants dependent on the orientation angles, geometry and physical properties of the piezoelectric-crystal plate and the electrodes. With this the parameters of DM characteristic change[14]

$$m_\omega(jF,\theta) = \frac{\Delta f(jF,\theta)}{f_o} = \varepsilon_\omega K_\omega(F,\theta)\exp[j\varphi_\omega(F,\theta)] \quad ;(2)$$

$$m_A(jF,\theta) = \frac{\Delta U(jF,\theta)}{U_o} = \varepsilon_A K_A(F,\theta)\exp[j\varphi_A(F,\theta)] \quad ;(3)$$

where ε_ω , ε_A are the coefficients of the static FM and AM; Δf, ΔU are the frequency deviation and the amplitude variation of the XO v signal in the FM-AM process. Using this property we can derive the information about the frequencies $F_{npq}(\theta)=f_{npq}(\theta)-f_o(\theta)$ behavior from the XO output signal, contained in the parameters of the DM characteristics, and convert it into the control signal affecting the control outputs of the frequency and XO operational mode; it enables us to construct quartz-frequency stabilization systems on this basis. This is the essence of the combined trend.

With this the essence of the modulational method is formulated in the following way[17,18] : we modulate the frequency (amplitude) of the XO generated oscillations with the F_{npq} frequency and with the depth which is enough to increase the oscillations energy of the XO at the f_{npq} frequency to the value with which the force oscillations of the h_{npq} mode would acquire the reference properties[12] enough to stabilize frequencies of the h_{n11} mode; then we derive the signal of the force oscillations of the h_{npq} mode with the use of DM characteristic, then we convert the signal parameters (frequency, phase, amplitude) into the control signals and use the latter to control the XO operational mode and f_{npq} frequencies.

FUNDAMENTAL SIGNALS AND RELATIONSHIPS

To justify the essence of the method the function of the XO output stress in FM-AM for the first harmonic of the oscillations frequency, ignoring the modulating frequency harmonics, is presented in the form of[13]

$$u(t)=U_o[1+m_A\cos(\Omega t-\varphi_A)]\times \exp\left\{j\omega_o\int[1-m_\omega\cos(\Omega t-\varphi_\omega)]dt+\varphi_o\right\} \quad , \quad (4)$$

where $\omega_o=2\pi f_o$; $\Omega=2\pi F$; φ_o is the initial phase[19-23]. The functions of the DM characteristics with the modulation in the band of the XO spectral line, where $S_\varphi(f)$ exceeds tens of dB/Hz, are found as[14]

$$m = \varepsilon_A\sqrt{1+\left(\frac{U_c a_{12}+U_s b_{12}}{U_c a_{11}+U_s b_{11}}\right)^2} \quad ; \quad (5)$$

$$\varphi_A = \operatorname{arctg}\frac{U_c a_{12}+U_s b_{12}}{U_c a_{11}+U_s b_{11}} \quad ; \quad (6)$$

$$m_\omega = \varepsilon_\omega\sqrt{\left(1+\frac{\Delta\omega_c}{\Delta\omega}\right)^2+\left(\frac{\Delta\omega_s}{\Delta\omega}\right)^2} \quad ; \quad (7)$$

$$\varphi_\omega = \operatorname{arctg}\frac{\Delta\omega_s}{\Delta\omega+\Delta\omega_c} \quad , \quad (8)$$

where U_c, U_s, $\Delta\omega_c$, $\Delta\omega_s$ - are the ortogonal stationary increments of the amplitude and frequency connected with the nonuniformities $K_{QR}(F)$; a_{11}, a_{12}, b_{11}, b_{12} are the variable dynamic amplitudes determinited for specific XO circuits.

With the frequencies $F \cong f_{npq} - f_o$, $pq \neq 11$ the relationship between the amplitude of the upper component of the signal spectrum (4) (Fig.2d) and the average anplitude of the white Gaussian noise occurring in this frequency region of tuning away from the carrier frequency f_o should be as small as possible which apriori provides an insignificant contribution of the modulation signal into the f_o frequency instability of the thermodynamic interaction mechanism[24-26]. The probabilities distribution densities of the DM characteristic functions (5) - (8) are found as[12]

$$W_1(m_{A,\varphi}) = \frac{m_{A,\varphi}}{\sigma} I_0(m_{A,\varphi} M_{A,\varphi}/\sigma^2) \times$$

$$\times \exp[-0{,}5(m^2_{A,\varphi} + M^2_{A,\varphi})/\sigma^2], \quad \forall\, m_{A,\varphi} \geq 0 \quad ; \quad (9)$$

$$W_1(\varphi_{A,\omega}) = \frac{\pi}{2} \exp\left(-\frac{M^2_{A,\varphi}}{2\sigma^2}\right)\left\{1 + \sqrt{2\pi}\,\frac{M_{A,\varphi}}{\sigma} \times \right.$$

$$\times \cos(\varphi_{A,\omega} - \psi_{A,\omega}) \Phi\left[\frac{M_{A,\varphi}}{\sigma}\cos(\varphi_{A,\varphi} - \Psi_{A,\omega})\right] \times$$

$$\left. \times \exp\left[\frac{M^2_{A,\varphi}}{\sigma^2}\cos^2\frac{\varphi_{A,\omega} - \Psi_{A,\omega}}{2}\right]\right\}, \quad \forall\, -\pi \leq \varphi_{A,\omega} \leq \pi \quad (10)$$

where $M_{A,\varphi} = \sqrt{M^2\{X_1, Y_1\} + M^2\{X_2 Y_2\}}$; $tg\Psi_{A,\omega} = M\{X_1 Y_1\}/M\{X_2 Y_2\}$; $M\begin{Bmatrix} X_1, & Y_1 \\ X_2, & Y_2 \end{Bmatrix} = \frac{U_B}{U_o}\left[(v_1 - 1)\begin{Bmatrix}\sin\varphi_B \\ \cos\varphi_B\end{Bmatrix} \mp \right.$

$$\left. \mp v_2\begin{Bmatrix}\cos\varphi_B \\ \sin\varphi_B\end{Bmatrix}\right] \mp \{m_{Ao,\varphi o}\}\begin{Bmatrix}\sin\varphi_{Ao,\psi o} \\ \cos\varphi_{Ao,\psi o}\end{Bmatrix}$$; σ^2 is the dispersion of the white Gaussian noise; $I_o(x)$ is the modified Bessel function of the zero order; $\Phi(x)$ is the probability integral; U_B, φ_B is the amplitude and phase of the upper component of the FM-AM spectrum; v_1, v_2 are the coeffecients determined by the XO circuit.

Mathematical expectations of the processes (9), (10) are found as[13]

$$m_{Ai} = \sqrt{m^2_{Ao} + 2m_{Ao}K_i\cos(\varphi_A - \varphi_i) + K^2_i} \quad ; \quad (11)$$

$$\varphi_{Ai} = \operatorname{arctg}[(m_{Ao}\sin\varphi_{Ao} + K_i\sin\varphi_i)/$$

$$/(m_{Ao}\cos\varphi_{Ao} + K_i\cos\varphi_i)] \quad ; \quad (12)$$

$$m_{\omega i} = \sqrt{m^2_{\omega o} - 2m_{\omega o}K_i\gamma\cos(\varphi_{\omega o} - \varphi_i) + K^2_i\gamma^2} \quad ; (13)$$

$$\varphi_{\omega i} = \operatorname{arctg}[(m_{\psi 0}\sin\varphi_{\omega 0} - K_i\sin\varphi_i)/$$

$$/(m_{\varphi o}\cos\varphi_{\omega o} - K_i\cos\varphi_i)] \quad ; \quad (14)$$

where $\gamma = \Omega/\omega_o$; $i \equiv npq$; $m_{\varphi o} = m_{\omega o}/\gamma$; $m_{\omega o}$, $\varphi_{\omega o}$, m_{A0}, φ_{A0} are the DM characteristics parameters with $F \neq F_{npq}$ determined from (5)-(8) with $F \gg \Pi^H_{n11}/2\pi$, where Π^H_{n11} is the passband of the loaded of the tuned-circuit of the h_{n11} mode; K_i, φ_i are the coefficient and the phase determined by the XO circuit[14]. The analysis of the phase DM characteristic dispersions (12),(14) of the processes (9-10)

$$\sigma^2_{\varphi A,\omega} = \pi^2/3 + 4\pi\sum_{l=1}^{\infty} l^{-2}(-1)^l c_{l_A, l\omega} \quad ; \quad (15)$$

where $c_{l_A, l\omega} = \left(\frac{M_{A,\varphi}}{\sigma}\right)^l \Gamma\left(1 + \frac{l}{2}\right) {}_1F_1\left(\frac{l}{2}, l+1; -M^2_{A,\varphi}/4\sigma^2\right) /$ $/\left(n!\,2^{l/2}\right)$; $\Gamma(1+l/2)$ is the gamma-function; ${}_1F_1(\alpha,\beta;x)$ is the degenerated hypergeometric function, showing its ratter wear fixing capabilities with $m_{F,\varphi}/\sigma = 1$, increasing with the increase of $m_{A,\varphi}/\sigma$ according to the Tikhonov law[19]. In practice the dispersions (15) of DM characteristic (11)-(14) can be neglected with $m_{A,\varphi}/\sigma$ exceeding 20 dB.

The next important condition influencing the reference properties of the DM characteristic is the choice of the regeneration factor of the self-excited system of XO μ_1. Its increase according to[13] hand the self-excited system to the state when all the necessary conditions of the oscillation at the frequency (1) of the h_{npq} mode are satisfied which does't had to a multifrequency mode due to the faihere to meet sufficient conditions of the oscillations-stability. With this XO state the losses in the unharmonic tuned are compensated by the circuit, its equivalent Q-factor and the DM characteristic slopes increase significantly which results in the spikes in the AF modulation characteristics of FM and AM and the slopes rise in their phase frequency modulation characteristics with the change in the behavior[13]. And the restriction for such variations is imposed only by nonlinear effects in the self-excited system.

As a result in the space of variations μ_1 and Ω a particularly pronounced resonance[13] is observed, let us call it the regenerating one. Such a phenomenon is of significant practical value, it enables

to optimize stabilization systems proceeding from the Q-factor of the equivalent tuned-circuite of the h_{npq} mode and the sensitivity of the f_{npq} frequency to the vector $\vec{\theta}$ changes. And it is necessary to choose the excitation level of the fundamental oscillation in an optimum way in order to avoid the influence of nonlinear effects either with large[27] or with small[28,29] amplitudes of the piezocurrent.

To carry out a comparative assessment of the stabilizing factors of the DM characteristics let us receive the assessment of the ΔK_A passbands and the slopes $\partial\varphi_A/\partial F$ of the amplitude DM characteristics. In particular, for the Colpitts oscillator with the varicap capacity modulation in the spectral line band these assessments according to[14,19-23] can be found in the form of

$$\Delta K_A\begin{Bmatrix} a > b \\ a < b \end{Bmatrix} = \frac{2\left\{ \begin{matrix} r_2(\mu_1-1)[2+(\mu_1-2)^2/9] \\ (2r_2+1,5)(\mu_1-1)^2 \end{matrix} \right\}}{r_2[2+(\mu_1-1)^2/9]-(2r_2+1,5)(\mu_1-1)} \times$$

$$\times \left[\begin{Bmatrix} 1 \\ \nu \end{Bmatrix}\mathbb{K}(k)-\mathbb{E}(k)\right] \quad ; \qquad (16)$$

$$\left.\frac{\partial\varphi_A}{\partial F}\right|_{F=0} = \frac{r_2[2+(\mu_1-1)^2/9]-(\mu_1-1)(2r_2+1,5)}{r_2(\mu_1-1)[2+(\mu_1-1)^2/9]} \quad ,(17)$$

where $a^2=r_2^2[2+(\mu_1-1)^2/9]^2/(2r_2+1,5)^2$; $b^2=(\mu_1-1)^2$; $\nu = a^2/b^2$; $r_2 = C_o C_{Bo}^{-1}[1+C_o(C_1^{-1}+C_2^{-1}+C_{Bo}^{-1})]$ is the capacity ratio of the circuit ;

$k = \begin{Bmatrix} (a^2-b^2)^{0,5}/a \\ (b^2-a^2)^{0,5}/b \end{Bmatrix} < 1$; $\mathbb{K}(k)$, $\mathbb{E}(k)$ are the full elleptical integrals of the 1st and 2nd kind; C_o, C_{Bo}, C_1, C_2 is the static capacity of the crystal resonator, initial capacity of the varicap and the phasing capacity of the XO circuit.

Analogous functions of the spikes parameters of DM characteristics at the frequencies of unharmonic resonances according to[13] can be written as

$$\Delta K_A = \left\{\int_{F_1}^{F_2}\left\{m_{Ao}^2 + 2m_{Ao}k_1(F)\cos[\varphi_{Ao}-\varphi_1(F)]+k_1^2(F)\right\}^{0,5}dF - m_{Ao}\Delta F\right\}/\left|m_{AM}-m_{Ao}\right| \quad ; \qquad (18)$$

$$\left.\frac{\partial\varphi_A}{\partial F}\right|_{F=F_{npq}} = \frac{\partial}{\partial F}\left[\frac{m_{Ao}\sin\varphi_{Ao}+k_1(F)\sin\varphi_1(F)}{m_{Ao}\cos\varphi_{Ao}+k_1(F)\cos\varphi_1(F)}\right]\Bigg|_{F=F_{npq}}, \qquad (19)$$

where $\varphi_1(F) = \operatorname{arctg}\{[(\alpha_1-1)\sin\varphi_\Omega - \alpha_2\cos\varphi_\Omega]/[(\alpha_1-1)\cos\varphi_\omega+\alpha_2\sin\varphi_\Omega]\}$; $k_1(F)=[(\alpha_1-1)^2+\alpha_2^2]^{0,5}U_B/U_0$; m_{AM} is the maximum value of the DM characteristic at the spike frequency; $\Delta F = 3\Pi_{npq}^H/2\pi$; $F_1 = F_{npq} - 1,5\Pi_{npq}^H/2\pi$; $F_2 = F_{npq} + 1,5\Pi_{npq}^H/2\pi$; Π_{npq}^H is the passband of the loaded of the tuned-circuit of the h_{npq} mode; $\alpha_1=\alpha_1(F)$; $\alpha_2=\alpha_2(F)$ are the coefficients determined by the XO circuit[13].

Generalized functions of the DM characteristics parameters (16) - (19) of the Colpitts circuit with the varicap capacity modulation are shown in Fig.4,5. During the linear mode of operation with $\mu_1 = 1.0$ the losses in the mode h_{n11} tuned-circuit are compensated by the circuit, the DM characteristics band $\Delta K_{An11}/\Pi_{n11}^H$ is approaching zero; and the slope of the phase DM characteristic $\partial\varphi_{An11}/\partial F$ tends to infinity. With the same value of μ_1 the losses introduced by the circuit into the fundamental tuned-circuit of the h_{n11} mode are increasing and into the ungarmonic one are decreasing. The equality of the DM characteristic parameters of the two tuneds corresponds to the value $\mu_1 = \mu_{1x}$. Further increase of μ_1 results in the increase of $\partial\varphi_{Anpq}/\partial F$ for $pq \neq 11$ up to the value of μ_{1npq} with which the losses in unharmonic tuned are fully compensated by the circuit; the extremum value of the amplitude frequency modulation characteristics of AM $\Delta K_{Aext.npq}$ and $\partial\varphi_{Anpq}/\partial F$ is increasing significantly. The band of the hold-in ΔF_{max} of the phase DM characteristic in this region μ_1 is decreasing up to a very small quantity and with the further increase of μ_1 tends to infinity stepwise. The variation range of $\Delta\varphi_{Anpq}$ is increasing and with $\mu_1 > \mu_{1npq}$ remains equal to 2π.

While inplementing the modulational method it is necessary to choose μ_1 in the range of $\mu_{1kp} < \mu_1 < \mu_{1npq}$, which is easily inplemented in practice as with precision XO the lower boundary of μ_1 for the nearest odd unharmonic resonances is in the region of $1,02 < \mu_1 < 1,1$.

PRINCIPLES OF SELF-CONTAINED FREQUENCY STABILIZATION

Analyzing the fundamental signals and relationships with FM-AM and taking into account that the DM characteristic parameters are in the functional dependence on the $\vec{\theta}$ vector we can formulate the basic principles of constructing quartz-crystal stabilization frequency systems which can be realized on basis of the modulational method.

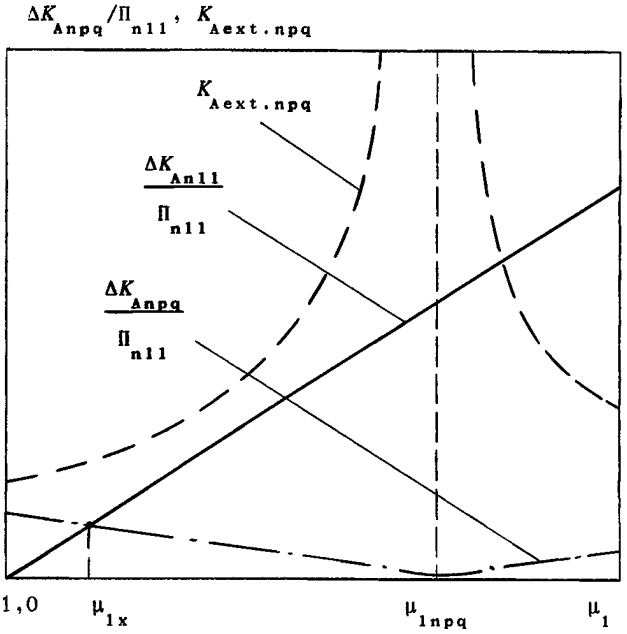


Fig. 4. The dependence of the XO DM characteristics ΔK_{Anpq}, $K_{Aext.npq}$ on the regeneration factor μ_1.

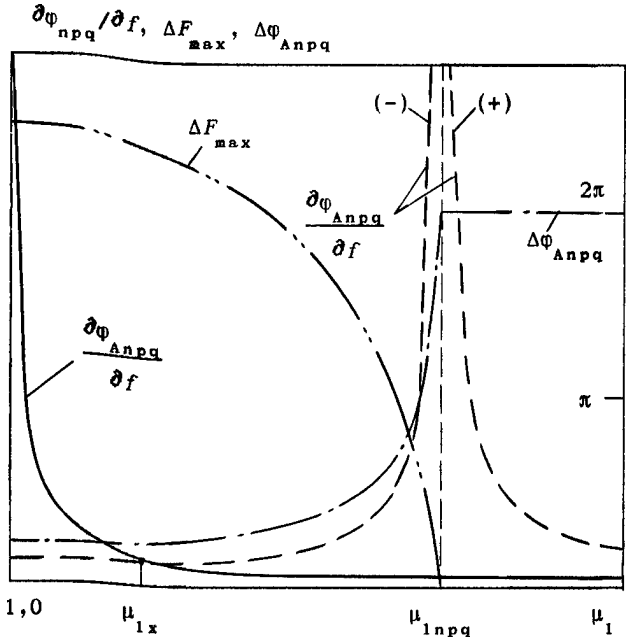


Fig. 5. The dependence of the XO DM characteristics ΔF_{max}, $\Delta\varphi_{Anpq}/\partial f$, $\Delta\varphi_{Anpq}$ on the regeneration factor μ_1.

The first principle is the nonmonofrequency of the crystal resonator. It means that the basis of the system construction is served by the phenomenological unharmonics properties of the quartz resonator.

The necessary conditions of the system realization on the basis of the first principle are as follows:

- in the stationary naturally fluctuating environmental field and with the chosen parameters of the XO the fixing capabilities of the phase DM characteristics of the h_{npq} modes must be higher than those of the h_{n11} mode tuned; it means that

$$\left.\frac{\partial\varphi_{A,\omega}}{\partial f}\right|_{f=f_{n11}} < \left.\frac{\partial\varphi_{A,\omega}}{\partial f}\right|_{f=f_{npq}} ; \quad Q^{n}_{n11} < Q^{n}_{npq} ;$$

$$\left.\Delta K_{A,\omega}\right|_{f=f_{n11}} > \left.\Delta K_{A,\omega}\right|_{f=f_{npq}} ; \quad \delta f^{AV}_{n11} > \delta f^{AV}_{npq} , \tag{20}$$

where Q_{n11}, Q_{npq} are the loaded Q-factors of the fundamental and unharmonic tuned-circuits in the XO circuit; δf^{AV}_{n11}, δf^{AV}_{npq} are the Allan variances of the forced oscillations of the h_{n11} and h_{npq} modes in the XO circuit.

- introduction of the modulating signal in the XO circuit must not change the main XO parameters significantly.

The sufficient conditions are as follows:

- availability of the measuring mechanism of unharmonic tuneds parameters, in the capacity of which the modulational method is used, and with the help of it the above-mentioned parameters are converted into the variations of the DM characteristic of the output signal.

The second principle is the parametric stability of the unharmonic mode or a combined oscillation of the crystal resonator in the nonstationary fluctuating surroundings. It means that the system can be constructed if there is at least one unharmonic mode or a combined oscillation of arbitrary modes in the quartz unit spectrum, which has reference properties in relation to the fundamental frequency f_o with the change of the vector $\vec{\theta}$ parameters.

The third principle is the parametric sensitivity of the unharmonic mode or a combined oscillation of the crystal resonator in the nonstationary fluctuating surroundings. It means that the system can be constructed if the sensitivity of at least one unharmonic mode or of a combined oscillation of arbitrary modes due to the change in the vector $\vec{\theta}$ parameters is greater than that of the fundamental oscillation.

In essence the second and the third principles mean that when constructing the system it is necessary to choose ungarmonic resonances with the frequencies sensitivities to the vector $\vec{\theta}$ parameters variation in view of (20) which are different from the fundamental oscillation. In the opposite case the effeciency of the system is an isolated case.

GENERALIZED MODEL OF STABILIZATION SYSTEM

The generalized vector model of the quartz-crystal frequency stabilization system in which the natural nonuniformity of the XO DM characteristics is realized on the basis of the modulational method [18] and the above-formulated principles is shown in Fig.6, where AR1, AR2 are the adapted regulators; AD is the adapting device; Output elect. is the band-pass crystal filter.

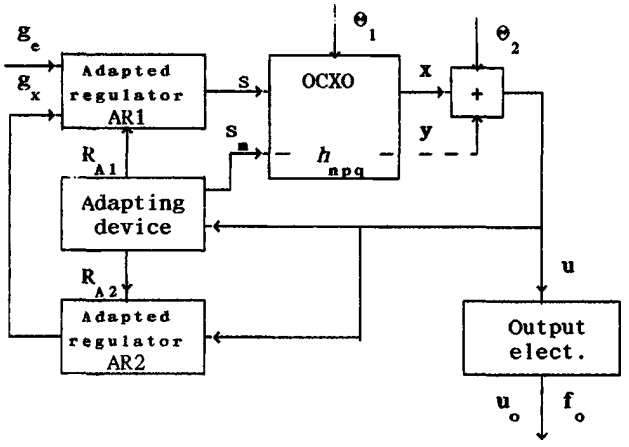


Fig. 6. The generalized model of the quartz stabilization system realizing the natural nonuniformity of the XO DM characteristic on the basis of the modulational method .

The system functions in the following way. The adapting device carries out FM-AM of the XO by the signal $\vec{s}_M$, carrying the information about the behavior of the unharmonic modes $\vec{h}_{npq}$, pq≠11 over to the DM characteristic of the XO signal $\vec{x}$ through the connection conditionally designated as $\vec{y}$, derives the information about the modulation parameters from the signal $\vec{u}$ and uses it to sinchronize the $\vec{s}_M$ signal and to form the measuring and control signals $\vec{R}_{A1}$ and $\vec{R}_{A2}$. The influence of the $\vec{\theta}$ vector on the $\vec{u}$ signal instability is deminished by the main "coarse" regulation tuned which includes AR1, XO and AR2. With this the control is carried out by the $\vec{s}$ signal, formed by the reference and measuring signals $\vec{g}_e$ and $\vec{g}_x$. The residual instability of the $\vec{u}$ signal is diminished by the $\vec{R}_{A1}$ and $\vec{R}_{A2}$ signals entering the control inputs of the regulators AR1 and AR2 . By this way the "coarse" stabilization system is adapted to the medium of operation. With this the main objective of the adaptation in the limits of invariance consists in increasing the stability of the XO fundamental frequency f_o . To filter the FM-AM products on the XO output the pass-band filter is connected in the channel of the latter.

In the point of the operating condition of the XO regeneration (in the residual instability region) the system can be linearized, and the equations of its dynamic state with, for example, passive adaptation written with respect to the signal increments of the main "coarse" tuned ($^{\mathcal{E}}$cymbol) and the tuned of the reference model - tuned of the adapting device (^{M}symbol) can be written in the matrix form

$$U^{\mathcal{E}} = (I + K^{\mathcal{E}})^{-1} K^{\mathcal{E}}_{\kappa p1} K^{\mathcal{E}}_{p1} G_e + (I + K^{\mathcal{E}})^{-1}\Theta \quad ; \quad (21)$$

$$U^{M} = (I + K^{M})^{-1} K^{M}_{\kappa p1} K^{M}_{p1} G_e + (I + K^{M})^{-1}\Theta \quad ; \quad (22)$$

where $K^{\mathcal{E}} = K^{\mathcal{E}}_{\kappa p1} K^{\mathcal{E}}_{p1} K^{\mathcal{E}}_{p2}$; $K^{M} = K^{M}_{\kappa p1} (K^{M}_{p2} K^{M}_{A2} + K^{M}_{A1})$; $U^{\mathcal{E},M}(p) = [u_1^{\mathcal{E},M}(p)\ u_2^{\mathcal{E},M}(p) \dots u_n^{\mathcal{E},M}(p)]^T$; $G_e(p) = [g_{e1}(p)\ g_{e2}(p) \dots g_{en}(p)]^T$; $\Theta(p) = [\theta_1(p)\ \theta_2(p) \dots \theta_n(p)]^T$ are the matrix columns of the vectors increments of the system output signal, control /signal and a random action; $K^{\mathcal{E},M}_{\kappa}(p) = [K^{\mathcal{E},M}_{\kappa ij}(p)]$, $i,j \in \overline{1,n}$ is the square matrix of XO controllability coefficients; $K^{\mathcal{E},M}_{p1}(p)$, $K^{\mathcal{E},M}_{p2}(p)$, $K^{M}_{A1}(p)$, $K^{M}_{A2}(p)$ are the diagonal matrices of the transfer coefficients P1 through the $\vec{R}_{A2}$ input, P2 through the outputs $\vec{R}_{A2}$ input, adapting device through the outputs $\vec{R}_{A1}$ and $\vec{R}_{A2}$; I is the identity matrix; $p = j\omega$.

The corresponding (21) and (22) equation of the adaptation statics, which has the form (23)

$$K^{\mathcal{E}}_{\kappa p1} K^{\mathcal{E}}_{p1} K^{\mathcal{E}}_{p2} = K^{M}_{\kappa p1} K^{M}_{p1} (K^{M}_{p2} K^{M}_{A2} + K^{M}_{A1}) \quad , \quad (23)$$

enables us to find the parameters of the "coarse" and adapted regulation circuits with which the efficiency of the fundamental frequency stabilization is maximum. In particular, with the zero $\vec{R}_{A2}$ vector $K^{\mathcal{E}}_{p1} = K^{M}_{p2}$ is provided after that the adapting device parameters are specified with which the identity $K^{\mathcal{E}}_{k} K^{\mathcal{E}}_{p2} = K^{M}_{k} K^{M}_{1}$ is satisfied and that similarly to (23) enables to achieve the required stabilization quality.

In the specific case the state of the system is possible in which the transmission channel of the fundamental signal $u^{\mathcal{E}}$ is open, i.e. $K_{p2} = K_{A2} = 0$. With this the system is organized on the basis of the combined connection of the tuned-circuit of the h_{s11} mode and the adapting device circuit and its equation of the instability movement has the form

$$U^{\mathcal{E}} = (I + K)^{-1} K^{\mathcal{E}}_{\kappa p1} K^{M}_{p1} G_e + (I + K)^{-1}\Theta \quad \text{where} \quad K = K^{\mathcal{E}}_{\kappa p1} K^{M}_{A1}$$

is the square matrix of the size $n \times n$. Here the signal $\vec{u}$ stabilization is reached

similarly to the tuned-circuit of the h_{511} mode (21) by increasing the corresponding components of the matrix K .

The second specific case is characterized by the fact that the closed system cannot be constructed in principle, because $K_{p2} = K_{A2} = 0$ and it is impossible to realize the operation of the instabilities measuring of the fundamental frequency f_o and the signal $\vec{u}$ with the vector $\vec{\theta}$ variation. Here the single possibility to achieve the effect consists in constructing the compensation system on the basis of P1 , XO and adapting device circuit combination. With this the equations of the signal movement in the main circuit and in the adapting device have the form

$$U^{\beta} = K^{\beta}_{k} K^{\beta}_{p1} G_{e} + \theta \quad ; \qquad (24)$$

$$K^{M}_{A1} = K^{M}_{A1} U^{M} \quad ; \qquad (25)$$

and the action of the K^{M}_{A1} signal is directed towards compensation, as a rule, by the regular component $\bar{\theta}$ of the destabilizing effect $\vec{\theta}$. Setting the matrix U^{β} in (24) equal to zero, we change over to the equation of the complete compensation of the matrix $\vec{\theta}$ components influence on the $\vec{u}$ signal

$$K^{\beta}_{k} K^{\beta}_{p1} G_{e} + \bar{\theta} - K^{M}_{A1} U^{M} = 0 \quad ; \qquad (26)$$

which can be solved relative to the matrix K^{M}_{A1} components with which the f_o frequency becomes invariant to the vector $\vec{\theta}$ $\forall\, \tau_{\theta_i} > \tau_j$, $i,j \in \overline{1,n}$, where τ_{θ_i} is the correlation time of the most quickly changing i component of the vector $\vec{\theta}$; τ_j is the time constant of the j channel of the system

PRACTICAL RESULTS

Let us present the most general results of the practical use of the modulational method to reduce the influence of temperature and aging on the frequency of precision quartz-crystal oscillators.

OCXO with the adaptation system of the temperature stabilization

The base circuit of the crystal oscillator is realized on the basis of the AT-cut crystal resonator, PK-187 (Phonon) by the Colpitts circuit with the proportional single-ring oven control system (Fig.7). With the stabilization temperature T^{o}_{o} = =65°C of the quartz unit and without the use of the modulation circuit had the temperature instability $\Delta f/f$ reaching 5×10^{-9} in temperature range of 20... 55°C (Fig.7) and the average slope of the temperature-frequency characteristic 1.4×10^{-10}/°C (Fig.8).

For the passive adaptation of the temperature stabilization circuit the mode h_{513} frequency was

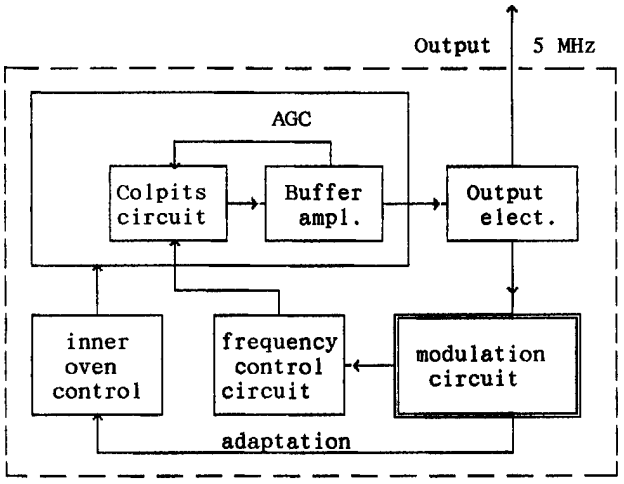


Fig. 7. The OCXO structure with an ambient temperature adaptation circuit .

used which was 160 kHz higher than the fundumental frequency of the h_{511} mode (5 MHz) and had the slope of the temperature-frequency characteristic in the T^{o}_{o} point of approximately 2×10^{-7}/°C .

The modulation circuit was realized on the basis of the formulated construction principles of the quartz-crystal frequency stabilization systems.

With this the instability movement in the circuit was described by a scalar equation of the type (22).

In the result of the adaptation of the temperature stabilization system to the ambient temperature via connection of the modulation circuit the frequency instability in the temperature range (20...55°C) reduced to

$$1 \times 10^{-10}$$

which corresponds to the average slope 2.8×10^{-12}/ /°C. With the temperature T^{o}_{o} = 63°C the instability was 5×10^{-9} ; a sharp instability rise was observed with the approach to the point T^{o}_{o} (Fig.8).

Quartz frequency standard "Ч1-88" with long-term aging compensation

The structure of the base part of a quartz frequency standard consists of OCXO and a quartz filter with the addition of peripheral units. The quartz frequency standard "Ч1-88" is realized by the Colpitts circuit on the AT-cut quartz resonator of the PK-187 or PK-327 (Phonon) type. The OCXO has a single-ring proportional oven control system.

To assess the long-term instability of the fundamental frequency 5 MHz according to the above method the modulation circuit and the computing unit have been introduced into the circuit (Fig.9).

At the standard calibration stage the aging functions of the fundamental frequency 5 MHz and the unharmonic resonance frequency h_{513} which is 160 kHz higher than the fundamental frequency are determined. The necessary constants and functions

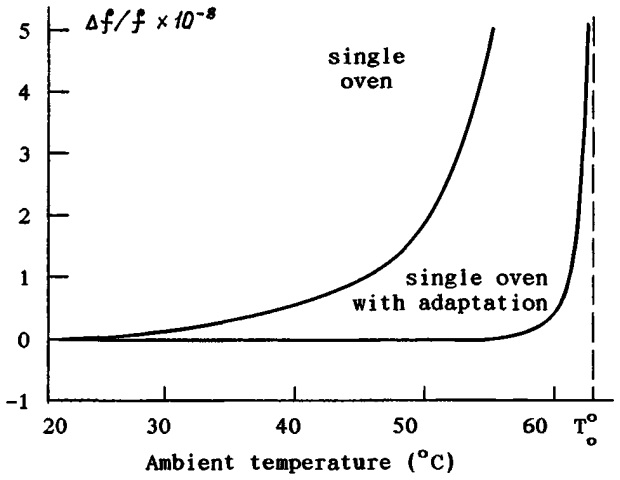


Fig. 8. The temperature characteristics of the base OCXO (single oven) and base OCXO with the adaptation to the ambient temperature .

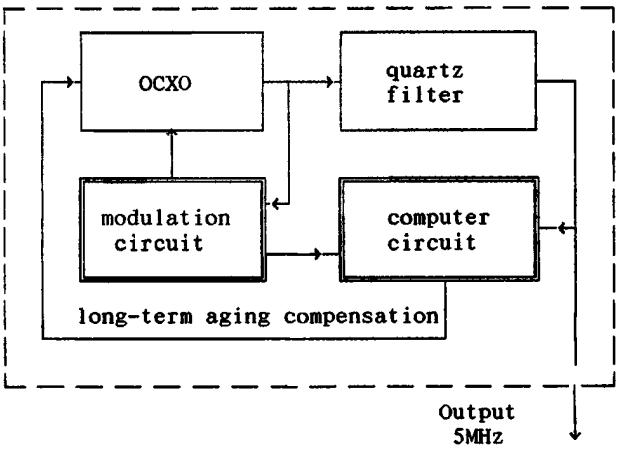


Fig. 9. The quartz-frequency standard structure with the long-term aging compensation.

used for the calculation of the compensation signal applied to the OCXO frequency control input are put down into computer memory.

During the standard operation process there is a possibility of the self-contained (without the use of quantum standards) frequency compensation both in the manual (by the aging indicator readings and with the use of the accurancy table) and in the independent mode of operation. With the axamined standards samples the regular component of the long -term frequency instability after the compensation didn't exceed 1×10^{-10} within a year (the same estimation is forecast for the whole period of operation. The 24-hourly Allan variance including the compensation error didn't exceed 5×10^{-10} and was connected with the dynamic temperature influence[26] .

The slope of the short-term frequency instabilities of the "Ч1-88", N02, PK-327 quartz-frequency standard are shown in Fig.10. Without the

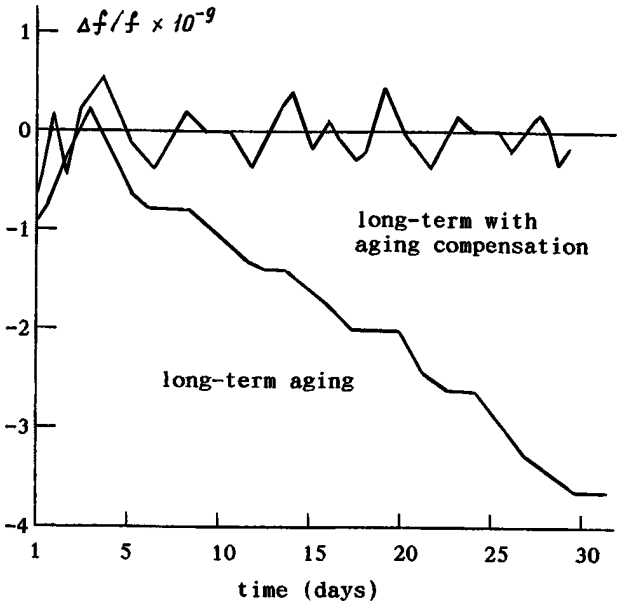


Fig. 10. The base standard aging characteristic (long-term aging) and the base standard long-term characteristic with aging compensation

use of the compensation circuit the standard has the 24-hourly Allan variance 1.5×10^{-10} , the 24-hourly aging speed 1.1×10^{-10} and the yearly frequency drift 4.2×10^{-8}.

With the use of the compensation circuit the frequency drift has been practically eliminated and didn't exceed

$$9\times10^{-11} \text{ within a year,}$$

and the 24-hourly Allan variance including the compensation error didn't exceed 2.8×10^{-10} . The Allan variance for the averaging periods from 100 μs to 100 s remains unchanged and in 1s is 2.8×10^{-12} for the N02 standard.

The results of long-term aging compensation in OCXO are presented in Figure 11.

CONCLUSION

Thus, the proposed combined trend of the self-contained frequency stabilization of precision XO substantiated theoretically and tested experimentally enables us, with its development, to reduce significantly the regular and slow random operational frequency instabilities of XO for each of the destabilizing factors. With this there is a real possibility of the development and industrial production of OCXO and quartz frequency standards with the characteristics which are on a par with quantum rubidium frequency standards.

The use for this purpose of OCXO based on BVA and QAS quartz-resonators as basic sources is the most promising.

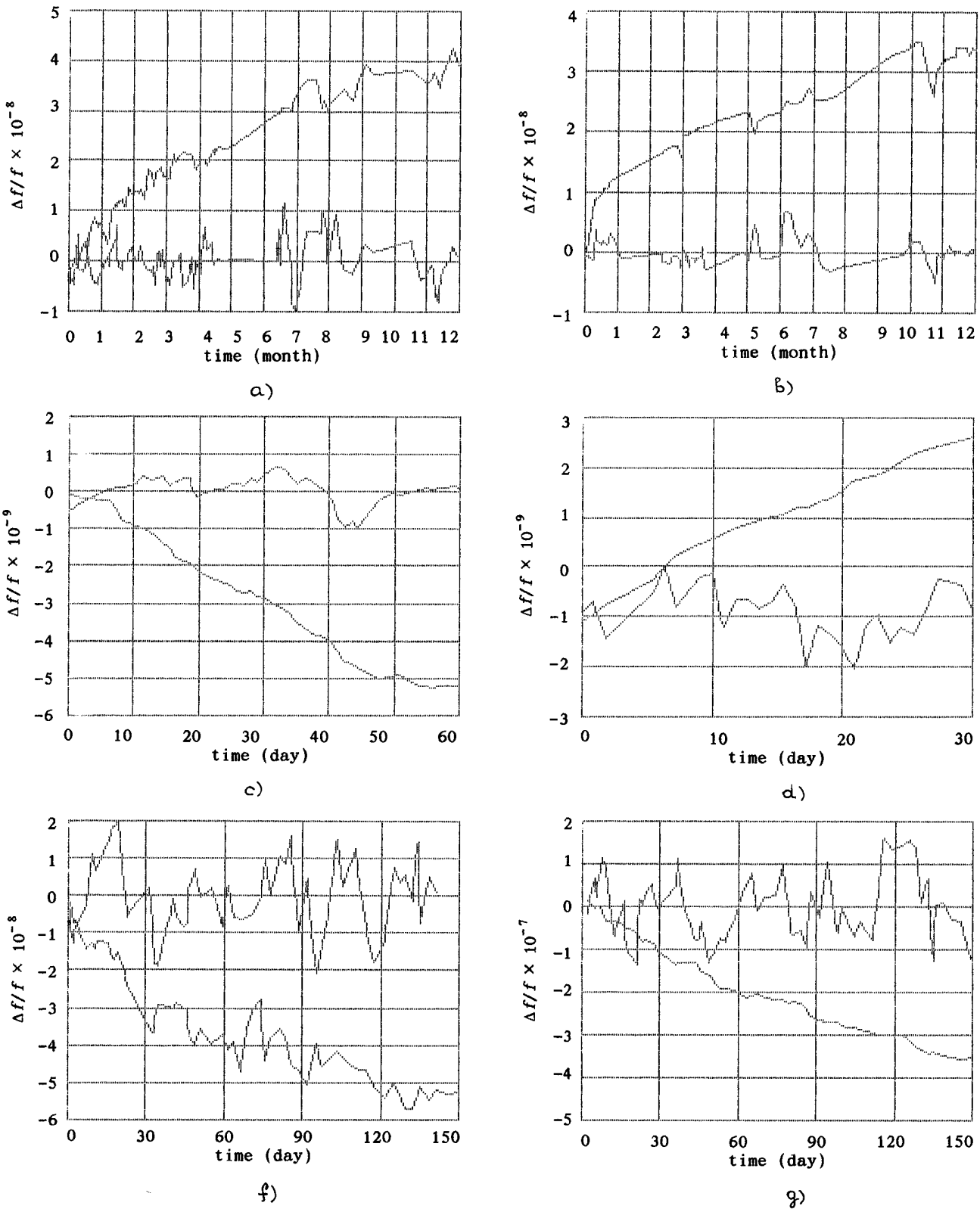


Fig. 11. The long-term characteristics of OCXO with and without aging compensation

ACKNOWLEDGMENT

The author wish to thank A.P.Kovel, S.A.Trushliakov, V.N.Romanko, G.V.Megeny, Yu.I.Yevdokimenko of Kharkov Military University; staff of "SICHRON" Centre, espesialy to thank A.Ph.Kurotshka, E.G.Sokolinsky, O.E.Rudnev; staff of S.Co.IRVA, espesialy to A.V.Marienko, A.V.Shemeko for experimental study of the modulational method, production and measuremens of OCXOs and frequency quartz standards.

REFERENCES

1. P.E.Kandyba,P.G.Pozdniakov, etc. Piezoelectrical resonators:Reference book,M.,Radio i sviaz,1992.
2. V.J.Rosati, R.L.Filler, S.S.Schodowski, J.R.Vig. State of the art in crystal oscillators, present and future. Proc.of 37th AFCS, p.386-390. 1983.
3. G.Caret, E.Gerard, G. Marotel. Oscillateur a quartz compense en temperature numerique (PSTN). Revue Technique Thomson-CSF, Vol.18, N2, p.301--324, 1986.
4. R.Besson. Les oscillateurs a quartz. L'Onde Electrique, V.66, N4-5, p.77-83, 1986.
5. R. L. Filler. The acceleration Sensitivity of Quartz Crystal Oscilators: A Review IEEE Trans. Ultrason. Ferroelec. Freq. Contr., Vol.35, N3, p.297-305, 1988.
6. J.P.Aubry, G.Caret, G.Marotel. Oscillateurs a quartz de faible sensibilite a l'environmement pour les applications Temps-Frequence. Bulletin BNM, N63-64, p.26-40, 1986.
7. G.Marotel, G.Caret, J.P.Aubry. An oscillator for the GPS program. Proc. of 39th AFCS, p.212-- 222, 1985.
8. A.Debaisieux, J.P.Anbry. E.Gerard, M.Brunet. A satellite oscillator for very precise orbitography. Proc. of 39th AFCS. p.202-211, 1985.
9. J. Ho. Hybrid miniature oven quartz crystal oscillator. Proc.of 39th AFCS, p.193-196, 1985.
10. J.-P.Valentin.Reduction de la derive de frequence des oscillateurs a quartz de haute stabilite. Revue Phy Appl., Vol 19, p.307-310, 1984.
11. D.L.Hessick, W.C.Euler. GPS user receivers and oscillators. Proc. of 38th AFCS, p.341-362,1984.
12. V.N.Romanko,Yu.S.Shmaly. Probability properties of the dynamic modulational characteristics of the quartz-crystal oscillators. Izvestiya VUZ. Radioelektronika, V.35, N.11, p.37-44, 1992.
13. Yu.S.Shmaly. The dynamic modulational characteristics splashes in a precision FM quartz-crystal oscillators. Radiotechnika, N.7, p.36--39, 1991.
14. Yu.S.Shmaly. Dynamic distortion in a FM auto-oscillatory systems. Izvestiya VUZ.Radioelektronika, V.29, N12, p.40-44, 1986.
15. H.F.Tiersten, R.S.Smythe. An analysis of contoured crystal resonators operating in overtones of coupled thickness shear and thickness twist. J.Acoust.Coc.Am., V.65, p.1455, 1979.
16. D.S.Stevens,H.F.Tiersten. An analysis of doubly-rotated contoured quartz crystal resonators. Proc. of 39th AFCS, p. 436-447, 1985.
17. Yu.S.Shmaly., Yu.I.Yevdokimenko. Modulational method of adaptation of the OCXO. M., VIMI,Ch.1, p.75-78, 1989.
18. Yu.S.Shmaly. The precision quartz - crystal oscillators frequency stabilization with the use of the natural unevenness of the dynamic modulational characteristics. Doctoral thesis (Doctor of the technical science). Kharkov, 1991.
19. V.I.Tichonov. Nonlinearity transformations of a random processes. M,.Radio i sviaz, 1986.
20. Yu.S.Shmaly. Estimation of a frequency-phase distortions in a FM oscillators. Radiotechnika, N5, p.30-33, 1988.
21. V.Ya.Bargin,Ph.Ph.Kolpakov,Yu.S.Shmaly. Dynamic characteristics of control of a quartz-crystal oscillator. Izvestiya VUZ.Radioelectronika,V.24, N8, p.67-72, 1981.
22. V.Ya.Bargin,Ph.Ph.Kolpakov,Yu.S.Shmaly. Frequency-phase distortion estimation in a control avto -oscillatory systems. Radiotechnika, V.37, N9, p.50-53, 1982.
23. Yu.S.Shmaly, V.Ya.Bargin. Decrease of influence of quartz-crystal unharmonics to distortions in FM oscillators. Radiotechnika, V.40, N1, p.46--48, 1985.
24. Yu.I.Yevdokimenko, Yu.S.Shmaly. Thermodynamical frequency instability in quartz-crystal oscillators. Izvestiya VUZ.Radioelectronika, N5, p.49--55, 1991.
25. Yu.I.Yevdokimenko, Yu.S.Shmaly. Thermodynamical frequency instability of volume acoustic vibrations of a quartz plate. Acustichesky Jurnal. V.40, N6, p.985-988, 1992.
26. Yu.I.Yevdokimenko, Yu.S.Smaly. A Thermodynamic resonance in piezoelectric crystal plates of thickness-shear vibrations. Proc. of 47th IFCS, p.193-201, 1993.
27. J.J.Gaqnepain. Evaluation non lineaire des elements motionnels du schema equivalent du quartz. C.R.Acad. Sc., Paris, Vol. 276, N7, Serie B, p.231-233, 1973.
28. S.A.Trushliakov, Yu.S.Shmaly. Influence of equivalent losses of quartz-crystal to excitation of oscillators. Radiotechnika, N7, p.41-44, 1989.
29. Yu.S.Shmaly. Probability characteristics of stationary acoustic vibrations in a quartz-crystal resonator under extralow excitation. Acustichesky Jurnal, V.40, N6, p.985-988, 1994.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

BVA resonators and oscillators : a review. Relation with space requirements and quartz material characterization.

Raymond J. BESSON, Jean-Jacques BOY, Marc M. MOUREY

LCEP / Ecole Nationale Supérieure de Mécanique et des Microtechniques
26, Chemin de l'Epitaphe - 25030 BESANÇON CEDEX - FRANCE

Abstract

BVA technologies for piezoelectric resonators have been introduced in the late 70's (1976, 1977, 1978). Since then, a lot of work dealing with BVA resonators has been done in our laboratory and also in several other laboratories including industry laboratories. BVA oscillators have been designed to match BVA resonators special properties. BVA resonators and oscillators are now commercially available from at least 3 companies. Due to their special structure, BVA resonators (monolithic and often "electrodeless") happen to be an efficient tool to obtain an interesting link between the properties of resonators and oscillators and the properties of piezoelectric material (usually quartz).

Under those conditions, a review devoted to BVA resonators and oscillators may be useful because there is a need of summarizing existing results and because some results have never been published yet. This is the case for some recent results dealing with space applications of BVA resonators and oscillators. This is also the case for connection of resonators and oscillators parameters with material's properties.

In this paper, after a fast introduction replacing BVA technologies in the general frame of resonator's technology evolution, various type of BVA resonators are presented (including tactical type) and BVA technologies are rapidly discussed with use of AT, SC and eventually BT cuts. Results obtained by various authors are presented.

In a second section, we explain why BVA oscillators must be specially designed so as to match resonator characteristics and obtain best results in $\sigma_y(\tau)$ and drift.

A special section is devoted to results obtained under hard environmental conditions (accelerations, vibrations, magnetic field, temperature and pressure changes, radiations, ...). Some results needed for space applications are presented.

Since, unfortunately, space applications are very demanding, piezoelectric material quality may be a serious limitation to satisfy certain space requirements. Then, a section is devoted to the link between properties of the material with parameters measured on resonators or oscillators. This includes possible contributions of material's defects and fabrication technologies to flicker noise level of oscillators.

As a conclusion, results are discussed in terms of future trends and evolution of quartz frequency standards.

I - Introduction

A piezoelectric resonator is a cavity for electroacoustic waves in a solid of given shape, configuration and dimensions prepared from high quality piezoelectric material under precise control of orientation shape and dimensions. Quartz resonators for frequency and time applications are built so that their frequency is as independant as possible from external parameters. On the contrary, dependance would lead to sensor applications.

Quartz oscillators are **indispensable secondary standards**. They are present in almost any time and frequency control system, their **cost is relatively low** and they appear as **highly dependable units. Ultra stable quartz resonators and oscillators** have been developed mainly for telecommunications space and military applications especially when excellent short term stability is needed.

Quartz oscillators have been used for **more than 70 years** after first unit was designed in **1920 by W. G. Cady**. Millions of quartz oscillators are built worldwide each year including a small quantity (a few thousands) exhibiting excellent performances eventually capable of competition with atomic rubidium standards. Even atomic clocks are in fact quartz oscillators enslaved to an atomic reference for long term drift and stability. Stabilities of

0-7803-2500-1/95 $4.00 © 1995 IEEE

quartz oscillators range between 10^{-5} and 10^{-13} (usually in the order of 10^{-8} or 10^{-9}) and **ultra stable** units exhibit stabilities ranging between **10^{-12} to some 10^{-14}** depending on frequency of the unit and integration time. Recently results down to a few 10^{-14} have been obtained. Frequencies may range between some kilohertz to some hundreds of megahertz but ultra stable units are rather 5 MHz or 10 MHz (eventually 100 MHz).

In fact, quartz oscillators simply correspond to a "special" case of **resonant circuit oscillators**, where "special" means that quartz resonators exhibit very high stable quality factors. Amplitude of resonator vibrations increases until non linearities reduce loop gain to 1. The obtained frequency corresponds to a total phase shift along the oscillating loop of 0 (or 2π). The **large crystal quality factor allows a large $\Delta\phi / \Delta f$** factor (where $\Delta\phi$ is the phase shift introduced by a frequency variation Δf). Thus, when a component in the loop causes a phase shift, phase oscillating conditions need an opposite phase shift $-\Delta\phi$ to appear in the loop and as a consequence a **very small frequency variation Δf**. Actually, resonators are the true heart of quartz oscillators.

Since the whole story is already a long one, a few milestones may be indicated :

1880 : Pierre and Jacques Curie - Discovery of piezoelectricity.
1893 : Lord Kelvin - Microscopic theory of piezoelectricity in quartz crystal.
1910 : Voigt - Lehrbuch der Kristall Physik.
1914-1918 : Langevin - Work on ultrasonic detection in water.
1918 : Nicholson (Bell Labs) - Patent on oscillating circuit with Rochelle salt crystal (US Paten 2212845, April 10 1918).
1920 : W.G. Cady - Patent on oscillating circuit with 3 vacuum tubes and quartz crystal resonator (in feed back circuit).
1921 : Pierce's first oscillator.
1926 : First crystal controlled radio station in New York city.
1939-1945 : 130 millions of crystal resonators and oscillators were manufactured during world war II.
1948 : Introduction of coated units by R.A. Sykes.
1952 : Introduction of energy trapping by A.W. Warner.
1975 : Doubly rotated SC cuts by E.P. Eernisse.
1975 : Ceramic flat pack for resonators by Wilcox, Snow, Hafner and Vig.
1976 : Electrodeless BVA resonator.
1980-1995 : • Important efforts for "better" piezoelectric material (quartz and/or others).
• Significant efforts to improve quartz oscillators behaviour under hard environmental conditions.

II - BVA technologies

In the general frame of piezoelectric resonator's evolution, BVA technologies represent an attempt to obtain, through new technologies (for instance ultrasonic, machining, chemical etching or ionic milling), **monolithic, highly reproducible structures** eventually using **"electrodeless"** techniques.

In BVA techniques, emphasis is given to a total reconsideration of resonator construction so as to fight established routines. BVA resonators (BVA stands for the French "inside a box with improved aging") were introduced after 1975 in BESANÇON, France [1] [2]. They are now well known [3], industrially produced and commercially available, though many patented models are not yet in production line.

The new BVA_n structures may use a rather **conventionnal bonding** and a special fixation **(n odd)** or may use "automounted" crystals, i.e. crystal with **monolithic structure** and fixations made out of quartz **(n even)**. The crystals are usually "electrodeless" but electrodes can also be deposited on vibrating crystal (patent n° 77 17 309) so leading to BVA_4 (sometimes said QAS) designs.
The denomination indicates two successive steps (n odd and n even), in the attempt to reduce the crystal's noise and frequency drift contribution in complete oscillator. Since the new structures have already been described elsewhere [1] [2] [3], we will only summarize here the most important features and results related to BVA_2 type resonators. The BVA_2 type resonator basically corresponds to the following :

An ***"electrodeless design"***. All problems that relate to electrode deposition, such as damping, stresses, contamination, and ion migration disappear.

*A **crystal mounting made of quartz***. Small "bridges" connect the vibrating part of the crystal to the dormant part, with the following key advantages :
(1) negligible discontinuities or stresses in the mounting points;
(2) very high precision in the shape and location of the bridges;
(3) symmetry and reproducibility when needed;
(4) versatility in the design of bridges, according to specific goals.

Additional parameters. The BVA_2 design exhibits the following additional construction parameters, as compared to classical designs :
(1) the electrode (and thus the electric field) can have a radius of curvature different from that given to the vibrating crystal;

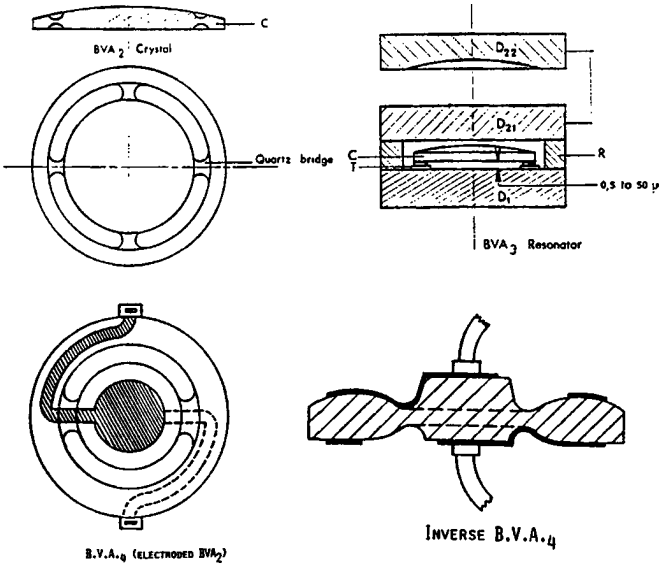
(2) heaters and sensors can be placed in vacuum close to the crystal without contacting the vibrating crystal;
(3) connecting bridges can have a great variety of shapes, locations, and other features.

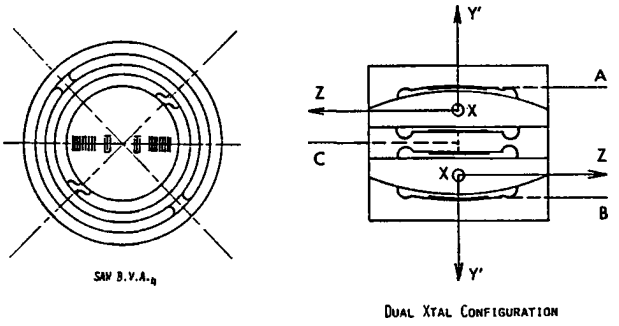
Provision for any material, crystal, cut, or frequency. Construction of the BVA_2 also allows very high frequencies operation and may use a quartz reentrant cavity to obtain fast warm up [4].

Reproducibility and versatility. The use of technological means (e.g., ultrasonic machining, ionic milling, chemical etching) allows both reproducibility and versatility. For example, the external shape of the crystal does not need to be circular or rectangular.
Very roughly speaking, the BVA design is capable of **an order of magnitude improvement** over conventional designs, in short-term stability, long-term drift, sensitivity to accelerations, vibrations, shocks, magnetic field and radiations.

At that point, best results have been obtained with "electrodeless" design usually combined with monolithic structures. Miniaturization may also be important for fundamental reasons [5] which means, in particular, that use of 10 MHz crystals for ultra stable oscillators now competes with use of 5 MHz crystals. In addition, when ultra stable crystal oscillators are used in an atomic passive clock configuration it may be useful to choose a higher frequency to decrease the influence of crystal oscillator noise on atomic interrogation [6].

Figure below gives an outline of some BVA configuration: the well known BVA_2 and BVA_4 designs and some others including BVA_3, inverse BVA_4, dual BVA_2 configuration and SAW BVA_4.





Let us recall that BVA_2 structure in a quartz reentrant cavity is a very useful solution for extremely fast warm up [7].

III - Tactical BVA resonators

Tactical BVA resonators have been progressively developped [8] [9] [10] [11] [12] [13] using a special symmetrical mounting according to Fig. 2.

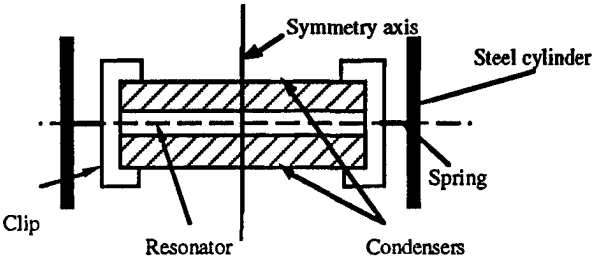


Fig. 2

The two condensers and the crystal are strongly clipped together and the clips are mounted in a groove inside a steel cylinder exactly in its middle by use of four springs. Mechanical parameters of clips and springs are carefully determined and checked. Usually the crystal is miniaturized (bva resonators). bva_2 10 MHz SC cut third overtone crystals have been space evaluated in a program under guidance of CNES and DGA (Centre National d'Etudes Spatiales et Délégation Générale pour l'Armement). Results are very promising for further space qualification.

IV - Typical parameters of a few BVA resonators

At that point it is probably usefull to give typical parameters of a few BVA resonators which have already been produced so far (see Table 1):

Frequency	Overtone P_n	Cut	Q 10^6	R_1 Ω	L_1 H
5 MHz	P_5	AT	2.6	70	5.8
5 MHz	P_3	SC	2.6	55	4.6
5 MHz	P_3	SC	2.6	48	3.97
10 MHz	P_5	AT	1.3	72	1.49
10 MHz	P_3	SC	1.36	100	2.16
50 MHz	P_5	BT	0.7	92	0.205
100 MHz	P_5	AT	0.11	65	0.011

Table 1 : Typical parameters of a few BVA resonators

These results represent **usual average results** obtained by us and by various authors in industry. Sometimes much better values are obtained (occasionnally 5 MHz units AT or SC exhibit a Q factor close to 2.9×10^6 or 10 MHz units AT or SC a Q factor close to 1.45×10^6) but it would be misleading to suppose they come out with a reasonable yield.

V - Design of oscillators and BVA oscillators

Metrological oscillators are usually of the Pierce, Colpitts or Clapp type more or less modified according to the application. Those three types of oscillators correspond in fact to the circuit of Fig. 3 :

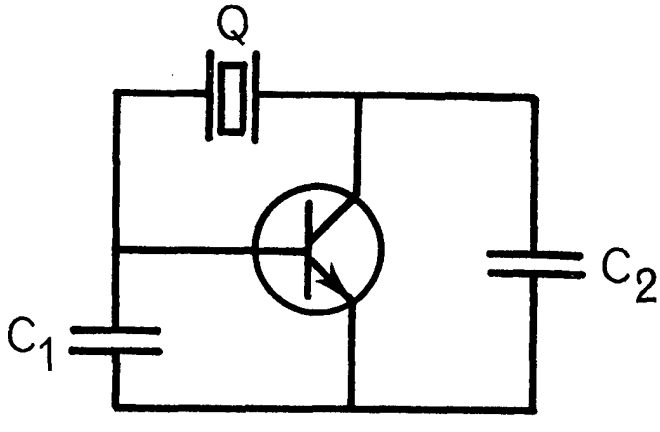


Fig. 3

DC voltages are not represented on Fig. 3 and the circuit is grounded either to emitter (Pierce oscillator) either to collector (Colpitts oscillator) or to base (Clapp oscillator). The oscillator finally choosen in industries or laboratories largely depends on available components and technologies [discrete components, surface mounted components (so said s.m.c), hybrid circuits or integrated circuits]. They also depend on applications and on personal experience of designers. S.m.c. techniques allow a substantial decrease of volume together with greater reproducibility.

One of the most important advance during the last twenty years has been introduction of measurement techniques which entitle to **separately measure the noise in resonator and in electronic circuits [14].** Under those conditions, it is possible to more or less identify contributions of resonator and electronic circuits to stabilities and drift as well. To make it simple aging and medium term stability (until 10^3 s) largely depend on resonator. On the contrary, short term stability of a quartz oscillator for integration times up to $\frac{2\,Q_L}{f_0}$ (where Q_L is loaded Q of resonator and f_0 frequency of oscillator) is mainly determined by additionnal noise of sustaining circuit.

Then in the region where $\sigma_y(\tau)$ varies as τ^{-1}, $\sigma_y(\tau)$ is generally considered as proportionnal to $\frac{1}{2\,Q_L\,S_R}$ where S_R is the signal to noise ratio. Between those two time domains, $\sigma_y(\tau)$ reaches a noise floor considered as approximately equal to :

$$\sigma_y(\tau) = 1.2510^{-20} \times \nu \qquad \text{Eq. 1}$$

where ν is frequency of oscillator.

Those results are usually summarized by consideration of several noise processes each of them being characterized by the slope in a log/log plot ($\sigma_y(\tau)$ versus τ and $S_\phi(f)$ versus f) according to following chart :

	slope $S_\phi(f)$	slope $\sigma_y(\tau)$
White phase noise	0	- 1
Flicker of phase noise	- 1	- 1
Flicker of frequency noise	- 3	0
Random walk of freq. noise	- 4	+ 1/2

Since white phase noise level is given by Eq. 2 :

$$S_\phi(f) = \frac{kTF}{3\,P_0} \qquad \text{Eq. 2}$$

It is also considered that stability may be improved by choosing a higher crystal drive level if this does not degrade performance like aging (or does not create unwanted non linear effects). BVA resonators are very suitable for this application because high drive levels up to mW range do not degrade aging too much [15]. For instance, BVA_2 5 MHz AT cut fifth overtone resonators yielded following results :

Unit	Drive level	Aging/day Unit started	Aging/day after 1 year	Aging/day after 5 years
2-77	1.6 mW	3.5×10^{-10}	3×10^{-10}	1.2×10^{-10}
2-119	2.8 mW	2×10^{-9}	1×10^{-9}	

VI - Design of BVA oscillators

A very useful concept to start with is to consider that an oscillator is composed of two complex dipoles $R+JX$ one $(R_1 + JX_1)$ representing the resonator and the other $(R_n + JX_n)$ the sustaining circuit (or negative resistor) according to Fig. 4.

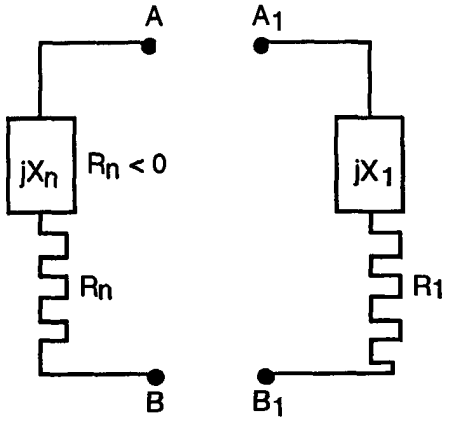


Fig. 4

The oscillator can be started by connecting A to A_1 and B to B_1. At steady state:

$$X_n = -X_1 \qquad (phase\ condition)$$
$$R_n = -R_1 \qquad (amplitude\ condition)$$

Let us suppose a resonator according to usual scheme (L_1, R_1, C_1, C_0). It is easy to see that relative frequency instabilities $\frac{\delta f}{f}$ of the frequency f will essentially be governed by Eq.3 :

$$\frac{\delta f}{f} = \frac{\delta X_1}{4\pi . L_1 . f} = \delta X_1 . \pi f . C_1 \qquad \text{Eq. 3}$$

when in the sustaining circuit a fluctuation δX_n opposite to δX_1 occurs.

On the other hand frequency pulling of resonator by a serie capacitance C_s is given by Eq.4.

$$\frac{\Delta f}{f} = \frac{C_1}{2(C_s + C_0)} \qquad \text{Eq. 4}$$

Then, it is easy to conclude that a large pulling will decrease stability and that stability is, as a first approximation, depending on L_1 according to Eq. 3 (at least when stability depends on resonator).
Let us suppose that, to start designing the oscillator, we use scheme of Fig. 5.

The scheme is first studied in "open loop" according to Fig. 6 for instance. Then, phase relation is computed between (1) and (2). Of course, the influence of C_1, C_2 and C (see Fig. 5) is also computed so as to choose the best possible values. The l, c circuit is devoted to elimination of B mode (it may be a band-pass filter or a notch filter). Of course, this circuit is suppressed when using a simply rotated cut resonator.

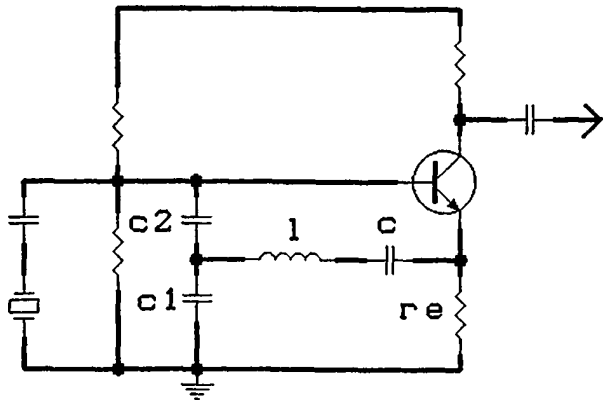


Fig. 5

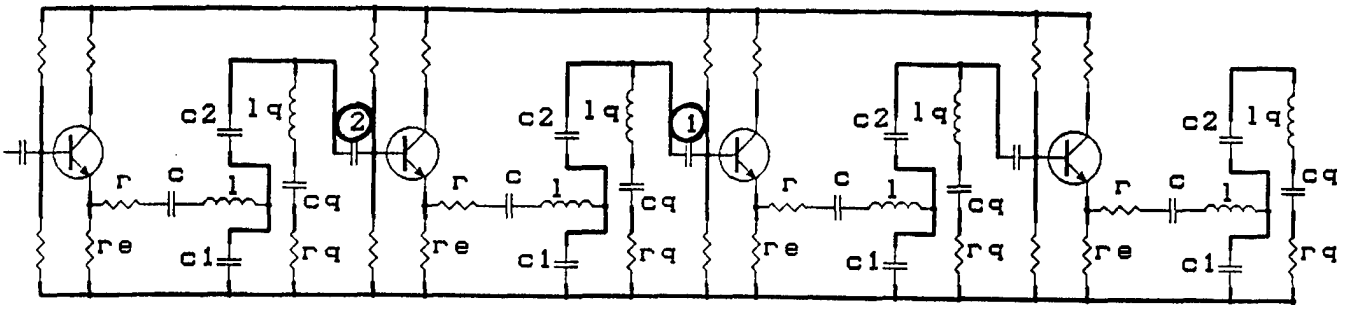


Fig. 6

It is very important that a high impedance is seen from the quartz resonator (this means that the coupling with sustaining circuit is weak).
A special simulation program (using PSPICE results on transfer function in open loop and computation under Mathematica) has been developed in our Laboratory. It yields $S_\varphi(f)$ according to parameters of quartz resonator and sustaining cicruit. In fact, using a BVA_2 resonator is convenient because the drive level can be choosen between 50 µW and 250 µW without really degrading aging. Then drive level is adjusted so as to optimize the flicker of frequency $\left(f^{-3}\ on\ S_\varphi(f)\right)$ level of noise. By the way resonator is predominant in $S_\varphi(f)$ only up to frequencies in the order of resonator's half bandwidth.

Until now, **our results have never been really limited by BVA quartz crystals but by transistor's noise**. This is specially the case with the use of 10 MHz SC, 3rd overtone tactical resonator. Under those conditions, the influence of the oscillating loop transistor has been studied using our specially devoted program giving $S_\varphi(f)$. Of particular importance is the 1/f noise pattern and its cut-off frequency. First consequence is noise selection of transistors, second consequence is of course search for "better" transistors.

VII - Tactical resonators under severe environmental conditions

Only one type of tactical resonator has been developed so far, along a ten years effort. It is a 10 MHz AT 5th overtone or SC 3rd or 5th overtone miniature bva_2 resonator. Only the SC 3rd overtone has been extensively studied in view of various programs including space programs. The resonator has been already described [9] [11] [12] [13] but a Process Identification Document (P.I.D.) has been established so that each detail of construction is documented. Symmetries, mounting and fixations, quartz bridges location, gap machining are carefully checked. Internal stresses are annealed and of course all manufacturing parameters are stored for statistical purposes. Under those conditions average typical results can be indicated (number of samples ranging from 20 to 30 according to parameter).

1/ - g sensitivity

From theoretical calculations g sensitivity of 10 MHz, SC cut, 3rd overtone plano-convex bva_2 tactical units can be evaluated. This theoretical g sensitivity is $1.8\times10^{-11}/g$ in most sensitive direction [18]. Average typical experimental value correspond to $9.3\times10^{-11}/g$ (worst direction). Discrepancy is connected to construction's imperfections that can eventually be traced back by carefully following all steps of P.I.D. [17]. Principal defects are misorientation of bridges, offcentering of contour, offcentering of bva_2 structure's grooves. Introduction of those defects in theoretical model yields theoretical value much closer to reality.
Some special bi-convex 10 MHz SC cut units of the tactical type have been designed and produced corresponding to a g sensitivity of $4.4\times10^{-11}/g$ worst case. For comparison purposes let us point out that best AT cut plano-convex units correspond to g sensitivity in the vicinity of $3\times10^{-10}/g$.

2/ - Vibrations

• Sinusoïdal vibrations

After its evolution our design is now capable of standing sinusoïdal vibrations of 80 g (20 Hz - 2000 Hz) during 15 minutes per axis on 3 axes after standing same treatment at 40 g, 50 g, 60 g and 70 g.

• Random vibrations
Latest design can stand following random vibrations :

+ 6dB/oct	(20 Hz - 100 Hz)
$1.0g^2$/Hz	(100 Hz - 600 Hz)
- 6dB/oct	(600 Hz - 2000 Hz)

3/ Shocks
The resonator can stand shocks of 3000 g during 0.1 ms on all 3 axes (6 shocks).

4/ - Pressure variation [0, 2 bars]
This measurement is not trivial because thermal effects have to be clearly separated from pressure effects. Pressure variation sensitivity varies from unit to unit (it may eventually be 0) but a typical value is 1×10^{-10}/bar.

5/ - Magnetic field
This sentitivity has been measured between -10 gauss and +10 gauss all values range between 2×10^{-12}/gauss and 3.9×10^{-12} /gauss average is 3×10^{-12} /gauss.

6/ - Irradiation effects
Tactical bva resonators have been submitted to high iradiation doses (up to 550 krad) [16]. Radiation suceptibility then depends upon Aluminum content, specific crystal defects and manufacturing process.
Radiation suceptibility of BVA-AT resonators [19] and BVA-SC resonators [20] has been tested in Applied Physics Laboratory (Laurel MD).
It appeared that, compared to usual resonators, BVA-AT resonators exhibited lower suceptibility to low doses of gamma radiation (i.e, simulated medium earth orbit radiation). Sensitivity did not seem to be correlated to the aluminium defect center concentration. BVA SC resonators exposed to 0.6 Rad (Si) for 42 min at a dose rate of 0.014 Rad (Si)/min exhibited sensitivities of some 10^{-11}/rad but interpretation of results (including insensitivity of one resonator) was not easy.

VIII - Link between material's quality and parameters measured on BVA resonators or oscillators

Some applications including space applications are very demanding, then piezoelectric material's quality has to be carefully checked before starting resonator fabrication. Of course, ordinary test methods for quartz material are routinely used (i.e, determination of infrared absorption coefficient, etch channels density, inclusions density, dislocations density). Results obtained with BVA resonators confirm usual results. However BVA_2 resonators do not have electrodes sitting on the crystal and their fixations are monolithic so interpretation of results may be easier. For better understanding we have decided to undertake some experiments.

a/ - Motional resistance R_1 versus temperature
The different behaviours existing between kinds of raw material can also be exhibited on the motional-resistance temperature curves. The following curves of figure 7 present typical behaviours for 5 quartz materials. For example, for a resonator in natural quartz, the resistance increases slowly with the temperature. On the contrary, for a swept synthetic material, the resistance is roughly constant. For all the resonators made in quartz said "US High Q", the resistance increases up to 50°C and decreases after.

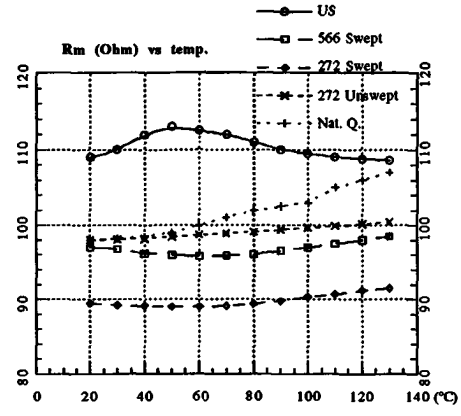


Fig. 7 : different resistance-temperature curves

Table 2 includes mean Q values calculated at 25°C for the previous five raw materials.

Series	17-18	20	21-23	23-25	26	27
Raw mat.	Nat.	272	272S	566S	Uns.	US
Q (*10^6)	1.2	1.36	1.37	1.32	1.28	1.18
Std dev.	0.08	0.02	0.02	0.03	0.03	0.04

Table 2 : different Q-factors (more than 10 units per serie)

b/ - Frequency changes under irradiation
Results concerning nine bva SC-cut resonators tested under irradiation are presented in table 3. The three first ones, issued from natural quartz of Madagascar present large frequency shifts; resonance frequency decreases quasi-linearly with time exposure. The two following resonators (samples of SICN 563, unswept) are less sensitive; the slope before saturation is lower than 10^{-12}/rad. Only one resonator (SC07, 08, made out of synthetic quartz from SICN too, run 309) tested in a high dose environnement exhibited a clear $\Delta F/F$ shift which was positive and in the order of a few parts in 10^6.

RESONATOR N.	17.04	17.08	18.05	15.07	16.08	07.08	19.06	19.11	20.05
IRR. DOSE (Krad)	550	300	300	300	550	300	350	300	350
On T.O. Point: DF(After-Before) In Hertz	-197	-102	-117	-3	-0.5	+99	≈ 0	-4	-1
DF/F /rad (10^{-12})	-36	-34	-39		<-0.1	+33	<-0.01		<-0.1
Saturation (Krad)	?	NO	NO	≈ 100	?	> 400	< 100	< 100	
Before Irradiation Turn-Over (°C):	84	87	87	80	80	90	78	82	79
After Irradiation	84	87	88	76	76	77	76	79	78
Before Irradiation F(T.O.) - F(25°C):	241	250	258	201	220	?	219	235	221
After Irradiation	217		254	192	192	198	187	201	196
Before Irradiation Q-factor at 25°C:	1.32	1.3	1.28		1.38	1.05	1.36	1.38	1.35
After Irradiation	0.65	0.58	0.53		1.31	0.77	1.35	1.36	
Before Irradiation Rm (Ohm) at T.O.:	104	97	102	102	100	100	93	94	102
After Irradiation	164	145	165	130	102	260	94	95	98

Table 3

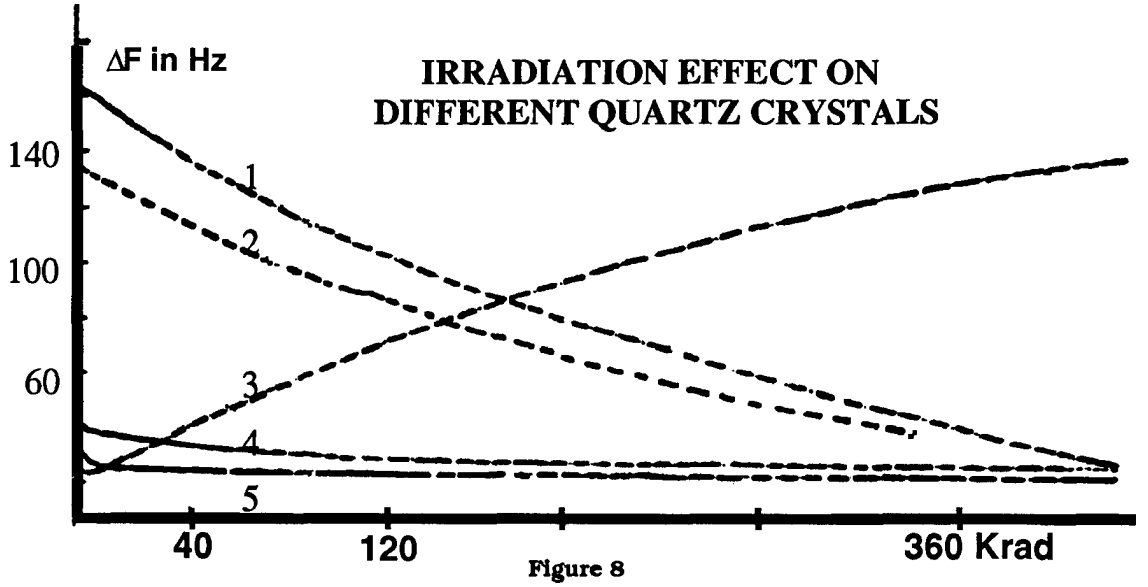


Figure 8

Figure 8 shows graphically the relative behaviors of 5 resonators (submitted to 300/400 Krad accumulated dose) :

1. 18.05 : natural quartz (from Madagascar)
2. 17.08 : natural quartz (- id -)
3. 07.08 : synthetic from SICN (309 unswept)
4. 15.07 : SICN 563 unswept
5. 19.11 : SICN 272 swept

From these curves, we can estimate if a saturation effect exists after a very high dose of γ-ray. The only possible conclusion is that swept synthetic quartz with low Al impurity contents is very rapidly insensitive to γ-ray.

γ-ray dose rate is about 3.5 gray/min (350 rad/min).

So during a 20 hours exposure, the quartz has received dose slightly higher than 400 krad.

c/ - Long term drift

Long term drift may be caused by a lot of phenomena. Various effects (humidity, pressure, oven temperature variations, aging of sustaining circuit components ... and so on) are often attributed to aging of resonator! A true

measurement of resonator's aging would in fact need a passive measurement exactly at turn over point (the resonator being kept in precise conditions of temperature and drive level). This has not been properly done yet. Nevertheless, we have studied bva oscillators drifts whose frequency f can be fitted versus time t by :

$$f(t) = f(t_0) + K_r \, Log\,(I + Bt)$$

Results seemed to indicate in resonators, two important causes of aging namely mass transfer from or to the vibrator inside the resonator's can and stress relaxation in the structure. At that point, we have been successfull to obtain very low aging of oscillators (always under 3×10^{-11}/day and often in the 10^{-12} range) but still we do not perfectly understand behaviours like those of Fig. 9a and Fig. 9b corresponding to first series of tactical bva_2 SC-cut 10 MHz 3rd overtone resonators.

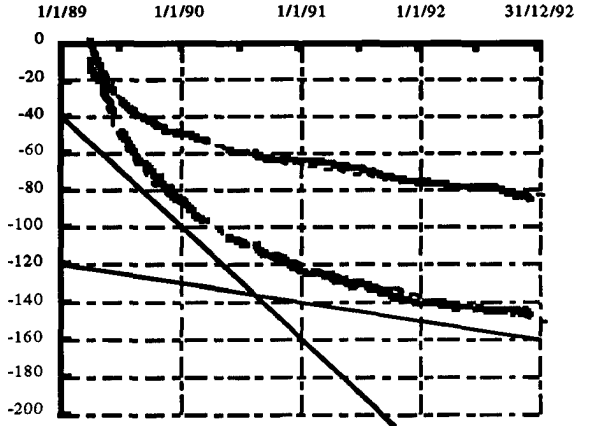


Fig.9a: L.T. drift on 07.04 and 07.11 (10^{-10})
$<2*10^{-11}$/day after 3 months and $<3*10^{-12}$/day after 1 year
Best fits : -22*Ln(0.03t + 1) and -40*Ln(0.03t + 1)

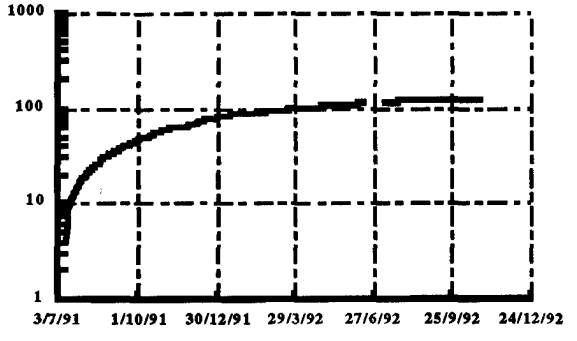


Fig.9b: L.T. drift on 16.00 (1.5 year) in 10^{-10}

However, with use of precise P.I.D. (in further series) behaviour of Fig. 9a almost totally disappeared. Furthermore, a serie of 10 units mounted in precision commercial oscillators, has been extensively tested outside our laboratory. All units aged in a way similar to Fig. 9b and, after 12 months, drift was 10^{-11}/day or less.

d/ - 1/f noise in quartz BAW devices

This **crucial** question has not yet got a definite answer though several attempts have been made [5], [22], [23].
Our work in the domain confirms the critical part played by surface condition and will be published later on. Basically, crystals may improve their flicker levl of moise when surface reconditionning has been properly operated. Unde those conditions, it may be understood why some crystals exhibiting higher Q factors nevertheless yield a higher level for flicker noise. Also since phenomena in SAW and BULK devices are rather different, it is suspected that theoritical models differ the one from the other. Still here, due to simplicity of construction, BVA type resonators may help to clarify among different mechanisms involved for 1/f noise in BAW devices and seem to proove influence of small surface defects (in the µm range) due to resonator's surface preparation.

IX - Some typical results, conclusions

We shall first give principal parameters for bva_2 10 MHz SC-cut, third overtone plano-convex unit because this resonator has been largely produced and extensively studied espacially for space applications.

Frequency at turnover	:	9 999 990 -5 +3 Hz
Enclosure	:	HC 40/U
Serie capacitance	:	30 pF (nominal)
Turnover	:	80 -3 +5° C
Q factor	:	1.35×10^6 (typical)
R_1	:	95 Ω (maximum)
L_1	:	2.2 H (typical)
Aging	:	better than 3×10^{-11}/day after 30 days
g sensitivity	:	10^{-10}/g (typical)
Pressure sensitivity	:	$< 1 \times 10^{-10}$/bar (typical)
Magnetic sensitivity	:	$\leq 3 \times 10^{-12}$/gauss (typical)

Results obtained with such a resonator in BVA tactical oscillator (inside 250 cm^3) are as follows :

Stabilities	:	between 1×10^{-13} and 3×10^{-13} ($10\ s \leq t \leq 100\ s$)
Spectral purity	:	-123 dBc/Hz at 1 Hz , -158 dBc/Hz at 100 kHz
Aging	:	between 5×10^{-12}/day and 3×10^{-11}/day
Stability versus temperature	:	$\leq 5 \times 10^{-12}$/° C from -20° C to +60° C
g sensitivity	:	$\leq 10^{-10}$/g (typical)

It is interesting to point out that, using BVA_2 5 MHz resonators Jerry Norton [21], could obtain stabilities down to 4×10^{-14} with a spectral purity of -130 dBc/Hz at 1Hz.

Some promising efforts have been started for improved results at 10 MHz because the resonator stands high levels of accelerations and vibrations and because miniaturization is suitable with a 10 MHz unit. We now believe that values clearly under 10^{-13} will be obtained for $\sigma_y(\tau)$ at 10 MHz together with aging rates under 10^{-11}/day and interesting behaviour in hard environmental conditions.

REFERENCES

[1] R.J. BESSON, *A new electrodeless resonator design*; Proc. 31st Annual Symp. Freq. Control, pp. 147-152 (1977).

[2] R.J. BESSON, Proc. 30th Annual Symp. Freq. Control. pp 78-82 (1976) and Proc. 10th Precise Time and Time Interval (PTTI) applications and planning meeting, pp 101-130 (1978).

[3] French patents n° 7601035, 7717309, 7802261, 7828728, 7918553, 8110006, 8215351, 8509097, 8814197, 9203175 and corresponding patents or patents pending in other countries.

[4] J.P. VALENTIN, M.D. DECAILLIOT, R.J. BESSON, J.J. GAGNEPAIN, *New approach of fast warm-up for crystal resonators and oscillators,* Proc. 38th Annual Freq. Cont. Symp., pp 366-373 (1984).

[5] F.L. WALLS, P.H. HANDEL, R.J. BESSON, J.J. GAGNEPAIN, *A new model relating resonator volume to 1/f noise in BAW quartz resonators,* Proc. 45th Annual Freq. Cont. Symp., pp 327-333 (1992).

[6] C.AUDOIN, *Transactions on Instrumentation and Measurement* ; IEEE, Vol. 40, pp 12 (1991).

[7] M. DECAILLIOT, R.J. BESSON, "Fast warm-up resonator", French patent n° 8110006 (May 1981).

[8] R.J. BESSON, R. DELAITE, G. GENESTIER, P. MAITRE, G. RENARD, D. THIEBAUD, J.P. VALENTIN, French Patent n° 8509097 (June 1985) corresponding patents in other countries.
R. DELAITE, "Une nouvelle génération de résonateurs piézoélectriques miniaturisés de haute stabilité, Proc. 1st EFTF (European Frequency and Time Forum), pp. 57-60 (1987).

[9] R.J. BESSON, R. DELAITE, G. RENARD, J.P. VALENTIN, J.L. VATERKOWSKI, French patent n° 8814197 (Oct. 1988) corresponding patents or patents pending in other countries.

[10] R. DELAITE, "Special miniaturized SC-cut quartz crystal resonator for severe environments applications", Proc. 3rd EFTF, pp. 296-300 (1989).

[11] J.J. BOY, A. BERTHAUT, G. RENARD and al, "Characterization of BVA in industrialization process", Proc. 7th EFTF, pp 57-62 (1993).

[12] M. MOUREY, R.J. BESSON, A space oscillator with cylindrical oven and symmetry ; Proc. 45th Annual Symp. Freq. Control, pp 431-441 (1991).

[13] J.R. NORTON, R.J. BESSON, Tactical BVA quartz resonator performance ; Proc. 47th Annual Symp. Freq. Control (1993).

[14] F.L. WALLS, A.E. WAINWRIGHT, IEEE Transactions on Instrumentation and Measurement, Vol. IM 24, pp 15-20 (1975).

[15] A. BERTHAUT, R. BESSON, "Résonateur à ondes de volume à très fort niveau d'excitation", C.R.Acad. Sc. Paris, t 289, Série B, pp 223-224 (19/11/79).

[16] J.J. BOY, A. BERTHAUT, D. THIEBAUD, K. YAMNI, "Long term drift and γ irradiation effect", Proc. 6th EFTF, pp 93-98, (1992).

[17] J.J. BOY, F. DEYZAC, A. BERTHAUT, "Accelerometric sensitivities of BVA resonators", to appear in Proc. 9th EFTF (1995).

[18] F. DEYZAC, personal communication.

[19] J.J. SUTER, A.G. BATES, J.M. CLOEREN, "The radiation susceptibility of BVA-AT resonators to simulated medium Earth orbit radiation", Proc. 2nd EFTF, pp 679-686, Neuchatel (1988).

[20] J.J. SUTER, A.G. BATES, J.M. CLOEREN, J.R. NORTON, "The susceptibility of BVA-SC resonators to photon ionization effects", Proc. 3rd EFTF, pp 11-21, Besançon (1989).

[21] J.R. NORTON, personal communication.

[22] T.E. PARKER, D. ANDRES, "Dependence of SAW resonator 1/f noise on device size", IEEE Inter. Freq. Control Symp., pp. 178-185 (1993)

[23] M.M. DRISCOLL, W.P. HANSON, "Measured vs volume model-predicted flicker of frequency instability in VHF quartz crystal resonators, IEEE Inter. Freq. Control Symp., pp. 186-192 (1993).

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SATELLITE MASTER OSCILLATORS

D. A. Emmons, D. W. Martin, and R. M. Garvey

Frequency and Time Systems, Inc.
34 Tozer Road, Beverly, MA 01915

Abstract

Quartz oscillators which serve as the time and frequency references on satellite missions are called upon to meet stringent performance requirements. Recent applications call for the best possible stabilities and low noise performance while at the same time demanding features such as high resolution remote digital tuning, which can potentially compromise ultimate stability. In this paper, we present test results and a discussion of some design criteria and strategies for meeting requirements of high precision, survivability and reliability.

The error budget for critical performance factors is examined within the context of a worst case analysis. This includes resonator aging vs. tuning range as it is impacted by tuning step resolution requirements. Frequency deviations can be driven by environmental factors, such as: temperature extremes and rate of change of the temperature on orbit; local magnetic fields; fluctuations in the level of ionizing radiation, and the constant accumulation of radiation dose.

Thermal control for the quartz resonator oven achieves small temperature coefficients. Thermal design requires trade-offs between the ideal of thermal isolation from the ambient, and rugged construction. Long term testing of Allan deviation is described, and aging rate results in the low e-11 range are presented. High resolution test methods and results for radiation sensitivity screening are discussed, with particular reference to low dose testing. Resonator sensitivities in the 1e-11 per rad range for a 4 rad/hour exposure have been observed for 4 MHz adhered electrode SC-cut resonators.

Introduction

Master Oscillators (MO) for space-borne applications require a wide variety of attributes for specific missions. Weight and power consumption are always important. Size is important, but small size is often not compatible with the required stability performance.

Although a given application must have design solutions which are uniquely tailored to satisfy form, fit and function, there are some attributes common to all. These include high performance in terms of frequency response to varying external environmental perturbations.

The major role of a MO is to provide a highly stable, low noise reference frequency. When requirements for remote tuning, monitor telemetry, on-off command, multiple outputs, frequency multiplication and other features are added, the trade-offs between performance, size, power, cost and reliability become more complex.

In this paper we discuss FTS Model 9500 oven controlled quartz oscillators designed for use in the 4-20 MHz range, having short term Allan deviation stabilities of less than 1e-12, and temperature stability of better than 2.5e-12/°C. These satellite application oscillators utilize 3rd overtone, SC-cut resonators, exhibit very low aging, and can have remote digital tuning capability.

Frequency Error Budget

Early in the design phase, it is useful to generate an error budget comprised of the factors which will contribute to changes in frequency over the life of the oscillator. This tool a) identifies and quantifies sensitivities to the expected environment; b) defines end-of-life effects; and c) identifies tradeoffs necessary to decide if external frequency tuning is needed. This error budget must address questions of stability, frequency aging, radiation effects, oven temperature control and other issues. Table I shows a version of such a budget.

The frequency error budget has some limitations, however. Many of the contributors to the error budget can be identified only with significant uncertainty. An example might be the frequency change associated with mechanical shock: its bounds can only be estimated based upon prior experience with similar designs, since the magnitudes of frequency shifts associated with mechanical shock are typically variable. Another

0-7803-2500-1/95 $4.00 © 1995 IEEE

example is the variability of frequency retrace after periods of power off, particularly when low temperature storage is involved. There is also the consideration of how to combine the various effects, whether as an algebraic sum or as a root sum square. Worst case estimates can drive the complexity and cost upwards while compromising performance.

Table I. Partial Error Budget

Parameter	Value
Aging (10 years)	
1st year (@ 1e-10 per day)	3.7 e-8
9 years (@ 4e-11 per day)	13.2 e-8
Radiation	mission dependent
Initial Frequency Centering	4 e-8
Shock and Vibration (Launch)	1 e-8
Zero gravity (earth-to-orbit shift)	2 e-9
Frequency Retrace	3 e-9
Algebraic Sum of Effects (Not RSS)	22 e-8
Root Sum of Squares (RSS)	14 e-8

Block Diagram

The block diagram of Figure 1 is an example of a specific configuration of the Model 9500 which provides two 5 MHz RS-422 clock outputs, and two sinewave outputs, one at 5 MHz and the other at 10 MHz by frequency doubling. The 10 MHz results discussed in this paper, however, were obtained with 10 MHz resonators and not via frequency doubling.

The resonator and the oscillator circuits, as well as oven control, digital-to-analog converter (DAC) and 12 volt regulator circuits, are located within the oven. Command, telemetry, D.C. power circuits, frequency multipliers and RF output drivers are located outside the oven. Thermistors are provided for oven temperature and case temperature monitoring.

Frequency Stability

Frequency stability of the MO depends strongly on the temperature stability of the oven. In addition, the inherent stability of the high-Q resonator must not be compromised by circuit effects or voltage regulator noise. Direct measurements have shown that the effective oven thermal gain is more than adequate at greater than 3000. Oven design and control are discussed in a later section.

Characterization of frequency stability for this series of oscillators has been carried out using a number of different test methods. In the short term, 1 to 100 second averaging times, Allan deviation is measured against a reference oscillator of performance comparable to the oscillator under test, using beat frequency comparison. The very good frequency stability seen in both 4 and 10 MHz Master Oscillators is consistent with the measured phase noise.

In the longer term, we have employed a heterodyne system in which 1 pps phase data is accumulated over time periods up to 30 days, with computer controlled data collection. Over long averaging times, variance measurements are affected by fluctuations in environmental conditions, due to oscillator sensitivities. Long term drift, characterized as linear aging, is tracked with respect to a cesium reference.

10 MHz Results

Allan deviation with tau extending to 2e4 seconds is shown in Figure 2. The data represent the combined noise of both reference and oscillator under test. This oscillator shows a daily aging in the low e-11 range, which is reflected in the long-term Allan deviation.

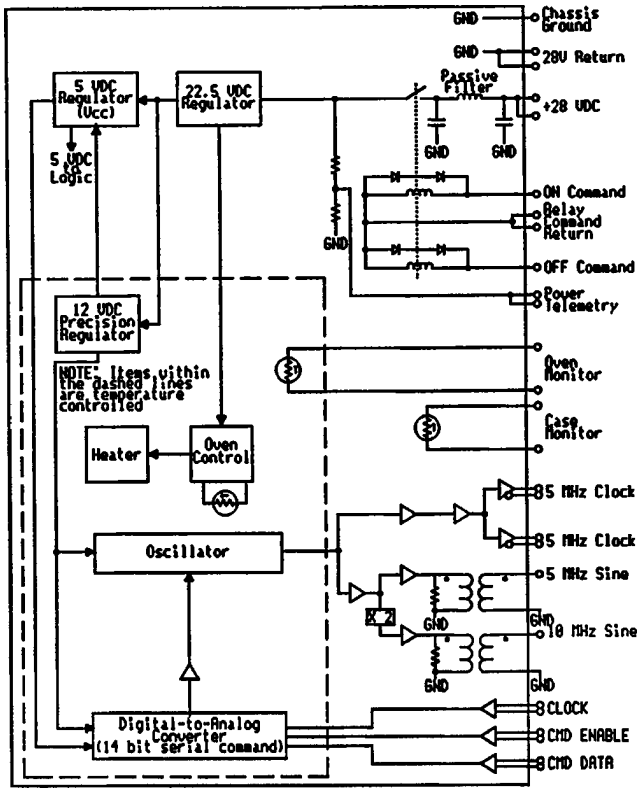


Figure 1. Block Diagram of Master Oscillator

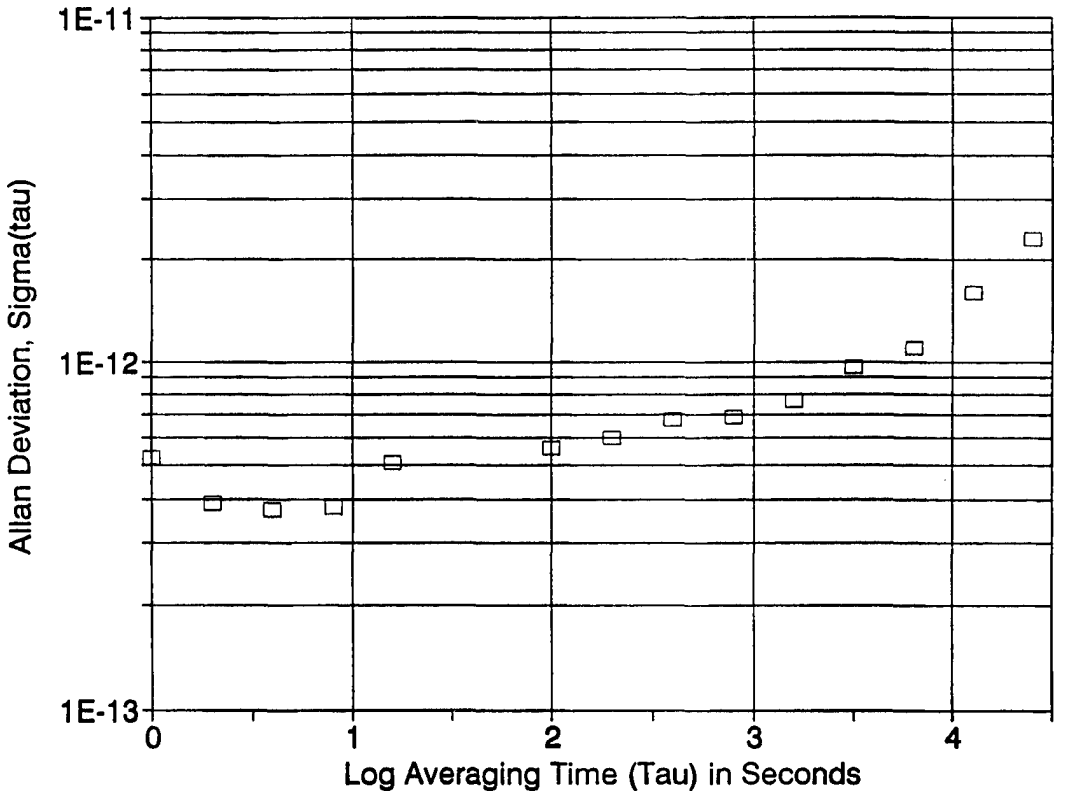


Figure 2. Frequency Stability, 10 MHz MO

30-Day Frequency Stability

Very-long-term stability was characterized for a 4 MHz MO. The measurement technique is to mix the 4 MHz output with a synthesized offset signal (referenced to cesium) set 1 kHz higher in frequency. The beat signal is divided down to 1 pps which is then phase compared with a cesium derived 1 pps, via a time interval counter. Averages over 1000 seconds are stored on the computer hard disk and can be analyzed at any time during the run [1].

Figure 3 shows the frequency history over 30 days in a laboratory environment. The linear aging is 5.2e-12 per day. Figure 4 is a plot of Allan deviation for this same data set, with drift removed. The stability of 1e-11 at tau = 1 day is remarkable, but in fact would certainly be better in a controlled environment (temperature, humidity, barometric pressure, etc.)

Testing in Thermal Vacuum

Allan deviation was also measured on 4 MHz oscillators in vacuum. Here one is faced with the possibility of periodic disturbances associated with temperature control of the baseplate, or vibration associated with vacuum pumps and refrigeration compressors. Results are shown in Figure 5. Since absolute frequency drift was of interest, the reference for the measurement was a cesium standard which contributes frequency noise at 100 to 300 second tau. The 4 MHz oscillator shows an Allan deviation below 4 e-13 between 100 and 1000 seconds in the vacuum environment.

Frequency vs. Temperature

Both static (equilibrium ambient conditions) and dynamic temperature coefficient characteristics are of interest for satellite applications. The static behavior of frequency vs. ambient temperature, in vacuum, is shown for three MOs in Figure 6. The equivalent coefficient for the 10 MHz unit is 5e-13/$^{\circ}$ C. Because the oscillator oven mass is large and the thermal isolation quite good, the thermal time constant for frequency equilibrium is more than 2 hours. This makes for lengthy test cycles. If the temperature coefficient were simply related to the theoretical shape of the SC-cut resonator f-T curve, then the large thermal gain of the design would ensure a nearly zero result. In fact, residual temperature sensitivities of elements other than the resonator contribute.

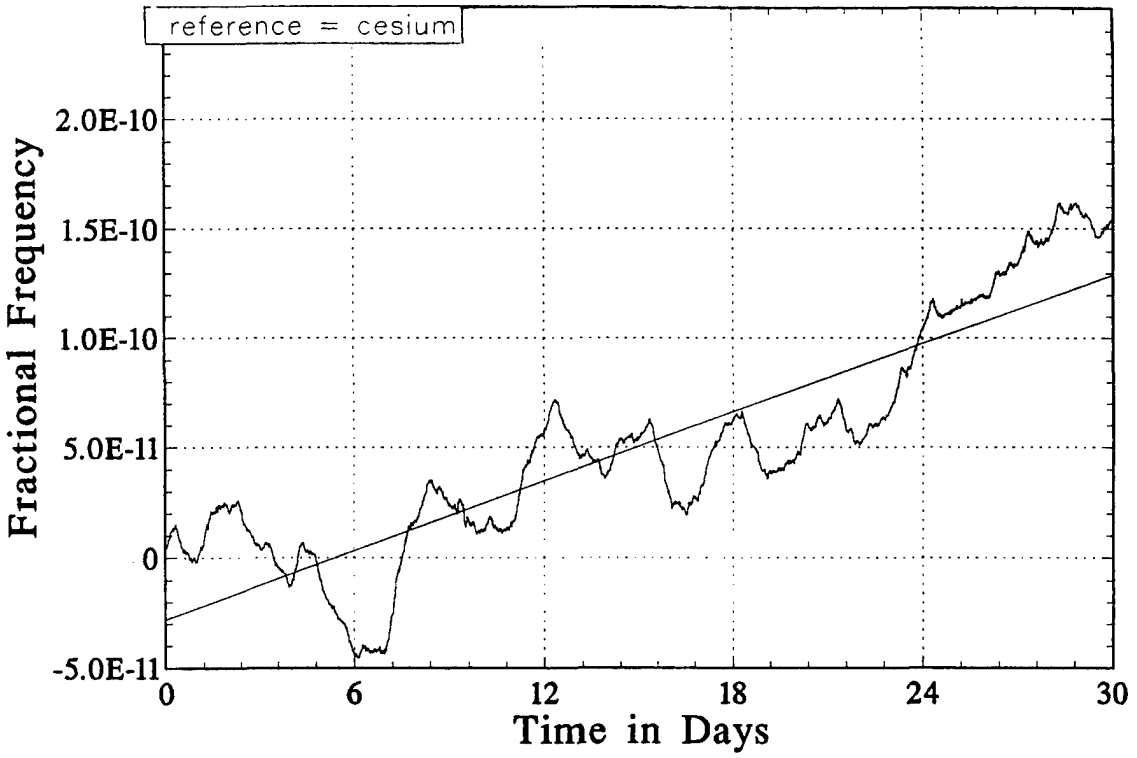


Figure 3. Thirty-Day Stability, 4 MHz MO, Lab Ambient

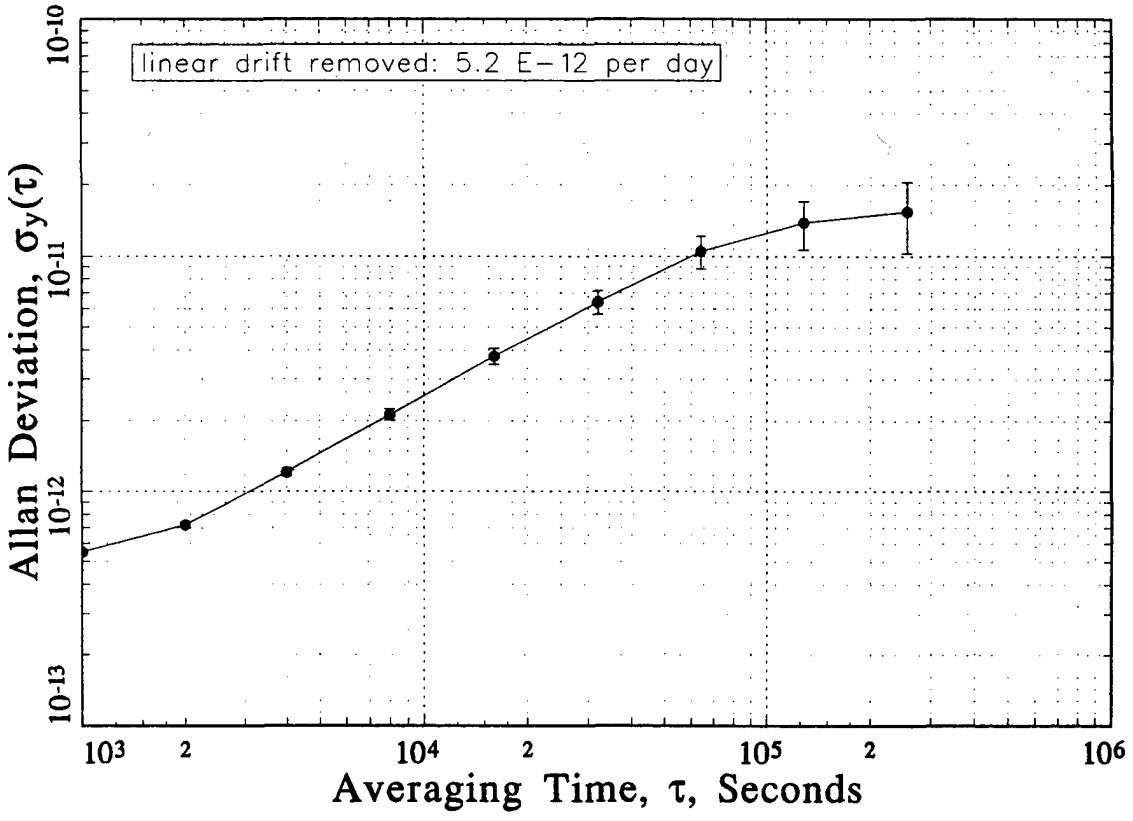


Figure 4. Allan Deviation, Thirty-Day Test

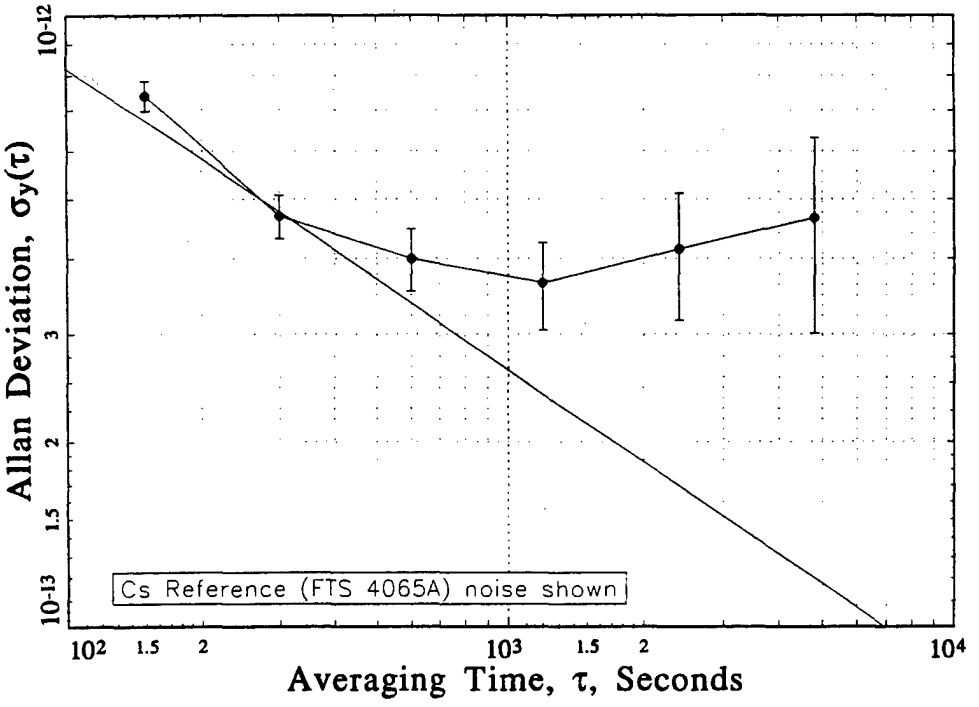


Figure 5. Stability in Vacuum Chamber

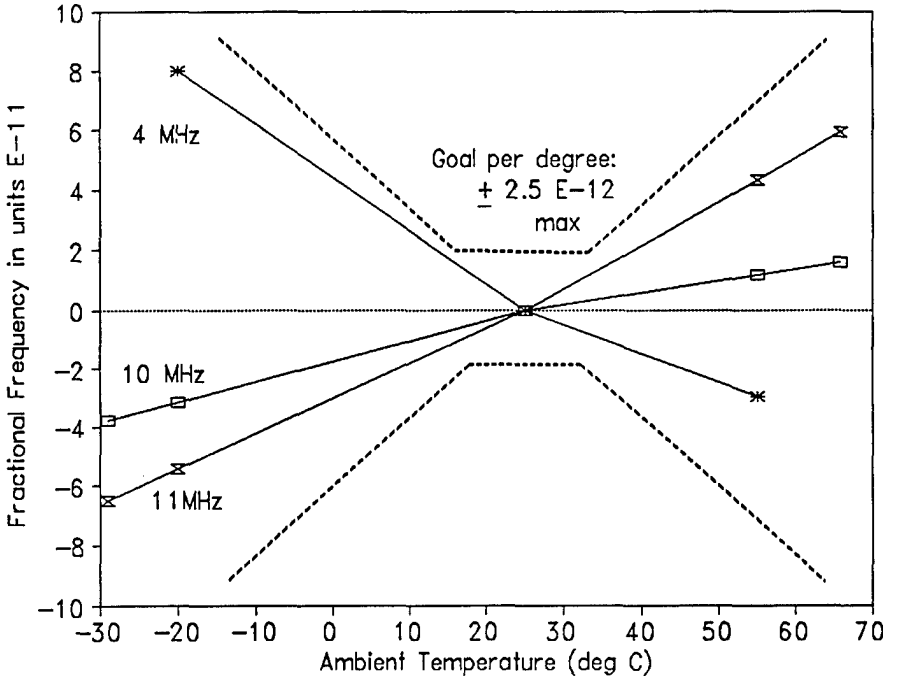


Figure 6. Frequency vs. Temperature in Vacuum

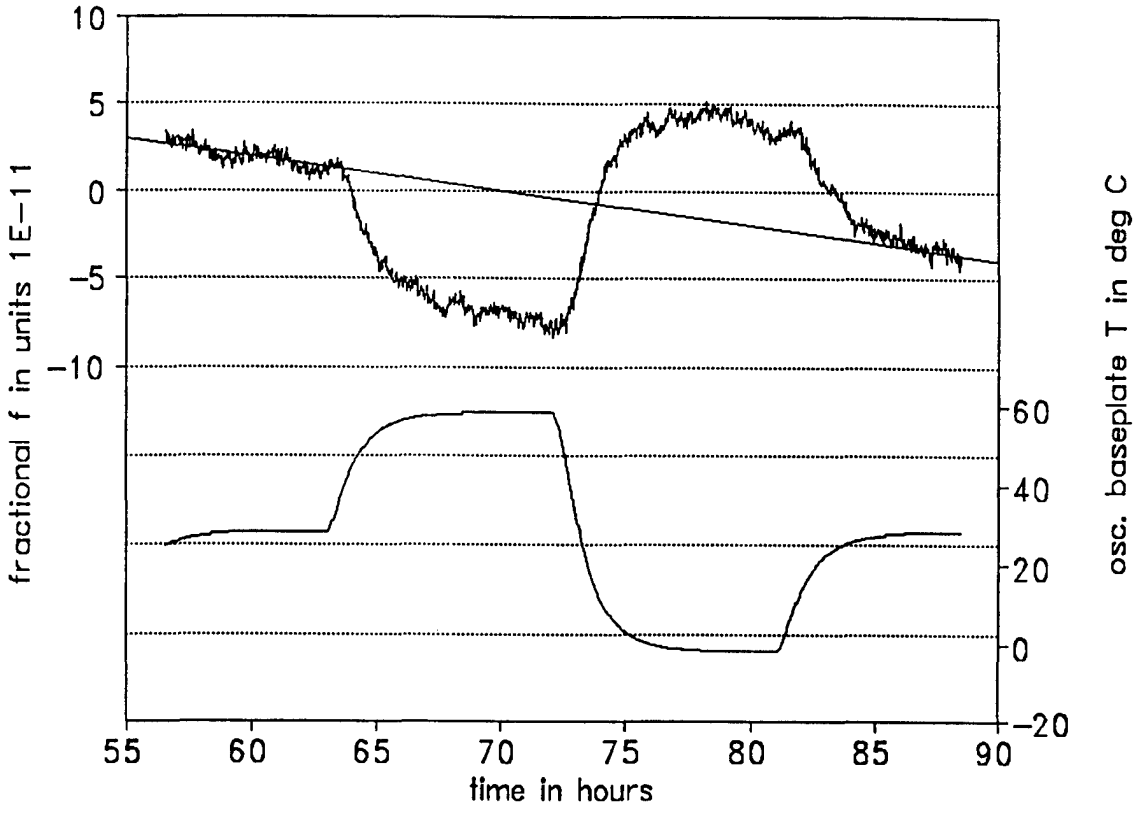


Figure 7. Stability, Temperature Cycling in Vacuum

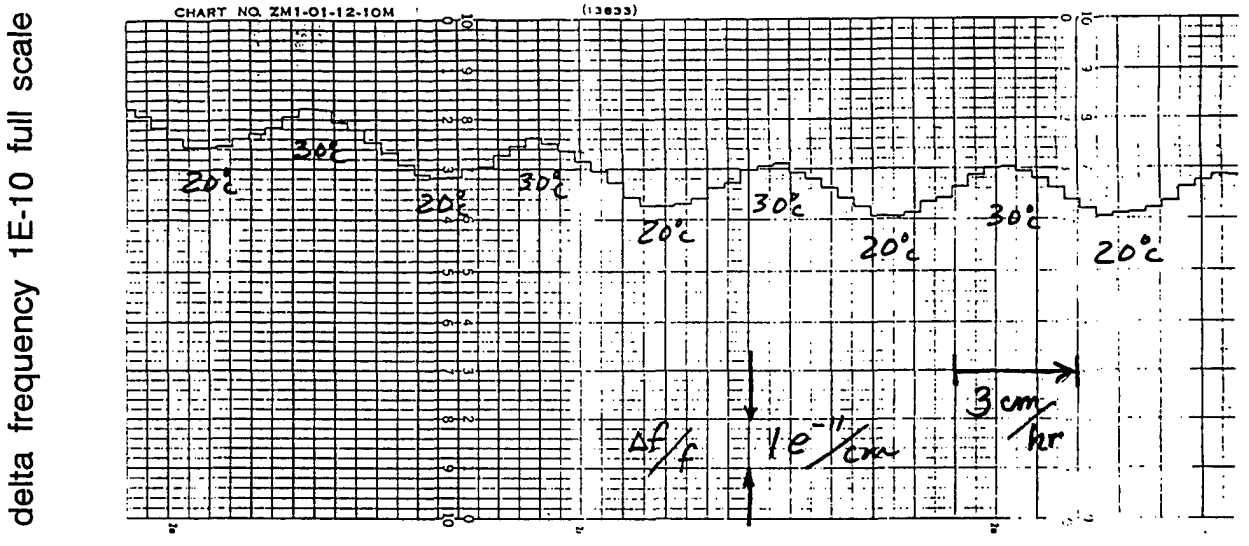


Figure 8. Square-Wave Temperature Cycling in Vacuum

Static and dynamic thermal response are shown for a 4 MHz oscillator in vacuum, in Figure 7. The low pass thermal filter behavior of the oven serves to attenuate the part of the crystal response which is driven by temperature rate of change.

Dynamic Temperature Environment

For low earth orbit MOs, the effects from an orbital period of approximately 100 minutes are of particular interest. A 4 MHz oscillator was subjected to periodic temperature fluctuations, in vacuum. The chamber temperature profile was a square wave of 10°C peak-to-peak, at 25°C average. The result is shown in Figure 8. Peak-to-peak frequency excursions are 1e-11, giving an effective dynamic temperature coefficient of 1e-12/°C .

In a second test, the MO baseplate temperature was varied sinusoidally, 10°C peak-to-peak, with a 100 minute period. In this case the average temperature, in vacuum, was 30°C. The Allan deviation under this dynamic temperature condition is shown in Figure 9. Again, the cesium reference noise contribution is shown, and the performance of the MO between 10 and 100 seconds averaging time is seen to be 2 e-12 or better. The half-period enhancement in Allan deviation, at 50 minutes tau, is only 5 to 6 e-12. Low resolution for the data beyond 1000 seconds makes the apparent peak position misleading.

Phase Noise Measurements

Phase noise results for both 4 and 10 MHz units are shown in Figure 10. In each case, the reference oscillator was of comparable noise performance to the unit under test. The 4 MHz unit shows (-124 - 3) dBc/Hz at 1 Hz; and for the 10 MHz unit, the result is -118 dBc/Hz at 1 Hz. The noise floor at 10 kHz is -165 dBc/Hz for both units.

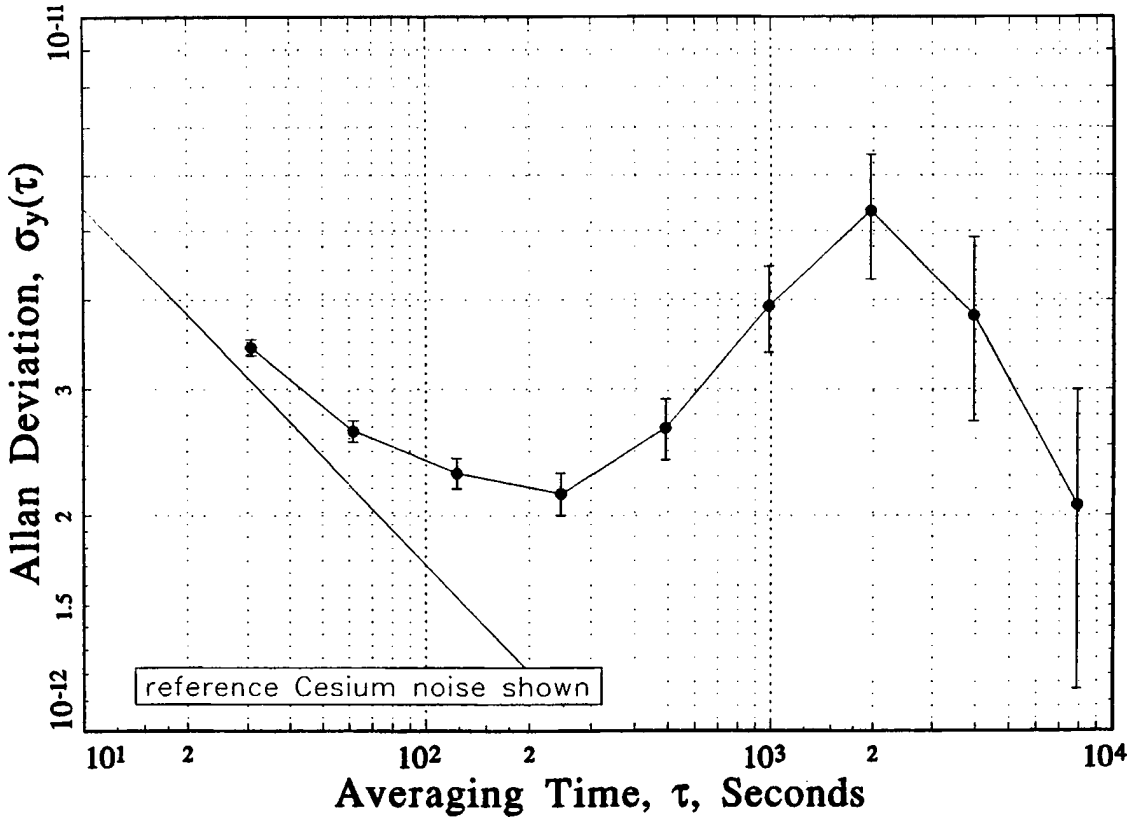


Figure 9. MO Stability in Vacuum, Sinewave Temperature, 100 minute Period

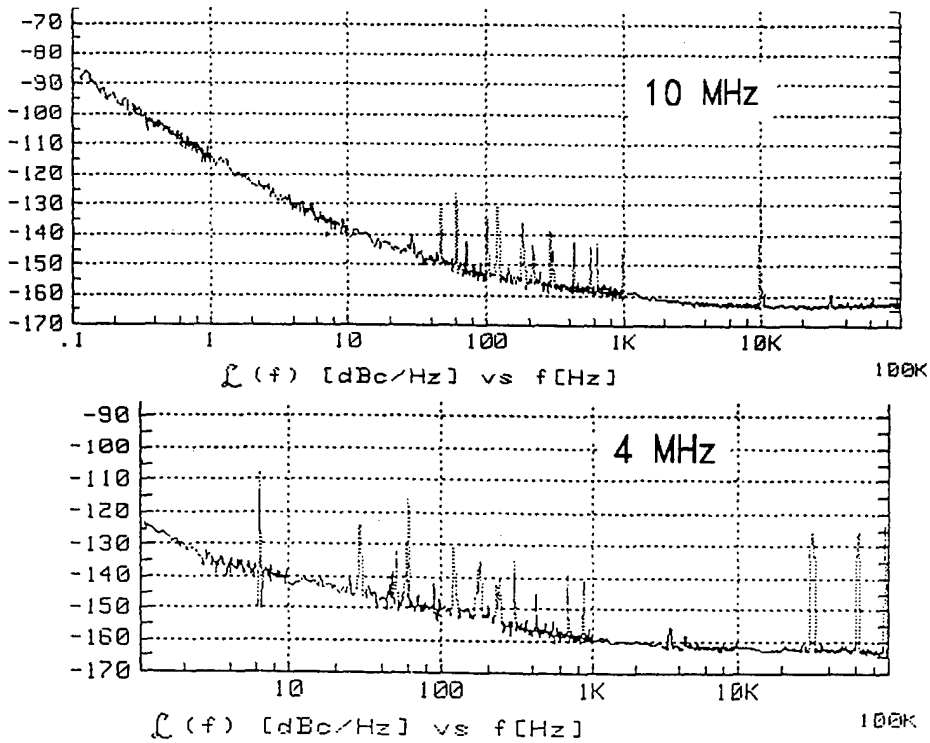


Figure 10. MO Phase Noise

Retrace

Frequency retrace is not usually considered along with other stability parameters. However, when a MO is required to meet very stringent frequency error budget demands, it is of interest to characterize changes in frequency due to periods of dormancy and power cycling. For best stability, the oscillator should simply be powered on all the time. However, it may, for any number of reasons, be turned off at some point by ground command. The subsequent power-on retrace then has aspects of "dormant aging" mixed with the effects of thermal-mechanical stress hysteresis, as they affect resonator frequency. A conservative estimate is usually attached to the retrace frequency error budget. However, a quite small retrace result was seen for a 4 MHz MO which was turned off for a period of 56 hours. This result is shown in Figure 11, on a folded logarithmic plot, where "zero" frequency offset means in this case, all values between 1e-12 and -1e-12. Frequency retrace was within 1e-11 of the frequency measured prior to turn-off.

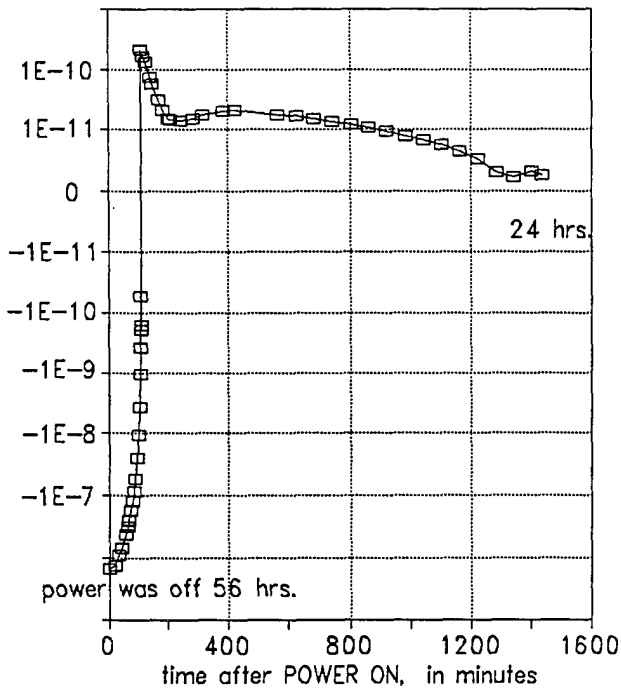


Figure 11. Frequency Retrace

Mechanical Design

The Model 9500 is designed to permit hard mounting to the spacecraft by means of an aluminum baseplate. The oven assembly is contained within a glass Dewar, a design which is derived from proven rugged and reliable FTS designs. The Dewar is wrapped with a cylinder of high permeability iron alloy which serves to shield the oscillator circuitry from external magnetic fields. (It also adds to the radiation shielding.) The resultant sensitivity is 1e-12 per Gauss, an important consideration when changing ambient magnetic fields are encountered. Other details of the Dewar assembly are discussed in the thermal isolation section below. A side compartment of the MO chassis contains power conditioners, buffer amplifiers, and connectors.

Mechanical Isolation

In some cases, analysis of the mechanical threat may show a need for shock and vibration protection for a specific component or subassembly, such as the quartz resonator. Figure 12 shows a photograph of FTS Model 9521 which has vibration isolating elastomer mounts. These afford a high-frequency attenuation of approximately 20 dB per decade above 250 Hz. For the specific application this gives a large margin of survivability for the quartz resonator in a critical frequency range of shock and vibration.

Since the oscillator baseplate would be thermally isolated in this case, thermal links in the form of flexible metal braid are provided to carry a portion of the two to four watt (typical, depending on ambient temperature) heat load to the spacecraft.

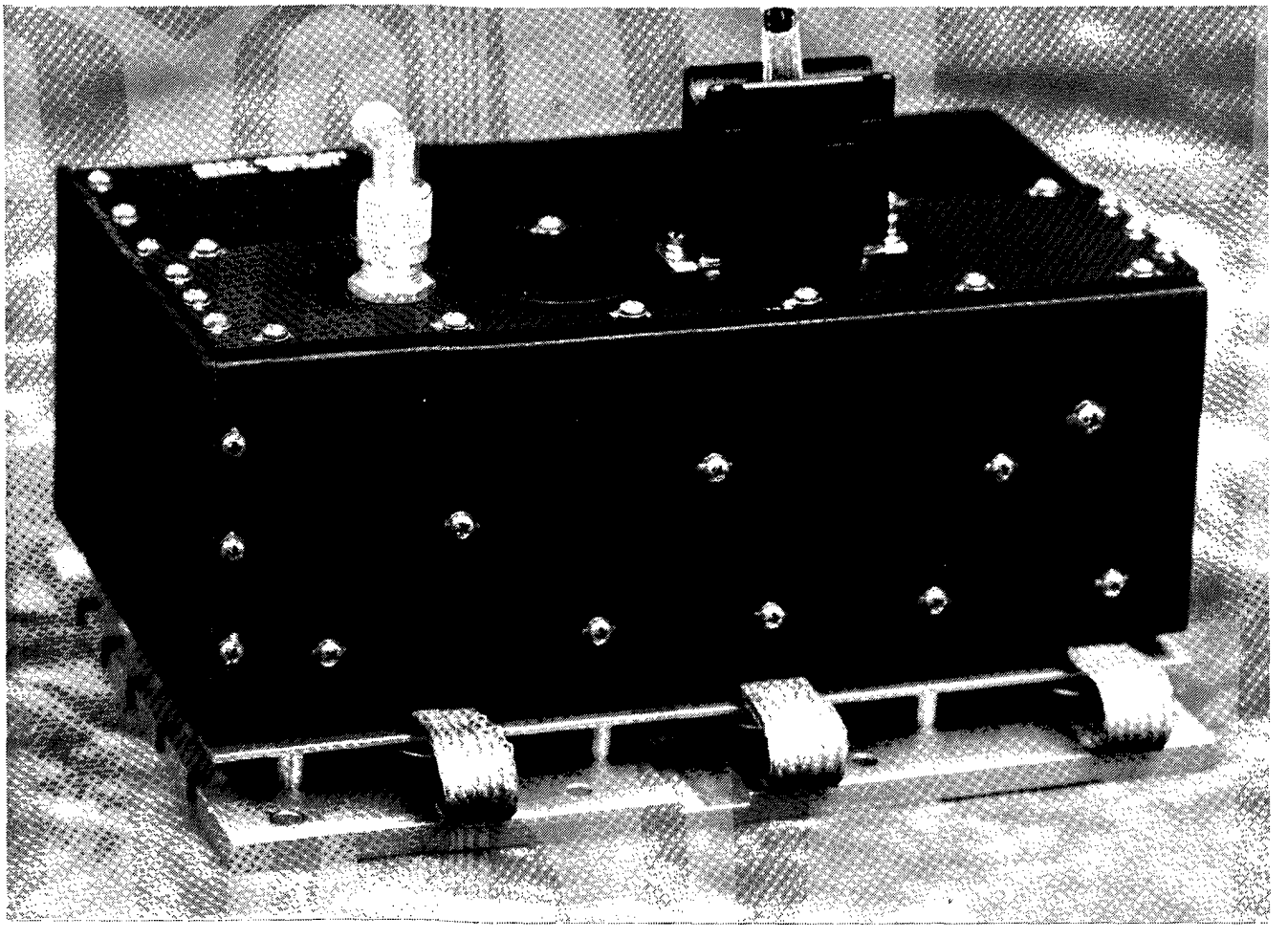


Figure 12. FTS Model 9521

Oven Temperature Control

The ideal thermal design for stabilizing the resonator temperature in a very high stability oscillator is one in which there is a high degree of isolation from environmental temperature changes. However, to survive the rigors of launch, a satellite oscillator must sacrifice some thermal isolation for rugged construction. In addition, it is useful for consistency of testing and verification of performance to have a thermal design which is not radically affected by the transition from atmospheric pressure to vacuum.

Thermal Isolation

We have chosen to use an evacuated glass Dewar to provide high thermal resistance to the ambient without requiring excessive thickness of surrounding insulation (in response to overall size constraints). The glass Dewar is surrounded by rigid foam, which also serves to capture a cylinder of close-fitting high permeability magnetic shielding. The oven thermal time constant is typically 300 minutes. Special construction of the Dewar makes it extremely stiff against vibration, so that even when loaded with the oven, it is calculated to have a fundamental vibration frequency for the inner glass cylinder above 5 kHz.

Oven temperature control is provided by a proportional heater circuit consisting of a thermistor bridge, an op amp gain stage and a power transistor bonded and mechanically attached to the oven. Because of the oven configuration, and its design for gradient-free temperature control of the resonator region, only a single heater is needed.

Additional thermistor sensors, at intermediate thermal distances between the oven and the ambient, can be used to enhance the dc oven response. [2] We have developed a very simple circuit configuration which accomplishes this. A drawback of this enhancement is that the time response for the compensator is intrinsically shorter than that for the primary oven control loop, and this requires additional circuit complexity to avoid unwanted transients. The obvious way to achieve the very long time constants required is by digital means. We have not found additional thermistor compensation to be necessary.

Oven Control and Effective Thermal Gain

Special effort is made in the design to provide a gradient-free local environment for the resonator. Oven thermal gain by the usual calculation of the product P (watt/degC) R (deg/watt) gives 1500, in a typical design considered here. In fact, the effective gain can be much larger than this. However, if the oven behaves strictly classically, G = 1500 would mean dT_o/dT_{amb} = 0.75 millideg/deg. For a resonator whose frequency characteristic is typically -4e-9 $(dT)^2$ at the turn-over point, this would mean a frequency stability of -2.4 e-13 per degree C ambient.

Figure 13 gives some results of oven testing where the stability is measured in oven degrees delta vs. ambient. The slope of the curve is the reciprocal of effective oven thermal gain.

Although oven thermal stability, even in a proportional-control design, can be nearly ideal in terms of effective thermal gain, in practice other frequency pushing effects limit the frequency-temperature behavior of the oscillator.

Oven Control Margin.

At low ambient temperature, the available oven power (equal to the oven portion of allowable warmup power) must be sufficient to maintain the differential between the crystal operating temperature and the ambient. In worst case, this may be impacted by degradation of the power transistor gain due to accumulated radiation dose on orbit. Dynamic gain is not affected since feedback from an emitter current-sense resistor maintains constant loop gain. At high ambient temperatures, control is lost when the oven heater shuts off. This occurs at a differential temperature which depends on overhead power dissipation from the oscillator circuits, and on the thermal resistance of the insulation.

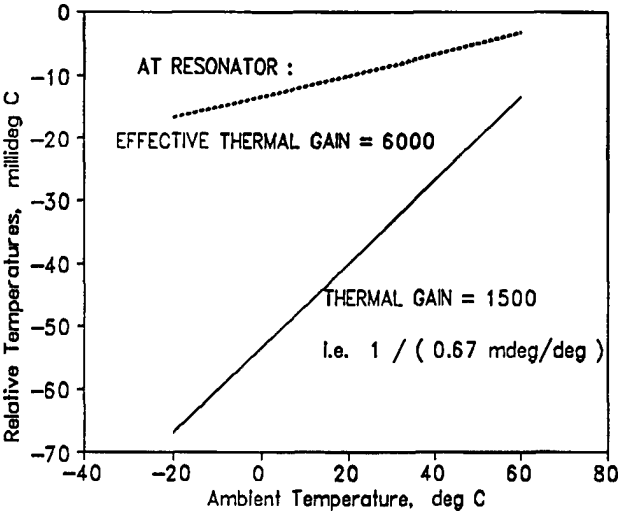


Figure 13. Oven Temperatures vs. Ambient

Remote Tuning

Resonator Frequency Aging

In the event that the frequency error budget falls outside the allowable frequency window over the mission lifetime, then the oscillator must be remotely tunable. The tuning range must encompass all expected frequency drift and offset effects including frequency retrace following storage and launch. This typically requires a tuning range on the order of 4e-7 with initial frequency set in the middle. The direction of aging drift is not predictable, even based on pre-launch measurements, and may change later. None of the parameters which drive the Error Budget are in fact deterministic as to sign. However, it might be noted that FTS experience with oscillators flown with the FTS 4401 GPS cesium clocks showed that the combined effects of frequency aging, even with radiation induced shifts, gave a net result of just a few e-12 per day aging, typically, for 5th overtone, AT-cut resonators.

Excessively large tuning ranges may compromise in-circuit crystal operating Q resulting in an overall loss of frequency stability performance; similarly, a large tuning range makes the oscillator more susceptible to voltage noise on the tuning signal. If, in addition, the tuning is implemented digitally, then additional constraints are involved in the specification of DAC performance.

Tuning may be implemented through either analog or digital voltage control. Noise and ground loops in the analog tuning can compromise the MO performance. The resolution required by a specific application dictates the number of digital tuning steps necessary to span the range. In order to most efficiently utilize the DAC range, it is useful for the oscillator tuning to be linear with voltage. The varactor tuning network is designed to afford an approximately linear reactance versus voltage characteristic. The non-linearities can be made as small as a few per cent over the full tuning range by properly tailoring component values for each specific resonator.

Digital Tuning Resolution & Temperature Sensitivity

Suppose the Error Budget dictates that the MO have a full scale tuning range of 4e-7. If a digital tuning resolution of 1e-10 is desired, then a minimum of 12 bits is required for the tuning DAC. The required temperature coefficient of the oscillator (2.5e-12/°C) dictates that the DAC temperature coefficient be less than 6 ppm/°C which necessitates that the DAC be provided with some form of thermal stabilization (inside the oscillator oven). To avoid undue perturbation of the resonator temperature stability, the DAC must have low power consumption, and not affect oven control. To avoid the heat leak of multiple conductors leading into the oscillator oven, a serial data format to the DAC is desired. An additional consideration for the DAC is that single-event-upset protection must be provided either within the DAC or in associated storage registers. Added to these requirements is the necessity to procure the DAC as a reliable space qualified component. These constraints combine to make the choice a narrow one.

DAC Candidates

Twelve candidate low power DACs were identified; various considerations narrowed the field to two devices. Total dose radiation testing of sample parts gave indications that one part is more robust, exhibiting negligible degradation after 20 krads exposure. To ensure flight integrity, total dose testing of samples from the production lot of flight DACs was required. These tests were successfully conducted in accordance with Method 1019 (including accelerated aging) of MIL-STD-883.

Special Testing and Analysis

In order to provide a very high probability of mission success, several engineering analyses are typically performed for satellite electronics. Electronic components are analyzed for applied stress with derating requirements in accordance with MIL-STD-975. A mechanical analysis is performed to ensure survival during the shock and vibration of launch. Survival of the crystal resonator is investigated with shock testing of sample devices. A thermal analysis is performed to verify the component thermal derating. A Worst Case Analysis is performed to ensure that the MO will perform over the mission life.

Shock and Vibration

The MO was successfully tested for shock survival in an operating condition. The applied shock level was 144 g. There was no discernible change in output frequency to a resolution of ±1e-10.

Random Vibration

The MO was successfully tested, without isolators, under random vibration exposure (20 Hz - 2 kHz; 19 g_{rms}) for 2 minutes per axis. The pre- to post-exposure frequency change was 2e-10.

EMI Test Results

The MO was successfully tested for compliance with the radiated emission requirements of MIL-STD-461, Method RE102. Sensitivity of the MO to ac magnetic fields was tested; the result was 1e-12/gauss. To be able to reliably detect levels of this magnitude, low frequency ac fields and phase sideband measurement techniques were employed.

Parts Program

There are many aspects of electronic components which are derived from the need for high reliability, long life and high performance in a demanding environment. Minimization of size and weight is also a driving factor. The choice of the DAC for the 9500, as discussed above, is illustrative of some of the considerations involved.

Launch survivability of the resonators depends on quartz enclosure design as well as workmanship. We have successfully qualified individual resonators using a 750 g, 300 microsecond controlled hammer blow shock test. Surface mount device technology is used throughout the design wherever possible, for size and weight savings, mechanical integrity and reliability.

Radiation Hardening

The subject of radiation hardening of precision quartz oscillators for satellite use can be cast into considerations of natural and man-made/weapons environments. Man-made/weapons environments will not be discussed in this paper.

Within the context of the natural radiation environment there are two aspects: 1) the degradation of electronic components as a function of total dose and single event upset and 2) the frequency effects of radiation.

Electronic Components

Modeling of the details of the specific satellite orbit over the satellite life can result in a radiation shielding curve which provides the mission total dose radiation exposure as a function of shielding thickness. Depending upon the location within the oscillator, different electronic components experience different total dose exposures. Geometric models of the oscillator are used to identify the total dose exposure of the individual components.

For the FTS 9500, components within the oven are protected by significant shielding. The oven housing, the magnetic shield and the MO external housing provide a shielding of 6 grams/cm^2.

Once the threat to the individual components has been identified, the degraded performance of the components is identified, usually from prior test data or from tests on sample components. These degraded performance parameters are used in a Worst Case Analysis to evaluate mission performance of the oscillator.

Quartz Resonator Effects

The effect of radiation on the frequency of the MO is driven primarily by the radiation sensitivity of the quartz resonator. This topic is complex, involving considerations of dose levels and rates, anneal intervals, prior irradiation history, etc. [3]

For total dose considerations, low earth orbit and geosynchronous orbit environments are rather different [4]. Low earth orbit satellites experience a dynamic radiation environment during the orbital period. In particular, for satellites which pass through the South Atlantic Anomaly, the environment can be defined as rather short (20-30 minutes) dose periods followed by essentially flux free periods of 70 to 100 minutes [5]. This alternating dose-anneal cycle leads to significant uncertainties in the evaluation of radiation effects.

Electronics inside satellites in geosynchronous orbit experience a continuous radiation exposure which accumulates to approximately 1 to 10 krads over a typical 10 year mission life [4].

Demanding MO specifications occur for low earth orbit satellites which require good frequency stability within the period of an orbit. The environmental sensitivities of the oscillator must be low enough to sustain the frequency stability specifications in the presence of a changing orbital environment: changing radiation flux environment, changing temperature, etc.

Further, the alternating dose/anneal cycles of low earth orbits present a complex scenario which makes difficult the determination of specific values for quartz sensitivity for use in analysis. Radiation sensitivity is non-linear and is greatest at low accumulated doses. The details of low dose rate sensitivity in the presence of intervals of anneal periods is not well understood. Low dose effects are process and manufacturing lot dependent. High dose effects are material dependent [6]. In addition, there is wide variability within a lot, for low dose sensitivity. There is evidence that pre-irradiation (preconditioning) is advantageous in reducing the frequency sensitivity of quartz resonators [6]. For

critical applications the prudent approach is to use premium Q air swept synthetic quartz, precondition the resonators, characterize them individually at dose rates comparable to those expected on-orbit and incorporate design margins.

Radiation Testing and Results

Concerns about radiation effects on frequency stability cover a wide range. Of particular concern for this class of MOs are low level radiation and low dose rates in low earth orbit.

Preconditioning

To insure the dynamic radiation performance of the 4 MHz oscillator, resonators were first preconditioned with Cobalt 60 radiation. The resonators were then individually screened for low dose rate radiation sensitivity. Three lots of resonators were procured to ensure availability of suitable resonators. All resonators tested exhibited sensitivities of less than 1e-10/rad; several resonators exhibited sensitivities of less than 1e-11/rad.

Test Methods

Screening for minimal radiation sensitivity at the 1e-11 response level requires that the resonator be in a stable, operating oscillator test-bed. We have performed such screening using FTS commercial ovenized oscillators in which the crystals are mounted to permit exposing the crystal, but not the oscillator electronics, to incident radiation.

Frequency measurement is performed by mixing the MO output with an offset reference frequency, giving a beat frequency which is period-averaged to reduce short-term noise. To track absolute frequency changes, and possible changes in aging behavior, the reference is a cesium frequency standard, translated via synthesizer. The offset (heterodyne factor) and period averaging can be adjusted to give 1e-12 sensitivity and a 5e-13 reference noise floor, typically.

Results

Shown in Figure 14 are some results collected during short exposures with 4 rad accumulation. The resonators (from a given lot) exhibit a variety of signatures [6]. In general one cannot expect uniformity of signature type to be a guarantee of subsequent predictability of response; however, a small sensitivity observed in the screening can be taken as a predictor of good behavior during the flight mission.

During the pre-irradiation stabilization period for some radiation tests, it was apparent that some frequency wander could be correlated with barometric pressure variations, as has been reported [7]. It is no surprise that atmospheric pressure changes cause frequency shifts, since the gross air-to-vacuum frequency shift of resonators is typically a few to many parts per billion, for a wide variety of resonator packages. If the pressure shift were linear, which is probably not the case, then the typical 5% to 10% fluctuations of atmospheric pressure would be expected to produce frequency fluctuations of at least several e-11 up to several e-10, which we, in fact, have observed. This contribution to frequency wander is of course absent in the vacuum of space. An approach to minimizing pressure and humidity effects is to seal the oscillator.

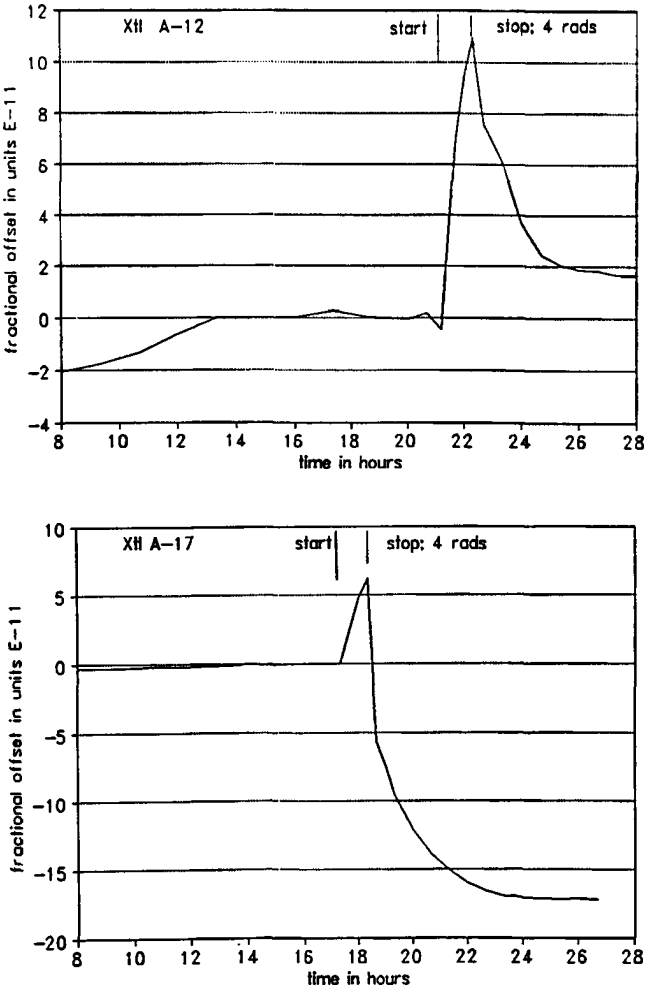


Figure 14. Low-Dose Radiation Sensitivity Testing

Summary

Design considerations and performance results have been presented for MOs at 4 MHz and 10 MHz, developed for satellite applications requiring very good frequency stability. Design tradeoffs have been discussed within the context of a frequency error budget, particularly for low earth orbit applications. Some results of radiation screening of 4 MHz resonators have been presented. The requirement for high resolution remote digital tuning has the adverse potential for compromising performance. However, test data for digitally tuned 4 MHz MOs have shown very good performance.

Particular attention has been paid to very long term testing of frequency stability, and to behavior under periodically fluctuating temperature conditions which are expected on orbit. For 10 ^{o}C peak-to-peak temperature ambient in vacuum, the measured Allan deviation for a 4 MHz MO was 5e-12 at tau = 50 minutes (the half-period of the sinusoidal temperature disturbance), and this is consistent with a dynamic temperature coefficient of frequency response less than 1 e-12 per degree. Summary performance results are shown in Table II.

Table II. Performance Summary

Parameter		Units	4 MHz	10 MHz
Short Term Stability	1 sec	e-13	6	6
	10 sec		4	4
	100 sec		6	6
	1000 sec		4	7
Phase Noise	1 Hz	dBc / Hz	-127	-118
(single source)	10 Hz		-143	-141
	1 kHz		-163	-163
	100 kHz		-165	-165
Frequency Stability vs. T		e-12 / deg C	2.3	1.5
Aging per day		e-11 / day	1 to 5	3 to 6
Delta f (load)		e-11 /10%	< 1	0.3
Delta f (Supply V)		e-13 / 2%	8	6
Delta f (acceleration)		e-9 / g	1.2	1
Delta f (Magnetic field)		e-12 / Gauss	1	
Output variation over T		+/- dB / T range	< 0.2	0.4

Acknowledgment

We would like to thank David Swant for his help in making radiation test measurements, and Peter Vlitas for contributions to the electronic design.

References

[1] The PC-based analysis program "Stable", Hamilton Technologies, Hamilton, MA, is a valuable tool.

[2] F. L. Walls, "Analysis of high performance compensated thermal enclosures," Proc. 41st Annual Symp. Freq. Control, pp. 439-443,1987.

[3] J. R. Norton, "Performance of Ultrastable Quartz Oscillators Using BVA Resonators", Proceedings 8th European Frequency and Time Forum, pp. 457-465, March 1994.

[4] J. Bourrieau, et al, "A Semi-Quantitative Model of the Dynamic Behaviour of Space USOs under Fragmented and Continuous Irradiations," Proceedings 7th European Time and Frequency Forum, pp. 269-278, March 1993.

[5] R. H. Maurer and J. J. Suter, "Total Dose Hardness Assurance for Low Earth Orbit," IEEE Trans. Nuc. Sci., Vol NS-34, p. 1757, Dec 1987.

[6] J. R. Norton, J. M. Cloeren and J. J. Suter, "Results from Gamma Ray and Proton Beam Radiation Testing of Quartz Resonators" Proceedings 38th Ann. Frequency Control Symposium, pp. 63-72, 1984.

[7] J. R. Norton, "Ultrastable Quartz Oscillator for Spacecraft" Proceedings 21st Annual PTTI, pp. 509-518, 1989.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ADVANCED TECHNOLOGY OSCILLATOR FOR SMALL SPACECRAFT

Jerry R. Norton
The Johns Hopkins University Applied Physics Laboratory
Johns Hopkins Road
Laurel, Maryland 20723-6099
Telephone 410-792-5000
Fax 301-953-5883
E-Mail: Jerry_Norton@jhuapl.edu

ABSTRACT

A small, low-mass, ultrastable oscillator with excellent frequency stability (1×10^{-13} at 100 s) is being developed for use in small spacecraft. The very nature of a small spacecraft places a very high premium on mass, size, and power of all instruments. Therefore, the primary task of this design was to reduce by 50% the mass and size of a flight-proven ultrastable oscillator design without seriously degrading the oscillator's frequency stability or immunity to environmental stress—a very challenging design problem. The mass of key components of an oscillator fabricated to test new design ideas was reduced by over 78%. A series of tests, including vibration testing, met or exceeded our expectations. On the basis of the results of the vibration tests, a vibration isolation system will not be used for this oscillator. A second thrust of this development is to optimize the oscillator electronics. Phase noise has been improved by 6 dBc to −159 dBc 1 kHz from the carrier. An oscillator configured to permit easy circuit changes is being used for this work.

With the data and knowledge gained from this development, the probability of fulfilling the design goals are very high.

INTRODUCTION

An oscillator with a very stable output frequency is required for the radio science experiment aboard the Pluto Flyby mission for which this oscillator development was initiated. The science objective of this experiment is to measure the pressure and temperature structure in the lower atmosphere of Pluto. Change in the radio propagation path length as the signal passes through Pluto's atmosphere must be measured to an accuracy of ≈100 μm to determine pressure and temperature of the atmosphere. A very stable reference oscillator is required because the accuracy of path length measurement is directly related to frequency stability. It is equally important to minimize mass, size, and input power because every kilogram added to the Pluto spacecraft increases cost by about $250,000 and adds one week to the flight time.[1] The primary objective for this exploratory design was to demonstrate that a quartz oscillator with greatly reduced mass and size could be realized while retaining excellent frequency stability and immunity to environmental effects. Two tasks were critical to realizing this goal: (1) to reduce the mass and size of the oscillator resonator oven assembly and (2) to demonstrate that the oscillator could survive the launch environment without a vibration isolation system. The Applied Physics Laboratory (APL) has a rich heritage of building small, highly reliable oscillators with excellent frequency stability; even so, to reduce the mass and size of an oscillator by 50% and retain high performance is an extremely challenging design problem.

Our experience has shown that the quartz resonator dominates oscillator frequency stability in a well-designed oscillator.[2] The tactical BVA resonator was chosen for this design because it can provide very good frequency stability and its mechanical resonance is ≈2000 Hz, the upper limit of the Pluto vibration specification.[3] The resonator is a 10-MHz, 3rd-overtone, stress-compensated (SC) cut device housed in an HC-40 enclosure.

RESONATOR OVEN DESIGN

The resonator oven and its design is the most critical mechanical assembly in terms of oscillator performance. The resonator oven assembly, which includes the quartz resonator, printed circuit boards for the oscillator and the

0-7803-2500-1/95 $4.00 © 1995 IEEE

oven controller, and the oven housing, establishes the minimum mass and size of the entire oscillator. The sizes of the mechanical enclosures, the supporting assemblies, and the insulating system are directly related to the size of the oven. The mass of the oscillator is likewise directly related to the size for a given material. The power required to maintain the oven at a given temperature is a function of surface area, which is again related to size. Ideal oscillator ovens should be fabricated from a material that has infinite thermal conductivity and thermal capacity. Unfortunately, materials that approach these characteristics are typically very dense. Therefore, a compromise must be made between mass and thermal properties. An aluminum alloy was selected for the resonator oven because it has relatively high thermal conductivity and relatively low mass. This oven is smaller in size than the ovens used in previous APL oscillators. The internal structure of the oven was also modified to reduce its mass. The design changes implemented for the oven reduced its mass to 28.6 g (78% less than the present oscillator oven housing). The frequency vs. temperature performance of the Pluto oscillator is excellent, with little or no degradation resulting from the oven design changes. The results from oscillator short-term frequency stability tests are not as clear.

Short-term stability 2 to 3 times better than the Pluto oscillator has been measured in APL oscillators using tactical BVA resonators. The difference in frequency stability most likely is due to the quartz resonator used in the Pluto oscillator and not the oven design. Detailed oscillator performance data are presented in a later section of this paper.

Magnesium has a density 33% less than aluminum and would produce a comparable mass reduction, but its thermal conductivity is 65% less than that of aluminum. With a conflict between these parameters the overall effect on the performance of an oscillator using a magnesium oven is not clear, but additional tests using a magnesium oven should be pursued in view of the extremely critical mass requirement for the Pluto spacecraft. If performance of a magnesium oven is not satisfactory, a composite oven should be investigated. A composite oven could also provide ionizing radiation shielding. Radiation effects on quartz oscillators are discussed later in this paper.

OSCILLATOR CHASSIS DESIGN AND VIBRATION TESTING

The oscillator housing is the second largest contributor to total oscillator mass. Again, minimizing size is the key to reducing mass. All previous APL oscillators required vibration isolation systems to prevent damage to the quartz resonator during launch. These isolation systems increase the mass of the housing in two ways: (1) space inside the housing must be provided to allow movement of the housing around the isolated resonator oven assembly; and (2) the housing must have thick, heavy internal structures and exterior walls to provide stiff, rigid supports for the vibration isolation system.

A tactical BVA quartz resonator that has been under development for several years has become commercially available.[4] Its specifications indicate it should survive the Pluto launch environment. A test oscillator with all components hard-mounted to the basic structure was built, and a series of tests was conducted in the APL Vibration Facility. The critical vibration-sensitive internal assemblies of the test oscillator, including the resonator, were similar to those that would be used for a flight oscillator. The rest of the test oscillator was assembled for convenience and low cost, bearing little resemblance to a flight oscillator. The oscillator was tested in two axes to the vibration levels expected during the Pluto launch, which are relatively high, as shown in Fig. 1. The vibration tests were conducted in the x and y axes, oriented relative to the quartz resonator as shown by the inset in Fig. 1a. The oscillator was powered and operated during the vibration tests. The 10-MHz output signal from the oscillator, which was monitored throughout the vibration test, was not interrupted. The oscillator was not damaged by the vibration tests and performance was not degraded after the test. On the basis of the results of this vibration test, a vibration isolation system will not be used for this oscillator. Eliminating the vibration isolator will greatly reduced the size and mass of the oscillator housing. Minimizing wall thickness in keeping with structural integrity also reduces mass. The estimated mass of a flight-qualified oscillator is only 320 g and its volume is 353 cm^3.

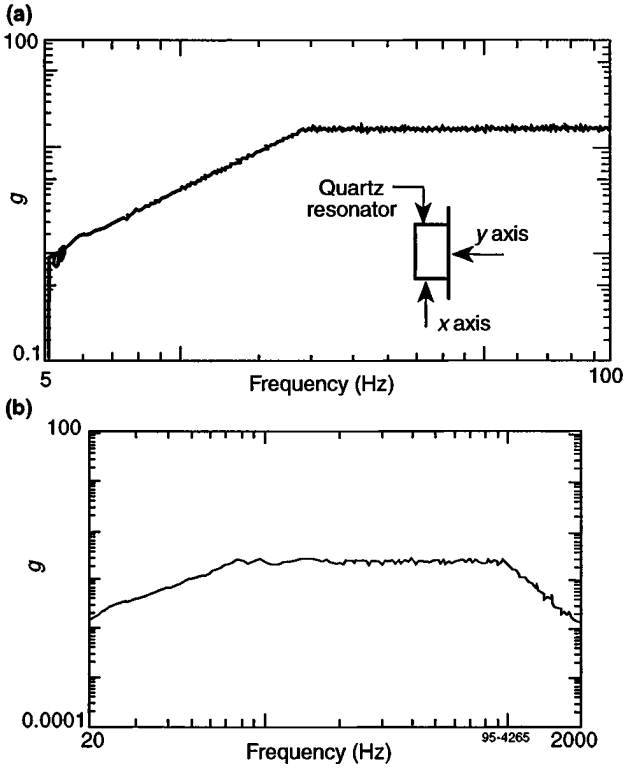


Figure 1. The levels of vibration expected during the Pluto launch. (a) Sine wave input. Inset: the vibration test orientation of the x and y axes relative to the quartz resonator. (b) Random input.

FREQUENCY STABILITY AND PHASE NOISE PERFORMANCE

To measure the change in the radio propagation path length to ≈100 μm, the onboard portion of the radio science instrument can be thought of as primarily a phase-measuring device with a measurement accuracy of about 0.01 rad. This corresponds to a frequency stability requirement of 1.8×10^{-13} for the onboard oscillator.[5] The design goal for this oscillator was 1×10^{-13} for a 100-s measurement period.

The measured Allan variance for the Pluto test oscillator was only 4.8×10^{-13}, as shown in Fig. 2. From the character and structure of the oscillator output frequency, the most likely reason for this relatively poor Allan variance is the quartz resonator. Allan variance data from the optimization oscillator using a different resonator are also shown in Fig. 2. The 100-s Allan variance for this oscillator is 1.0×10^{-13}, and it has extended periods of operation when the Allan variance is $< 1.0 \times 10^{-13}$. This is our first 10-MHz oscillator to achieve performance below 10^{-13}. While the number of tactical BVA resonators evaluated at APL have been limited, more than 35% of the units tested had an Allan variance of $< 2 \times 10^{-13}$. Resonator selection will be required to achieve an Allan variance of 1 to 2×10^{-13}, but this goal is achievable. Resonator selection was not possible for the Pluto test. The Allan variance data presented do not assume equal noise sources.

Phase noise ($\mathcal{L}_f$) data for both the Pluto oscillator and the optimization oscillator are presented in Table 1. Phase noise performance for the Pluto oscillator achieved most of the design goals. An objective for optimizing the oscillator electronics is to improve the phase noise performance. An improvement of ≈6 dB has been achieved (Table 1) with all design goals realized, some with a very comfortable margin. Circuit optimization work is continuing with a goal to improve both the Allan variance and the phase noise performance.

The 24-h aging rate for the Pluto oscillator was -6.1×10^{-13} 150 days after it was turned on, which is excellent for a quartz oscillator and approaches the stability of small atomic rubidium frequency standards.[6] Frequency retrace and the time required to establish an aging rate can be important parameters in some applications. The output frequency as a function of time for the Pluto test oscillator is presented in Fig. 3. The oscillator was operated in air at atmospheric pressure for the first 11 days. Changes to the oscillator were being made and various tests were being conducted, so the aging rate was not well established. On day 11 the oscillator was moved to a vacuum chamber and evacuation was initiated. A rapid frequency increase of $\approx 9 \times 10^{-10}$ occurred during the first day. During the next 20 days frequency decreased and established an aging rate of -3×10^{-12}/24 h. On day 46 the oscillator was turned off for 2 days; then power was turned back on. A frequency change or retrace to the frequency prior to turnoff was $\approx -2 \times 10^{-10}$. An aging rate of -3×10^{-12}/24 h was reestablished within 2 to 3 days after the oscillator was restarted. Three other retrace tests were conducted: on day 80 the oscillator was turned off for 4 days, on day 150 the oscillator was turned off for 7 days, and on day 278 the oscillator was turned off for 1 day. In each retrace test the recovery period was different, as shown in Fig 3. Both the retrace frequency offset and the time required to reestablish an aging rate can be important to small spacecraft missions where power is a premium quantity and it may not be possible to keep the oscillator continuously powered. On day 348 the oscillator was vented to atmospheric pressure, which produced a frequency decrease of $\approx 8 \times 10^{-10}$, which is very close to the frequency change that occurred during evacuation but with the opposite sign. The aging rate of the oscillator just prior

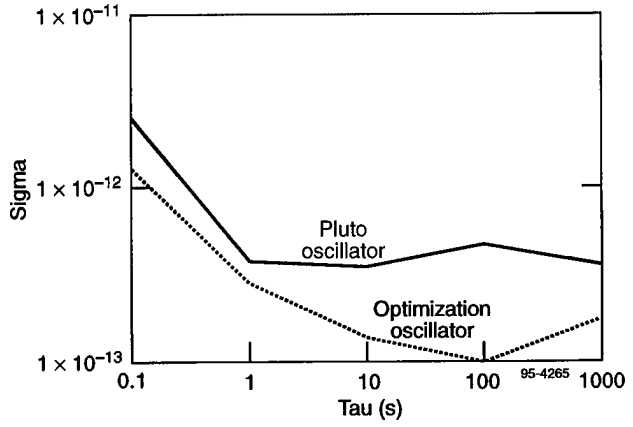


Figure 2. Allan variance data for the 10-MHz ultrastable oscillators. Measured data do not assume equal noise sources. Measurement bandwidth, 1000 Hz.

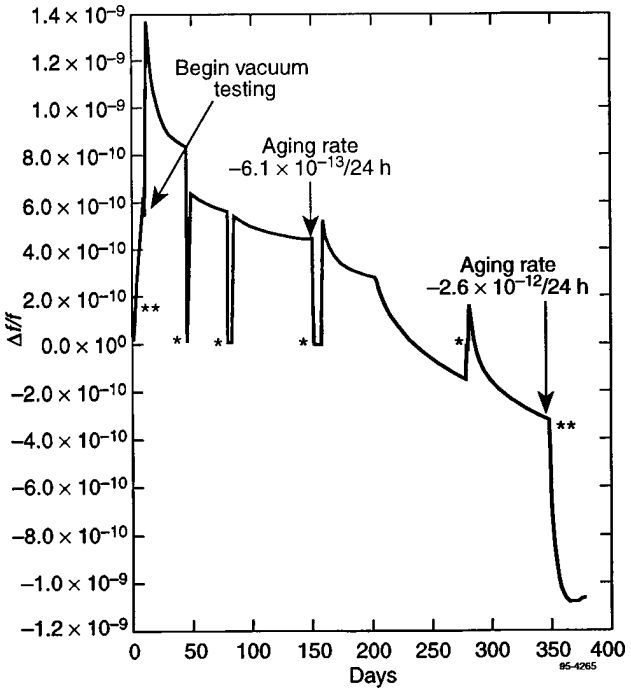


Figure 3. Output frequency of the Pluto oscillator as a function of time, showing aging rate and frequency retrace. * Power off for retrace test. ** Atmospheric pressure.

to venting was -2.6×10^{-12}/24 h. One other event shown in Fig. 3 is noteworthy. On day 204 a change in aging rate occurred. There is no known reason for this change: operating parameters were constant and there were no physical changes.

Table 1 presents performance data for both the Pluto and the optimization oscillators.

ENVIRONMENTAL EFFECTS ON FREQUENCY STABILITY

Temperature

Changes in the ambient temperature and environment in which the oscillator is operating will cause frequency changes. Oscillator design and operating temperature range are related to oscillator power consumption. Output frequency and input power consumption for the Pluto oscillator are presented in Fig. 4. Data for three configurations are presented. The amount of insulation or the thermal resistance the insulation provides directly affects oscillator performance. Configuration 1 had about half the insulation normally used in APL oscillators so that the oscillator subassemblies could be hard-mounted for vibration testing. Input voltage for this test was 19.5 V. Configuration 2 had the quantity of insulation normally used in APL oscillators, and the input voltage was 19.5 V. This configuration has only one data point at 28°C. Configuration 3 is the same as configuration 2 except the input voltage was reduced to 17 V. These data provide insight to the trade-offs that can be made in the design of an oscillator and how they relate to operating conditions, specified temperature range, and input power.

Table 1. Performance data for 10-MHz ultrastable oscillators.

Performance Parameter	Goals	Pluto Oscillator	Optimization Oscillator
Aging rate/24 h	2.0×10^{-11}	-2.6×10^{-12}	4.7×10^{-11}
Allan variance, tau (s)			
0.1	1.0×10^{-12}	2.5×10^{-12}	1.3×10^{-12}
1	3.0×10^{-13}	3.8×10^{-13}	2.9×10^{-13}
10	1.0×10^{-13}	3.6×10^{-13}	1.4×10^{-13}
100	1.0×10^{-13}	4.8×10^{-13}	1.0×10^{-13}
1000	2.0×10^{-13}	3.7×10^{-13}	1.8×10^{-13}
Phase noise			
Frequency offset (Hz)			
0.1	—	−88 dBc	−90 dBc
1	−120 dBc	−116 dBc	−120 dBc
10	−135 dBc	−140 dBc	−148 dBc
100	−145 dBc	−150 dBc	−155 dBc
1000	−150 dBc	−153 dBc	−159 dBc
10000	−150 dBc	−155 dBc	−159 dBc
Frequency as a function of:			
Temperature per °C (20° to 40°C)	1.0×10^{-12}	-2.0×10^{-13}	N.T.†
Load (50 Ω ± 10%)	2.0×10^{-12}	2.0×10^{-12}	$<2.0 \times 10^{-12}$
Input voltage	1.0×10^{-12}	1.0×10^{-12}	1.0×10^{-12}
Radiation	1.0×10^{-10} rad*	N.T.†	N.T.†
Magnetic susceptibility	2.0×10^{-12}/G	1×10^{-12} to 2×10^{-11}*	N.T.†
Acceleration	1.5×10^{-12}/g	1.0×10^{-10}/g	N.T.†
Vibration	5 – 2000 Hz	Survived*	N.T.†
Output characteristics			
Power level	0 dBm	7.4 dBm	5.6 dBm
Harmonic	−50 dBm	−52 dBm	−66 dBm
Spurious	−60 dBm	−86 dBm	−80 dBm
Power at 25°C (W)	0.8*	0.77*	
Mass (g)	320 (est.)		
Size (cm)	9.65 × 6.86 × 5.33 (est.)		

*See text.
†N.T., not tested.

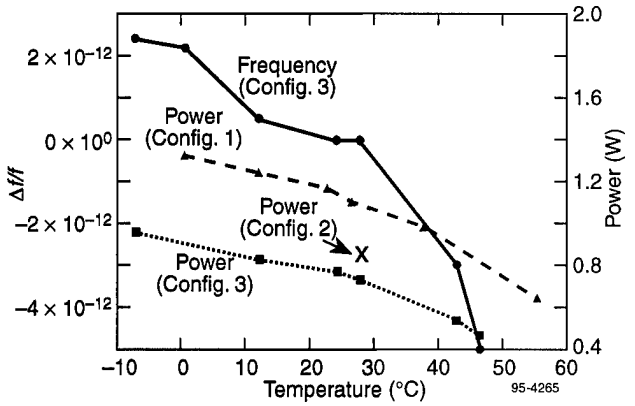


Figure 4. Output frequency and input power consumption as a function of temperature for three configurations of the Pluto oscillator.

The quartz resonator must be operated precisely at its turning point temperature, which is ≈85°C. The thermal control circuitry requires a temperature differential between the ambient temperature and the set turning point temperature to maintain control of the turning point temperature. The magnitude of the temperature differential depends on several design parameters, but the primary ones are the quantity of insulation surrounding the resonator oven and the required upper operating temperature. Input power requirements are different for the three configurations, as clearly shown in Fig. 4. Configuration 2 requires 24% less power than configuration 1, and configuration 3 requires 33% less than configuration 1. Recall that the only difference between configuration 1 and 2 is the quantity of insulation. Maintenance of a given oven temperature at a given ambient temperature requires a constant input power to the oven; additional power is dissipated in the control circuits and in the transistor controlling current to the oven heater. If the input voltage is reduced, the oven power remains constant, but the power dissipated in the control transistor decreases. This explains the power difference between configurations 2 and 3.

Another method of changing the power balance between the oven and the control transistor is to change the resistance of the heater. High ambient operating temperature is limited by and directly related to the quantity of insulation. Beyond this limit, control of the oven temperature is not maintained. The high-temperature limit for configuration 3 was ≈49°C, compared with ≈62°C for configuration 1. Again, the quantity of insulation is the primary cause for the difference. As the above data indicate, design trade-off options frequently conflict with each other, so intelligent compromises must be made, which include realistic oscillator specifications that do not exceed the real science requirements.

The dynamics of the thermal time constant are heavily influenced by the quantity of insulation, as shown in Fig. 5. The longer time constant achieved with more insulation (configuration 3) will delay or sometimes prevent the full magnitude of short-term ambient temperature changes from

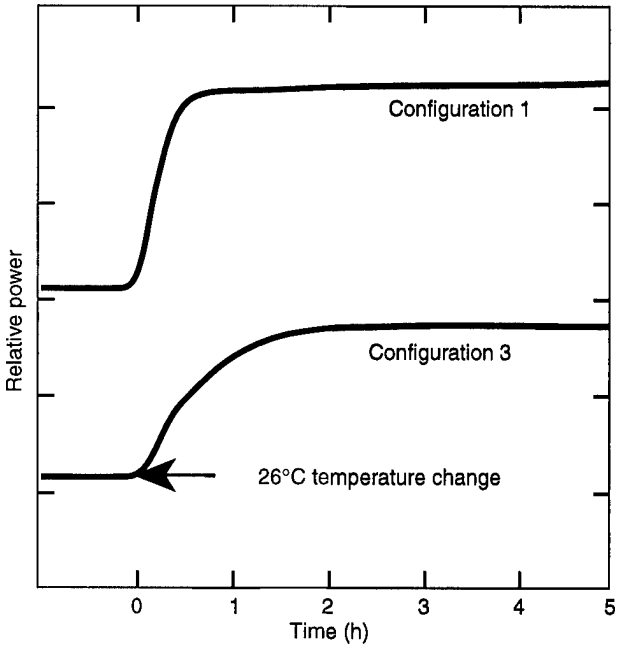


Figure 5. Influence of insulation quantity on the dynamics of the thermal time constant.

reaching the resonator oven. If the magnitude of the change in ambient temperature is large, the rate of change inside the insulation layers at the oven will be slower, allowing the control circuits to respond more accurately and thereby minimizing oven temperature change.

The Pluto oscillator has a very low temperature coefficient of 2×10^{-13}/°C between 20 and 40°C. The frequency change as a function of temperature from −10 to +50°C is shown in Fig. 4. These data are for the oscillator in configuration 3. Even small, rapid temperature changes can have a detrimental effect on oscillator output frequency. Therefore, in addition to the static temperature test in which a large step in ambient temperature was made and the oscillator output frequency change measured after stabilization, a dynamic temperature test was also conducted. The test temperature was changed in 10°C increments at ≈30-min intervals. With an oscillator thermal time constant of 3 to 4 h, the temperature control circuits and output frequency did not have enough time to stabilize before the test temperature was reversed. This simulates a real-world condition that often occurs in spacecraft, particularly small spacecraft, as they pass in and out of sunlight. The results of this test are shown in Fig. 6. The peak-to-peak frequency change is $\approx 3 \times 10^{-12}$.

Acceleration

A quartz oscillator's output frequency is sensitive to position changes in the gravitational field and to acceleration forces. Usually this sensitivity is not a problem after the spacecraft is placed in orbit, but if thrusting maneuvers are conducted by the spacecraft, acceleration sensitivity may be a design consideration. BVA resonators are generally less sensitive to acceleration than conventionally mounted reso-

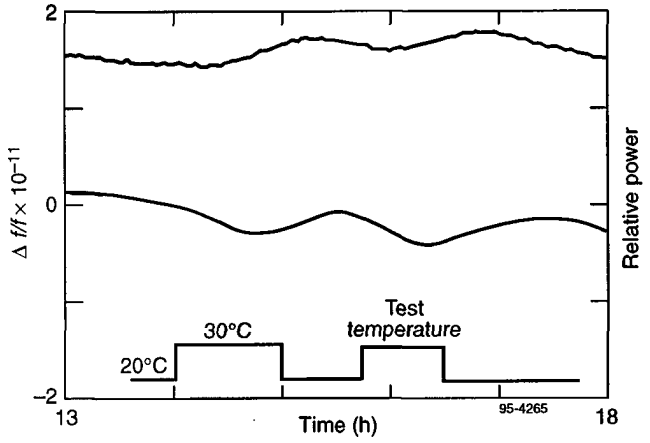


Figure 6. Results of the dynamic temperature test simulating the conditions of a small spacecraft passing in and out of sunlight.

nators.[3] The Pluto oscillator was tested by rotating the oscillator around both axes while observing the change in output frequency. This is called a 2-*g* tip-over test. A maximum frequency change measured for the Pluto oscillator was $1 \times 10^{-10}/g$. The *g* sensitivity is not linear and is different for each axis. In applications with known directional acceleration forces, the oscillator could be oriented in the plane with the minimum acceleration sensitivity.

Magnetic Sensitivity

Both the oscillator circuits and the tactical BVA resonator are sensitive to magnetic field changes. The largest magnetically induced frequency change measured for the Pluto oscillator was 2×10^{-11} for a 1-G change measured in a DC field. Again, the magnetic sensitivity of the oscillator is not symmetrical and depends on polarity. The tactical BVA resonator is less sensitive ($\approx 2.7 \times 10^{-12}$/G) than the oscillator circuits.[3] A low-mass magnetic shield is routinely incorporated into the oscillator if the magnetic field in the spacecraft varies. This shield reduces the magnetic sensitivity to $\approx 1 \times 10^{-12}$/G.

Radiation

Interest in the effects of ionizing radiation on quartz resonators and oscillators has been intense at APL since the late 1960s.[7] With improvements in oscillator performance and reductions in sensitivity to other environmental stresses, the radiation-induced frequency change is the largest and least predictable environmental effect in many orbits. The radiation environment for the Pluto mission has not been well defined; therefore, radiation tests, which are time consuming and expensive, were not conducted on the Pluto oscillator. The Pluto oscillator configuration has little radiation shielding. If radiation shielding becomes necessary to meet mission frequency stability requirements, the mass of the oscillator will increase because effective shielding materials are usually heavy. BVA-type resonators are generally less sensitive to radiation than conventionally mounted resonators. Radiation tests on tactical BVA resonators have been conducted with a spread of 2×10^{-10} to 2×10^{-11}/rad(Si) for a 0.6-rad exposure.[8]

CONCLUSION

The knowledge and data gained from this advanced technology oscillator development provide a solid basis for development of a small, ultrastable quartz oscillator for use in small spacecraft applications. The effort to reduce mass and size met or exceeded our expectations. We firmly believe we can meet the design goals for mass and size, environmentally induced frequency changes, and power requirements. The design goals for frequency stability are more difficult, but with circuit optimization and resonator selection these goals are believed to be within reach. Funding limitations prevented exploration of aspects of the oscillator design that are important but would be of little interest if the primary objective to reduce mass and size could not be achieved. By addressing the tasks listed above, a small, low-mass, flight-qualified oscillator with outstanding performance can be developed.

ACKNOWLEDGMENT

This work was supported by a contract from Stanford University. I am grateful to many colleagues, especially J. M. Cloeren, F. L. Walls, and R. J. Besson, for valuable discussions and contributions to this oscillator development.

REFERENCES

[1] R. L. Staehle et al., "Pluto Technology Challenge Team Review," Jet Propulsion Laboratory, June 1994.

[2] J. R. Norton, "Ultrastable Quartz Oscillator for Spacecraft," in *Proc. 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 1989, pp. 509–518.

[3] J. R. Norton and R. J. Besson, "Tactical BVA Quartz Resonator Performance," in *Proc. 1993 IEEE Frequency Control Symp.*, 1993, pp. 609–613.

[4] R. J. Besson, "A New Electrodeless Resonator Design," in *Proc. 31st Annual Symp. on Frequency Control*, 1977, pp. 147–152.

[5] G. L. Tyler et al., "Definition of an Uplink Radio Science Instrument to Measure Pluto's Atmosphere from the PFF Platform," Proposal to NASA in response to NRA 93-OSSA-5, submitted by Stanford Univerisity, Stanford, CA, 1993.

[6] J. R. Norton and J. M. Cloeren, "Precision Quartz Oscillators and Their Use Aboard Satellites," *Johns Hopkins APL Tech. Dig.*, Vol. 15(1), pp. 30–37, 1994.

[7] J. R. Norton, J. M. Cloeren, and J. J. Suter, "Results from Gamma Ray and Proton Beam Radiation Testing on Quartz Resonators," *IEEE Trans. Nuclear Science*, Vol. NS-31, pp. 1230–1233, 1984.

[8] J. R. Norton, "Performance of Ultrastable Quartz Oscillators Using BVA Resonators," in *Proc. Eighth European Frequency and Time Forum*, in publication, 1994.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

An Experimental Study of Long Term Aging of Quartz Oscillators

MITSUAKI KOYAMA,
Nihon Dempa Kogyo co. ltd., Sayama city, Saitama 350-13, Japan

YASUAKI WATANABE, HITOSHI SEKIMOTO and YOSHIMASA OOMURA
Tokyo Metropolitan University, Hachioji city, Tokyo 192-03, Japan

Abstract - Temperature-compensated or ovenized crystal oscillators (TCXO or OCXO) are widely used in self-contained equipment, especially in cases when a highly-stabilized frequency is needed. In these oscillators, AT-cut crystal resonators are generally employed because of the ease of resonator design and frequency controllability. Variation of oscillation frequency with time is called "aging", and is caused by the aging of the resonator itself. When the variation is smooth, regular calibration is effective. There are, however, many cases when the variation evidences a "jump." While many causes may be considered for this, the aging of a nearby mode is regarded as the leading cause when the frequency jumps suddenly. To avoid this nearby mode influence, which decreases the long term frequency stability, detection and avoidance of the mode must precede oscillator assembly. If a cause of aging in the spurious modes is also relaxation of lattice defects in a quartz plate or electrodes, then the phenomenon is related to integration of the provided electrical power.

Based upon this background the relation between the integrating drive level and the drive level-frequency dependence (DLD) of the nearby mode was examined for AT-cut quartz resonators used in high stability OCXOs. The authors had previously developed a precise measurement system for DLD characteristics. We have also pointed out that the frequency jump appears to depend on the resonator drive level. We used some quartz resonators which exhibited DLD frequency jumps. These are caused by the nearby mode. The experimental investigation included; (1) the resonator was vibrated by a constant driving current at regular intervals: (2) after one interval, the frequency-jump (current) was searched by DLD characteristic measurements. By continuing this sequence, we specified a relation between the drive level integration (or elapsed time) vs. the "frequency jump current". The results clearly indicated that the "frequency jump current" level increases as a function of drive current integration. We concluded that the drive level characteristics of nearby modes varied with the total (integrated) power provided to the resonator. The oscillation frequency then jumps suddenly.

Introduction

A great number of crystal oscillators are used in communications equipment, measuring instruments, computers, audio-visual equipment etc. When that high frequency stability of 10 ppb/day or more is required, oven controlled crystal oscillators(OCXOs) are generally used. In particular, OCXOs with AT-cut crystal units are widely used because of their low cost and ease of frequency adjustment.

One important measure of OCXO performance is long-term frequency stability, i.e. aging characteristics. The aging characteristics of an OCXO are determined by those of the crystal unit used. With the passage of time, the resonance frequency of the crystal unit changes gradually. But the aging characteristics of an OCXO sometimes show an abrupt jump in frequency, followed by quick recovery[1]. In most cases the frequency change is small, ranging from 1 ppb to 10 ppb, but the frequency jump great degrades the reliability of the devices making them unsuitable for critical applications such as equipment in airplanes and in communication networks and in the devices requiring high frequency stability. The frequency jump is thought to be caused by a change in oven temperature, the influence of a spurious resonance with step temperature characteristics, or change in oscillation circuit parameters.

On the other hand, we have shown that the frequency instability of an OCXO with the ambient temperature change is caused by nearby weak spurious resonances due to the current change in a crystal unit, and have developed a system which allows minute resonances to be detected through the current-resonance oscillation change[2]. With the above-mentioned aging characteristics, the oscillation frequency changes abruptly, which is considered to be caused by weak spurious resonances near the fundamental mode of the crystal unit used. But highly accurate measurement of these weak spurious resonances has been impossible due to the lack of a sufficiently accurate measuring instrument.

0-7803-2500-1/95 $4.00 © 1995 IEEE

On the other hand, the measuring system we have developed has a frequency resolution of 0.001 ppb and provides excellent measurement reproducibility. Therefore, we consider our system suitable for checking whether the frequency jump is caused by nearby weak spurious resonances.

Measurement system

The concept of current characteristic measurement using the transistor Colpitts crystal oscillation circuit[2] is shown in Fig.1. If the base electric potential V_B is changed in such a way that non-linearity on the active circuit side and base input capacitance change can be neglected, the crystal unit driving current can be changed just a little. In order to measure the current of crystal units accurately, a non-contact probe[4] is used, based on the Fr-meter method[3].

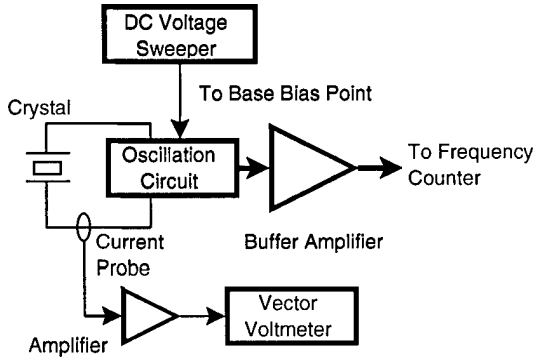


Fig. 1 Block diagram for measuring the resonance frequency vs. current characteristics of a crystal unit based on the oscillation circuit.

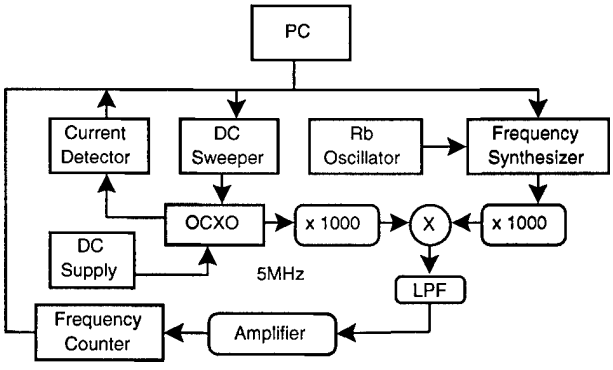


Fig. 2 Block diagram of a highly accurate measurement system for measuring the frequency vs. current characteristics of a crystal unit based on an oscillation circuit.

The block diagram of the automatic measuring system used for our experiment is shown in Fig.2. The base voltage is controlled by a d.c. voltage sweeper. The output from the crystal oscillator is multiplied 1,000 times by the multiplier, and a small fluctuation in the frequency of the crystal oscillator can be detected using the heterodyne detection method. In this system, the frequency resolution is 10^{-11}/sec and the current resolution is 1 μA.

Measured results and discussions

The specifications of the crystal units used in OCXOs for our experiment are shown in Table 1. There are four crystal units, Nos.1-4.

The results of the current-resonance frequency measurement for crystal unit No.1(Fig.3) show a frequency jump due to the change in driving current. This figure shows that the frequency of crystal unit No.1 jumps at around 0.86 mA. If this frequency jump is the cause of OCXO aging characteristics frequency jump, then the current causing the frequency jump should shift if the crystal unit is driven for a certain period of time. Therefore, we conducted an experiment by driving this crystal unit for one minute repeatedly applying 1 mA of current and measuring the drive level dependence(DLD) each time to watch the change in current causing the frequency jump.

Table 1 Specification of crystal units

Item	Type/Value
Nominal Frequency	10 MHz
Holder Type	HC-36/U
Overtone	3rd overtone
Quartz Crystal	Synthetic Quartz Unswept Crystal
Infrared Absorption	0.034
Etching Channels	12 m^{-6}
Cutting Angle	AT-cut
Equivalent Circuit Q-value	1.2×10^6
L1	1.12 H
C1	0.226 fF
Co	3.1 pF
Co/C1	13716
M	87.5

Figure 4 shows the current value causing the maximum frequency jump per specified crystal unit driving time. This figure reveals that the frequency jump causing the resonance current value increases with the passage of crystal unit driving time.

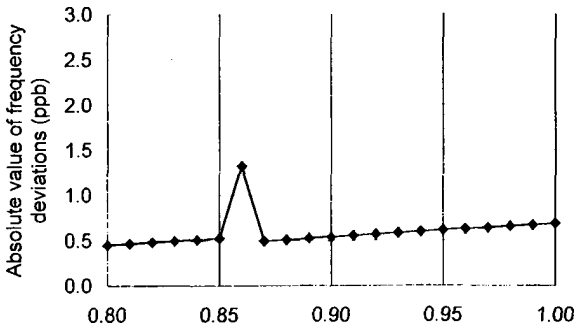


Fig. 3 The DLD characteristics of a crystal unit No.1.

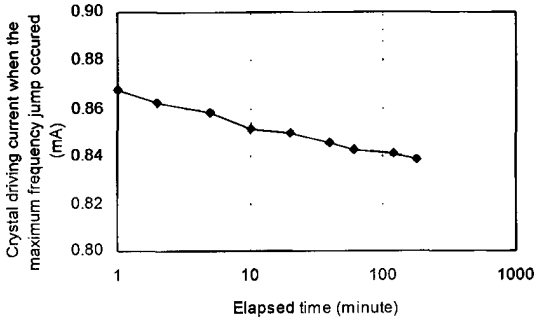


Fig. 4 The current value causing the maximum frequency jump vs. crystal driving time.

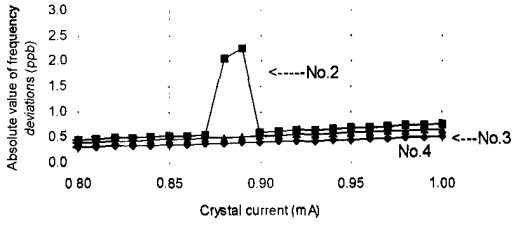


Fig. 5 The DLD characteristics of crystal units No.2 to No.4.

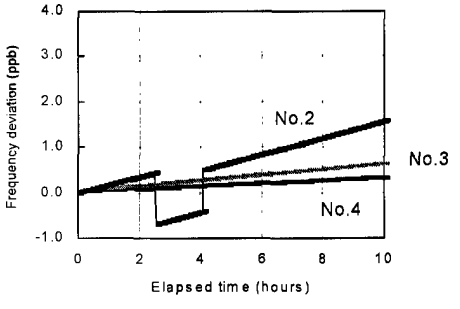


Fig. 6 Aging characteristics of the OCXOs using the crystal units which have DLD characteristics shown in Fig. 5.

Figure 5 shows the DLD characteristics of crystal units Nos. 2-4. Just like No. 1, No. 2 shows a resonance frequency jump at a specific current value. Other crystal units show no frequency jump.

Figure 6 shows the aging characteristics of OCXOs incorporating crystal units. The OCXOs used are transistor Colpitts ones with AGC. The current of the crystal units is set at approx. 0.8 mA to reduce experimentation time. For crystal unit No. 2, the frequency jumps 2 hours 30 minutes after the start of measurement and jumps again approx. 4 hours after the start of measurement before returning to the specified level. Other crystal units show no such jumps. We believe this is caused by the frequency jump causing the driving current to become negative and by the crossing of the OCXO driving current. It can be predicted that the frequency jump will occur after a much longer time when the current of the crystal unit is set at a lower value than in this experiment.

Our experiment has revealed that the cause of the discontinuous frequency jump in an OCXO is the combination of nearby spurious resonances with the fundamental mode with the passage of time.

Conclusions

We have made accurate measurements of the current-resonance frequency characteristics of AT-cut crystal units used in OCXOs in order to find the cause of the frequency jump that degrades the aging characteristics of OCXOs. Our experiment demonstrated that the cause is the combination of the nearby spurious resonances with the fundamental mode with the passage of time, and that testing the crystal units, current-resonance frequency characteristics can prevent the abrupt frequency jump. But we have not completely found out what causes the weak spurious resonances. This is a problem to be solved in the future.

References

[1] J.Vig:"Quartz crystal resonators and oscillators for frequency control and timing applications",Ann.Symp.on Frequency Control, TUTRIAL, SLCET-TR-88-1(Rev4.2),4(1991)

[2]M.Koyama, Y.Watanabe, H.Sekimoto and Y.Oomura; Proc.95EFTF(to be published)

[3]JIS(Japan Industrial Standard):JIS C 6701-1981

[4]M.Koyama,et al.:Report of the commission on high-precision frequency generating circuit miniaturization,IEE of Japan,No.17-4(July 1994,In Japan)

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

[1]Design and fabrication of a Capacitance Switch Array usable on a DTCXO.

W.Daniau*, G. Marianneau*, E. Girardet**, J. Marguier**

*LPMO/CNRS : Laboratoire de Physique et Métrologie des Oscillateurs, Centre National de la Recherche Scientifique. 32 av. de l'observatoire, F-25044 BESANCON CEDEX. FRANCE.
**AR Electronique. 14 rue de Franche-Comté, F-25048 BESANCON CEDEX. FRANCE.

Abstract

This work is based on a pre-industrial evaluation of a Capacitance Switch Array (CSA) The main idea of this study is the replacement in a DTCXO of the usual compensation system composed of a Digital to Analog Converter coupled with a varactor, in order to decrease the total power consumption and the supply valtage.

First, fundamentals of the CSA concept are exposed, followed by the conception methodology. Then, the technology that has been used is exposed and the characteristics of the final product are presented. The final CSA is a twelve bits device fully integrated in a twenty pins dual in line package with a dynamic range of variation of one hundred pF. Finally, simulated forecast characteristics of a DTCXO using the CSA are presented.

Introduction

The basic topic of this work is the study of a Capacitance Switch Array (CSA) that can be used in a Digitally Temperature Compensated Crystal Oscillator (DTCXO). The purpose of this approach consists in providing a good frequency stability versus temperature as in other DTCXO, but with a low power consumption and a low supply voltage. For a large application field such as telecommunications, there is a need for low consumption, high stability oscillators.

The original idea of this work has been presented in 1983 by Takehiko Uno and Yoshio Shimoda [1]. Then a study has been initiated by Daniau et al [2,3] in order to demonstrate the practical interest of such a DTCXO using a CSA instead of the usual compensation system (varactor coupled with a Digital to Analog Converters (DAC)). Since Digital to Analog Converters (DAC) are high consumption components and since for a great range of capacitance variation, a great range of tension must be applied on a varactor, it appears that a CSA is more suitable for these products as previous studies had demonstrated it.

The first goal of the present study was a preindustrial evaluation of the CSA in conjonction with improvements of the previous CSA structure[2,3]. The second goal is the fabrication of different quartz oscillators using the CSA, such as DTCXO or Digitally Controlled Crystal Oscillators which will balanced the VCXO.

This paper will be first devoted to the description of the fundamentals of the CSA and the design methodology. Then the requirements for the product will be discussed and the final product will be presented. Finally, time and frequency applications will be exposed and forecasted characteristics of a DTCXO using the CSA will be presented.

Fundamentals

Independently of the compensation system, a conventionnal DTCXO (as shown in fig.1) is based on the modification of the load capacitance of the crystal, by application of a tension to a varactor. Since the temperature compensation treatment is digital, a DAC is needed. Using a CSA allows to suppress the DAC because the CSA is directly controlled by digital data. As explained in the introduction, the interest of this approach consists mainly in the fabrication of low power consumption DTCXO with a low supply voltage.

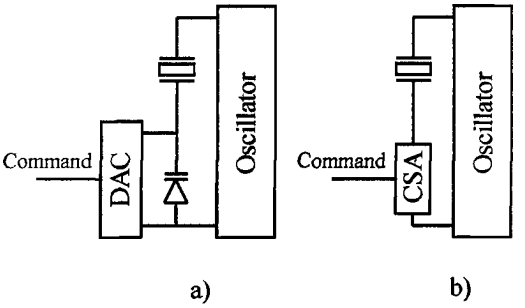


Fig. 1 : a) Conventional system b) CSA system

[1]This work was supported by French Ministry of Research under grant number : 92S0715 and 92S0716

0-7803-2500-1/95 $4.00 © 1995 IEEE

The fonctionnal scheme of the CSA is presented in fig.2. It is composed of parallel capacitors, each placed in line with a digitally controlled analog switch. If the switches are perfect and the capacitors values binary ponderated, the application of the operating word N will provide a practical output

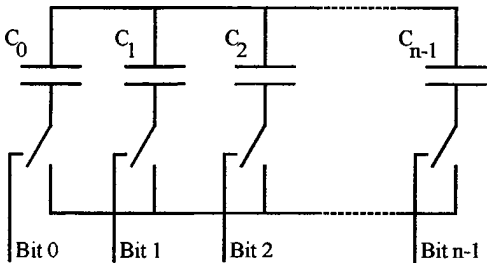


Fig. 2 : Functionnal sceme of the CSA

capacitance of $N.C_o$, where C_o represents the capacitance associated with the lowest significant bit.

Capacitors Technology

First experiments with discrete components and manual switches (which can nearly been considered as perfect) have shown the difficulty to ensure the reliability of the device because of the high tolerance on the commercial capacirors

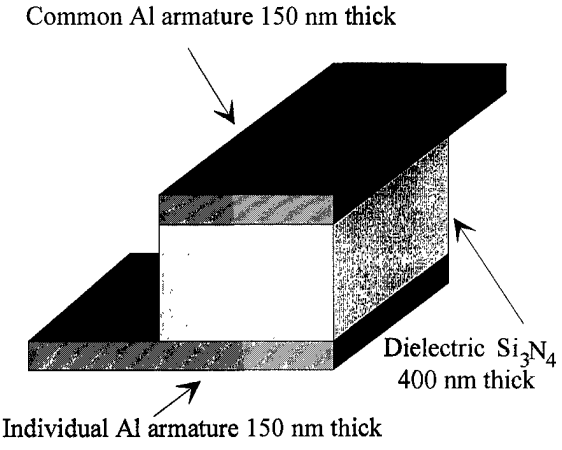


Fig. 3 : Capacitor's structure

inducing insufficient precision in the capacitor values. Moreover the exact values of capacitor selected for the device were not available. Since the first purpose of the study is to achieve an industrial product, the reliability of its characteristics appears as an absolute necessity. As a consequence, it has been decided to build all the capacitors simultaneously by using a thin film technology. This choice presents two principal advantages. The first one is that all the tolerances on the capacitor values will be correlated. The second one is that a great precision can be achieved not for the absolute value of capacitors but in the relationship between them, only by adjusting their surfaces.

As shown in fig. 3, all the capacitors are composed of three layers. The first aluminum layer (150 nm thick) is used as the individual armatures of the capacitors and the connections to the switches. The second layer is a Plasma Enhanced Chemical Vapor Deposition (PECVD) silicon nitride layer with a thickness of 400 nm used as dielectric of the capacitors. The third layer is an other 150 nm thick aluminum layer acting as the common armature of the capacitors.

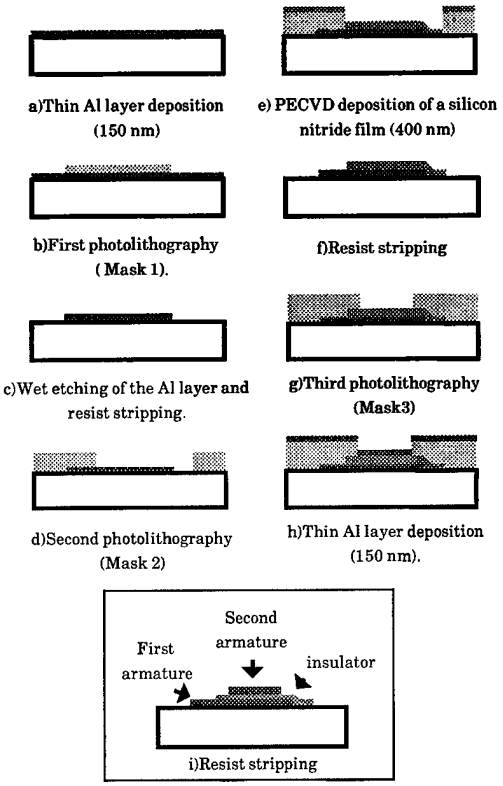


Fig. 4 : Flow Chart for the realization of a capacitor.

Fig.4 shows the flow chart of the fabrication process which needs three masks. This process only needs a wet etching step and two lift-off operations. It must be noted that the choice of the substrate is quite important. Since capacitors have to be implemented, it is absolutly necessary to have a thick dielectric substrate to reject the parasitic capacitances arising from the package. One experiment have been performed using an oxydized silicon wafer, which implies that the equivalent dielectric substrate was only 1 μm thick. In that case, the practicle capacitance was negligible in regard with the parasitic capacitances! As a consequence, all the other devices have been implemented on polished 1mm thick 2 inches diameter glass wafers. The parasitic effects have been experimentaly found negligible using this approach.

First experiments

Using the design defined in previous studies [2,3], wafers have been processed containing 80 CSA with binary ponderated capacitors values. the first task was to test the reliability from one CSA to another on the same wafer and from one wafer to another. The results are better than 5% dispersion in both cases.

Since the values of the smallest capacitances could be as small as a few multiple of 10 fF, measurements were difficult to perform using a simple capacimeter. Then they have been achieved by placing directly the array in the circuit of a quartz oscillator. An Ultra stable OCXO using a very well known crystal has been used for these measurements. The frequency shift induced by incrementation of the operating word has been measured and then related to the applied load capacitance and finally to th capacitance values of the array. This method of measurement has been fully automized with a computer and provides the characteristic curve of the device, exhibiting the capacitance versus the operating digital sequence.

It is now important to emphasize that the switches used (DMOS transistors) are not perfect. These transistors, according to the constructors documentation, can be modeled as shown in fig.5. One of the fundamental parameters influencing the choice of the transistors is the value of the on resistance which has to be as small as possible. The parameters of this non-perfect switch greatly influenced the resulting response by adding big steps on the characteristic curve for the binary powers operating word values (fig.6) which make the device non usable in this state, because there is a loss of capacitance values in the variation range.

A simulation program taking into account these parameters has been implemented to simulate the resulting curves for different configurations of the couple capacitor/switch. Since the switches were commercial products, no modification of their constitutive parameters was possible. Then the capacitors had to be adapted to the switches.

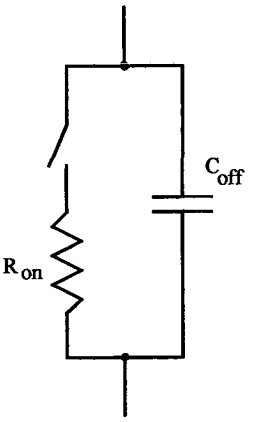


Fig. 5 : Modelisation of the switches

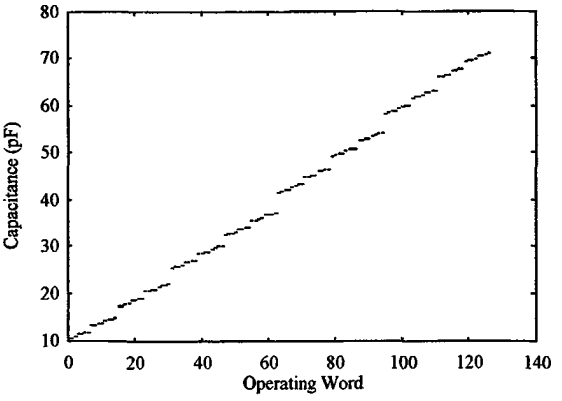


Fig. 6 : Capacitance versus Operating Word in first experiments

Design and fabrication of a new device

Since the simulation has been validated by a comparison between experimental data and the predicted results, it has been used to define a new device limiting the previous parasitic effects. The requirements were a twelve bit device with a total excursion close to 100 pF. Indeed, the minimum capacitance value of the array is determined by the C_{off} switch value, and cannot be as small as desired. This implies that for a twelve bit device, the total excursion will be largely greater than the requirements, and moreover, the dimensions of the device will be prohibitive.Then, a nine bit device has been designed with the required total excursion of 100 pF. Nevertheless using an other functionnal scheme for the array allows to produce a device in accordance with the requirements as it will be shown later in the paper.

The simulation has shown that adjusting the capacitors values allows to obtain three kinds of capacitance versus

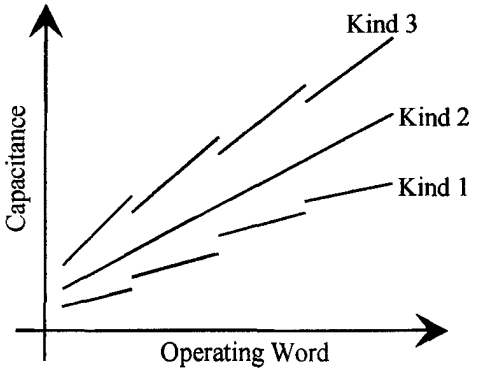


Fig. 7 : Different kinds of characteristic curves

operating word curves, as shown in fig.7. The first kind of curve with positive big steps is similar to the first experimental results, the second kind of curve corresponds to the ideal case and the third kind of curve presents negative big steps. Now, if the tolerance on the switches parameters values which is between 10 and 20% is taken into account, it appears to be hazardous to define the array to respond to the second kind of curve, because the resulting curve regarding to this tolerance will be in some cases of first kind and in other cases of third kind. Then it must be noted that the third kind of curve does not present any loss of capacitance values within the excursion. Indeed, it provides a redondance of certain values in a region around the binary powers values. Since a digital treatment will be performed behind the CSA, it will be possible to make a classification of the values and put it in a table. That is the reason why the new CSA has been designed to work in the third kind zone within the tolerance on the parameters switches. This will allows, associating a classification table (that can be simply obtain with an automized system) to each CSA, to obtain a quite linear response in all configurations.

Final device

Using the new functionnal scheme reported in fig.8, it has been possible to implement a twelve bit device without increasing the total excursion of the array. The characteristic curve of this final device is presented in fig. 9. Then, after a learning phase a classification table has been performed, providing the experimental results of fig. 10. There is of course alittle loss of operating words due to the elimination of redondances. That means that this is not really a twelve bit device, in fact the exact number of bits is not an integer and is obtain by taking the base 2 logarithm of the maximum operating word plus one. The device which characteristic is presented is then a 11.7 bits device.

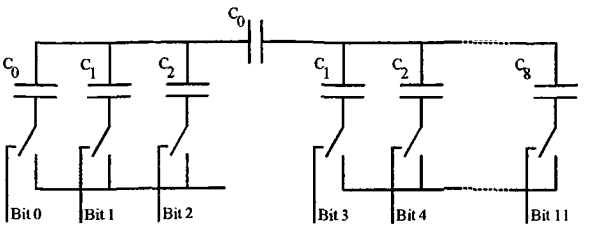


Fig. 8 : New functional scheme for a 12 bit device.

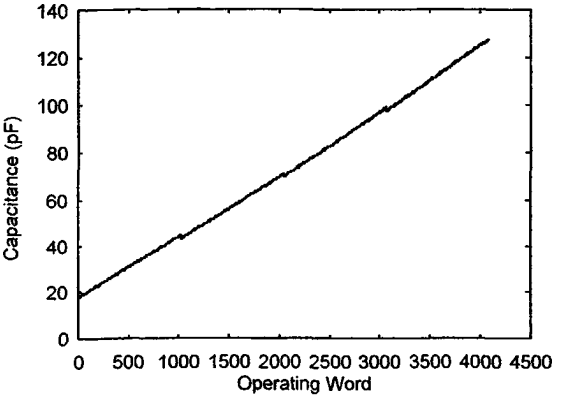


Fig. 9 : Capacitance versus Operating Word before classification for the 12 bit device.

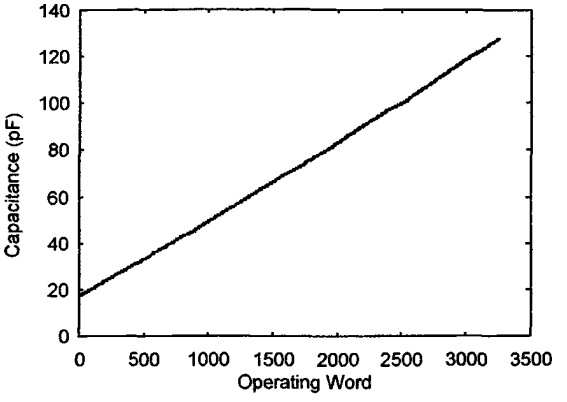


Fig. 10 : Capacitance versus Operating Word after classification for the 12 bit device

The product has been fully integrated in a 20 pins dual in line package, which contains the CSA and the DMOS switches. Fig. 11 shows a photograph of the final device.

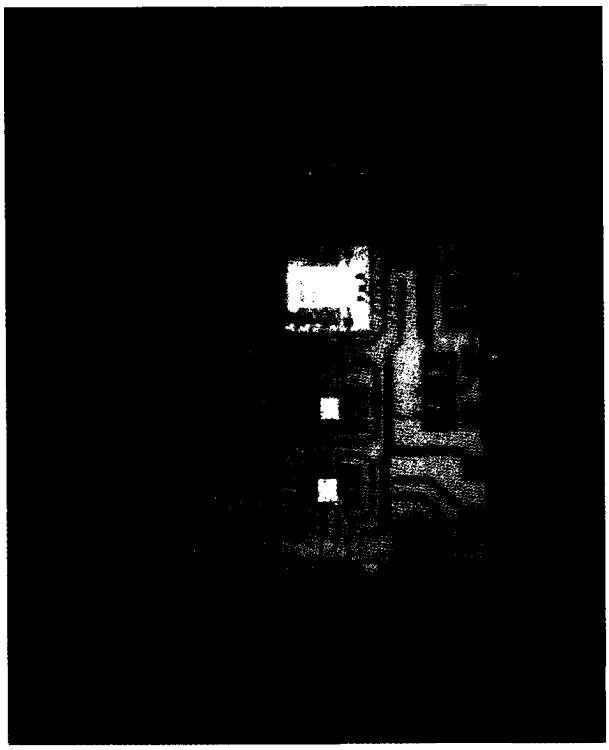


Fig. 11 : Photograph of the integrated CSA

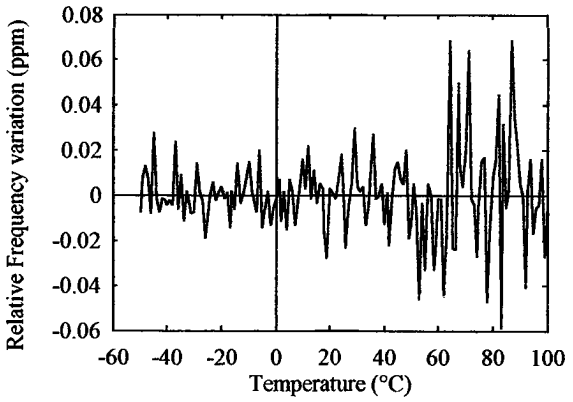


Fig. 12 : Simulation of a DTCXO using the CSA with a fundamental 16 Mhz AT cut crystal.

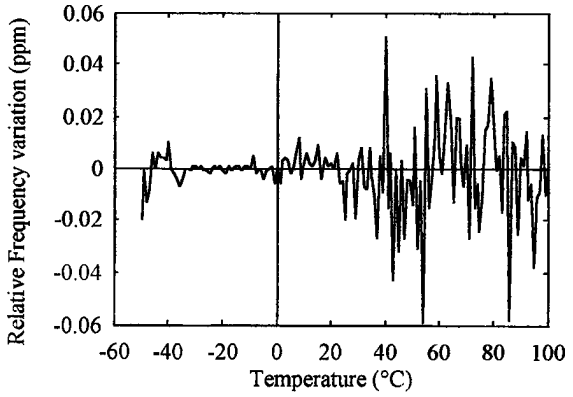


Fig. 13 : Simulation of a DTCXO using the CSA with a third harmonic 16 Mhz AT cut crystal.

Application

The CSA was originally designed for time and frequency applications and specially for DTCXO. At the time this paper is written, only simulations have been performed. One parameter used to characterize the compensation is the « efficiency » of the compensation which is the ratio between the free frequency shift of the oscillator in a given temperature range and the compensated frequency shift in the same temperature range. The higher this ratio is, the most efficient is the compensation system. This parameter is used because it is relatively independant of the crystal which is used.

With a fundamental 16 Mhz AT cut crystal, the simulation has shown an efficiency of 333 which means that the total frequency shift in the temperature range -40°C to 100°C is reduce by this factor. The result of this simulation is shown in fig.12.

One of the interest of the CSA in regard with the varicap is its big dynamic which allowing to compensate 3rd harmonic AT cut crystals, which have greater Q factor. In the same temperature range, the simulation has shown an efficiency of 285 (fig.13).

Limitations

The main limitation which had not been solved using the CSA is the noise induced by the switches when turned off. A stability measurement has been performed using a nine bit device incorporated in an ultra stable OCXO. In the first experiment, the applied operating word was 255 which means 8 switches turned on and 1 turned off. In this case with an integration time of 10s, the frequency stability was 5.10^{-12}. In the second experiment, the applied word was 256 which means 8 switches turned off and 1 turned on. Then the

frequency stability using the same integration time was 10^{-10}. Two orders of magnitude in frequency stability has been lost. This result shows that the CSA cannot be used if a stability of the oscillator lower than 10^{-10} is required..

Conclusion

The interest of a DTCXO using a CSA has been demonstrated and a preindustrial study of the CSA has been performed. Then a twelve bit device with a dynamic range of 100 pF has been implemented. Although it appears that for high stability applications, new low noise, low on resistance switches had to be found. The device has been fully integrated in a 20 pin dual in line package. Simulations of DTCXO's using this device have been achieved and it has been shown that this device allows the compensation of 3rd order quartz crystals. These simulations have shown the high eficiency that can be achieved for the compensation, this approach will allow the fabrication of high quality low power consumption low supply voltage DTCXO.

Bibliography

[1] Takehiko Uno, Yoshio Shimoda, « A new digital TCXO circuit using a capacitor switch array »
Proc. 37th AFCS pp 434-441, 1983.

[2] W. Daniau, « TCXO Numérique à réseau de capacité programmable », Thèse de docteur de l'Université de Franche-Comté, Faculté des Sciences et des Techniques. Université de Besançon, 1991.

[3]W. Daniau, G. Marianneau, « OSCCAR, Oscillator Compensated with a Capacitance Array »;
Proc. 6th EFTF (European Frequency and Time Forum) pp 165-167, 1992.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

NEW RESULTS ON HIGH PERFECTION LANGASITE CRYSTALS: STUDIES OF CRYSTALLINE DEFECTS AND MODES SHAPES.

A. ZARKA*, B. CAPELLE*, J. DETAINT**& D. COCHET-MUCHY***

*L.M.C.P. Universités P. VI et P. VII, CNRS, 4 Place Jussieu. 75252 .Paris .France.
**CNET Laboratoire de Bagneux, 196 Av.H.Ravera 92225 Bagneux France.
***CRISMATEC. 2 rue des Essarts, ZI des Mayencins, 38610 Gières France.

Abstract.
The langasite (LGS: $La_3Ga_5SiO_{14}$) is a new piezoelectric material obtained by the Czochralsky method. In this paper we present results obtained on new samples which are more representative of the crystalline quality achieved with this material. Observations using X-ray transmission topography have permitted to verify the quality of samples of different origins and to analyse the nature of their crystalline defects. Vibration modes of different types of devices made with L.G.S. have been studied by X-ray transmission topography This study was essentially concerned with plane or plano-convexe resonators. This material has interesting properties and among them compensated cuts with a large coupling coefficient and a very reduced angular sensitivity.

Material	Phase Transition	Solvent	Growth Method	Tempera-ture (°C)	d kg.m-3
Quartz	alpha/beta 573°C	Alkali (Acid salts)	Hydro-thermal NaOH/Na2CO3	320-450	2655.
AlPO4	alpha/beta 581°C	Acids (Salts)	Hydro-thermal H2SO4/HCl H3PO4	220-350	2640.
GaPO4	alpha/cristo. 933°C	Acids + ?	Hydro-thermal H3PO4/H2SO4	170-400	3570.
LGS	no	Acids Alkali.	Czochral. Pulling	Tf=1470	5764.

Table 1-Properties and growth methods of the four piezo-electric materials of class 32.

Introduction.

The langasite, as the quartz, gallium phosphate or berlinite belongs to the crystal class 32. The crystals are obtained by the Czochralsky method. There are no problem of seed (as for the growth of berlinite or gallium phosphate) or phase transition (1,2,3,4) The most important properties related to the crystal growth of these materials are summarized in the table 1 with those of quartz gallium phosphate and berlinite.

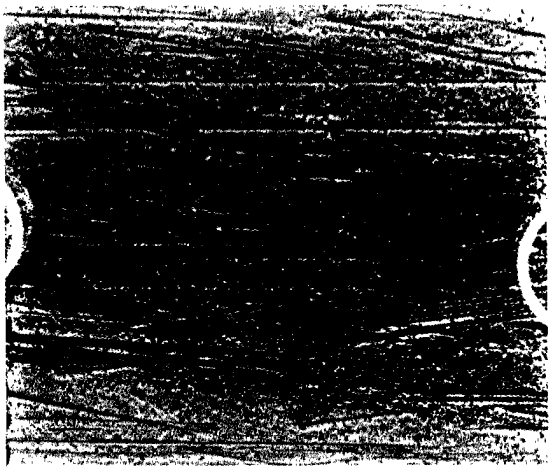
Quartz and its analogous, berlinite and phosphate gallium, have phase transitions below their melting point. They display non negligible solubilities in both alcaline and acid aqueous solutions at elevated temperature and high pressure. So that the natural choise for crystal growth for them is the hydrothermal method. Langasite presents no phase transition up to its melting point, so that it is usually grown using the Czochralsky method. The table 1 shows that quartz and berlinite have a comparable low density whereas gallium phosphate and particularly langasite have higher one, more than twice that of quartz for langasite.
For the studies of the quality of the samples and the corresponding resonators we have used traverse topography in the Laue setting with the synchrotron radiation at L.U.R.E. in Orsay (France) (5).

0-7803-2500-1/95 $4.00 © 1995 IEEE

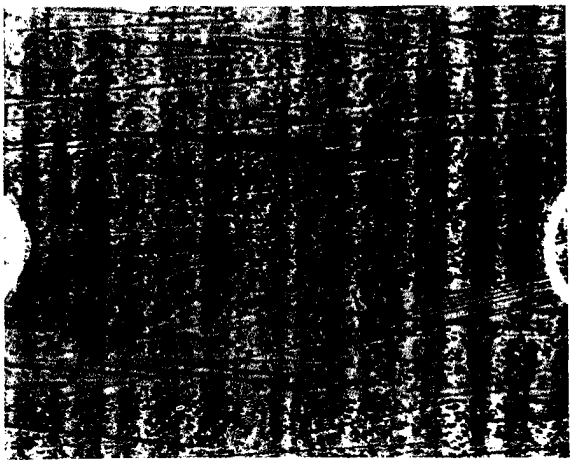
Quality of the crystals.

This section shows some results concerning the langasite crystals. For different samples the crystalline quality is analysed and different plane or plano-convex resonators are studied in the following.

The figure 1 presents two topographs of a plane resonator of about 0,5 mm thick. The surface of this Y plate is about 1 cm^2 and the electrode diameter is equal to 6 mm. These two topographs are obtained simultaneously in the same Laue pattern but with two perpendicular diffraction vectors.



a

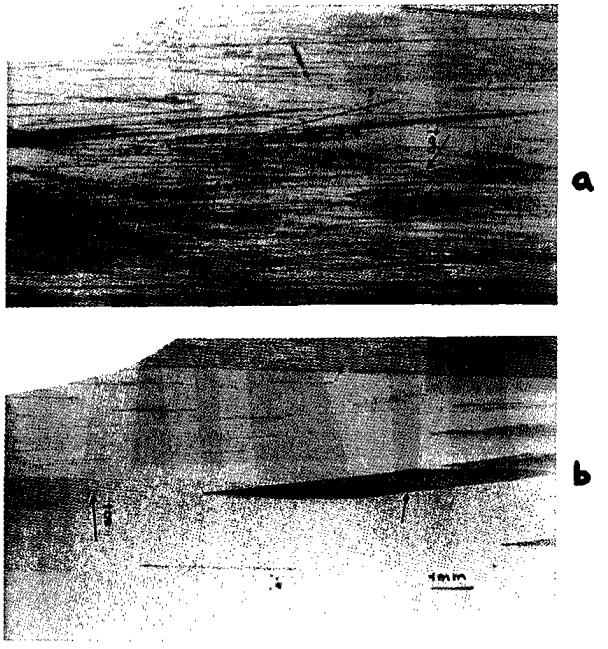


b

Figure 1-Two topographs of a langasite plane resonator of about 0,5 mm thick. The surface of this Y plate is about 1 cm^2 and the electrode diameter is equal to 6 mm. These two topographs are obtained simultaneously in the same Laue pattern with two perpendicular diffraction vectors.

The dislocations which appear are parallel with the pulling axis and their density is quite similar to the density of dislocations in the Z zone of a good quartz crystal. The little dots correspond to the images of precipitates which are due to a chemical dopping. The large black and white contrasts which appear in the topograph in the figure 1b are growth bands. They are perpendicular to the pulling axis and do not appear on the other topograph (Fig. 1a) because the diffraction vector is perpendicular to their displacement vector.

The second sample corresponds to a crystal grown by A. Gotalskaia from the Lantan Company (Moscow)(6). It is a slice of 0.5mm thick and different topographs were obtained on several Laue patterns. As an example two of these are presented in the figure 2. In Fig. 2a it can be observed that the quality of the sample is good in spite of the dislocations oriented in the pulling direction. In Fig. 2b is shown another topograph issue from the same Laue pattern. In this case the majority of the dislocations are out of contrast. In contrary, a stacking fault, very few visible in Fig. 2a, appears more distinctly in this topograph.



a

b

Figure 2-Two topographs of a langasite crystal. It can be observed the dislocation in a which are elongated following the pulling axis. In b the diffraction vector permits to observe more distinctly a stacking fault.

LGS PLANO-CONVEXE RESONATORS WITH THICKNESS EXCITATION

Experimental results.

Plano-convexe Y-cut resonators using 15x15mm square plates cuts in non-doped crystals elaborated by Crismatec were made using different designs (radii of curvature, electrode parameters,...). Some experiments were also made with Nd doped plates (10.mm diameter). Most of the studied resonators were made to have a thickness of the same order of magnitude as 10MHz 5th overtone quartz. Some samples nearly twice thicker or twice thinner were also experimented.

On table 2, the results of measurements made on a Y cut plano-convexe LGS resonator with thickness excitation an a curvature radius of 250m are displayed together with the design parameters. High Q factors were measured for the resonance frequencies of the 5th and almost the 7th and the 9th overtone. It was also observed that the electrical impedance level corresponding to these resonances are much lower than what is observed for the corresponding overtone of AT quartz. This fact results of the larger values of the dielectric constant and of the coupling coefficient. It was also observed that these resonator withstands extremely high excitation levels without showing any evidence of non-linear behaviour (typically +10 to +15dBm source power, 1/5 to 1/10 of that being dissipated in the resonator).

Plano-convexe Thickness field
Y cut 2h=0.725mm Rc=250mm 2Re=7mm

Ov.	$Fr_{(MHz)}$	$Q_{(Fr)}$	$Q_{(5MHZ)}$	$L_{(H)}$	$R_{(Ohm)}$
1	1.918	0.064	0.024	0.045	8.56
3	5.698	0.439	0.500 *1.356 (Fa)	0.132	10.89
5	9.497	0.536	1.019 *0.995(Fa)	0.154	17.15
7	13.289	0.605	1.608	0.193	26.73
9	17.097	0.387	1.325	0.209	45.0

* measurement at Fa

Table 2-Characteristics of a plano-convexe resonator with a curvature radius of 250mm

The response of the fundamental mode and the 3rd, 5th, 7th overtones are displayed in figure 3. The Q.f product measured at the resonance of the 7th overtone (Q.f=.785.10^{13} for this resonator) is of the same order of magnitude as the values obtained using the HBAR technique by G.D. Mansfeld (7) whose measurements constitute an other important indication of the possibilities to obtain extremely high Q factors from LGS devices.

In table 3, the results of the measurements made on a Y cut plano-convexe LGS resonator with a radius of curvature of 200mm are displayed together with the design parameters. For this resonator, which was the best obtained with this curvature, high Q factors were measured for the resonance frequencies of several overtones. The same conclusion as for the previous resonator can be drawn from these results. Still higher Q factors were measured at the antiresonances of some overtones The mode shapes corresponding to the different overtones were observed by X-Ray topography (see below). The measurements of table 3 were obtained at a source power of 15 dBm. No distortion of the amplitude response was observed even at high excitation power (up to +20dBm).

Thickness excitation LGS Y cut
2h=0.737mm 2Re=6mm Rc=200mm R=2.4 e-3

Ov.	$Fr_{(MHz)}$	$Q_{(Fr)}$	$Q_{(5MHZ)}$	$L_{(H)}$	$R_{(Ohm)}$
1	1.887	0.081	0.030	0.064	9.32
3	5.590	0.527	0.590	0.149	9.97
5	9.320	0.600 *1.471	1.230 *2.743	0.177	15.7
7	13.039	0.602 *0.772	1.570 *2.013	0.218	29.7

(*) Q factors measured at the antiresonance

Table 3-Measurements obtained for a resonator with Rc=200mm.

Very high Q factors at the antiresonance.

On figure 3 and in table 2 and 3 we can observe that several modes have the unusual property of having much greater Q factors at the antiresonance than at the corresponding resonance frequencies. Still higher values of the Q.f product were measured at the antiresonance of other resonators. An example corresponding to observations made for th 5th overtone of a thicker plano-convexe resonator is given in figure 4b.

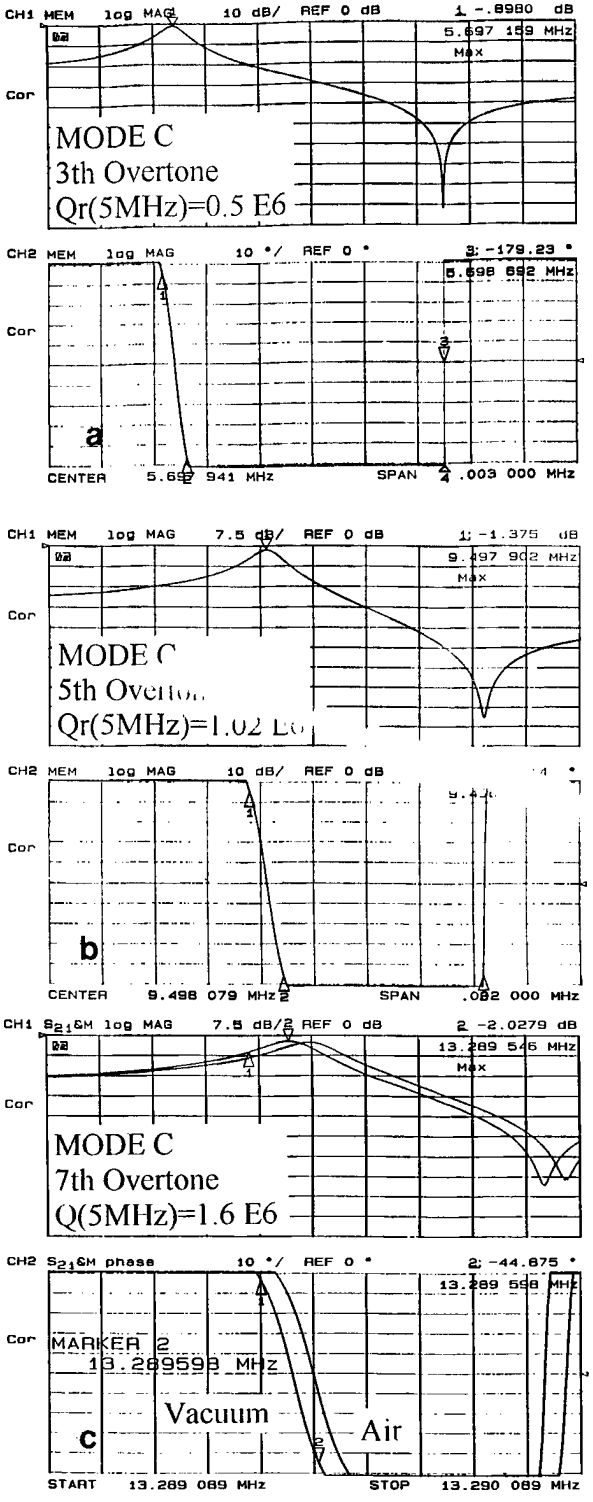


Figure 3-Electrical response of the different overtones of a plano-convexe resonator with Rc=250mm.

Although these measurements are much less accurate than at the resonance frequencies and require some particular cares (quartz holder, coaxial hardware with very reduced dielectric losses etc...), it was verified that very large values of the Q.f product can be observed for the antiresonances of the 3rd and 5th overtones of most of the measured resonators. This fact which can not be predicted by the usual equivalent scheme, seems to indicate that an unusual mechanism of losses exist at (or is predominant only near) the resonance, in such resonators. This fact is also a first indication that much higher values could be obtained at the resonance if the proper solutions could be found. In figure 4 we have represented the observations made for two resonators (4a measurement at 3dB for a resonator with Rc=200mm, 4b phase measurement for a resonator with Rc=175). Values largely over 3.106 at 5 MHz (Q.f>1.5 1013) were measured in several instances.

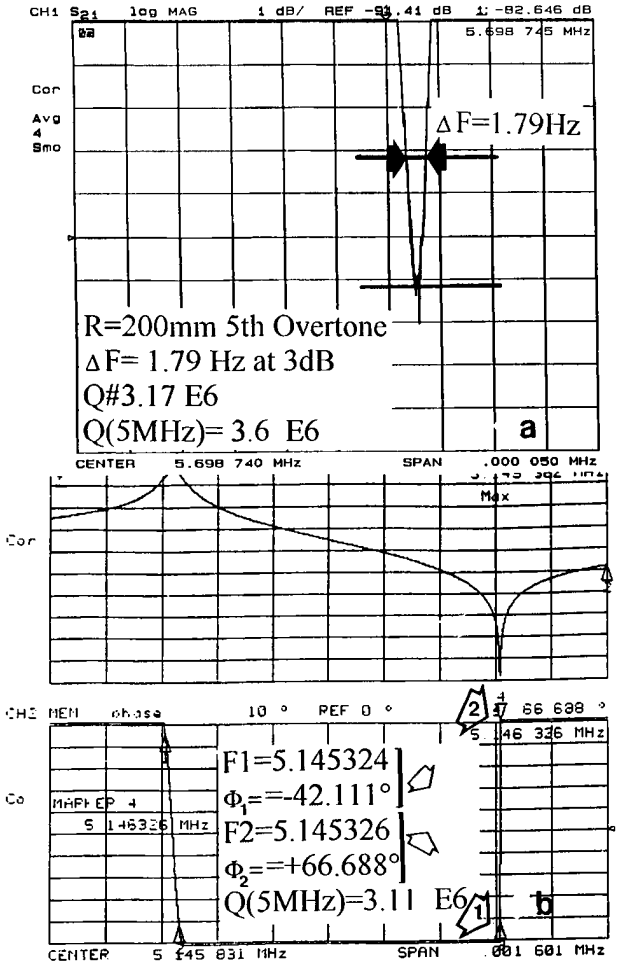


Figure 4-3dB(4a) and phase measurement(4b) of very high Q factors at Fa.

Anharmonic response.
On figure 5 we can observe that a much larger number of un-attenuated anharmonic modes propagate in these resonators (5th overtone of a Rc=200mm) than in a quartz having a similar design. This results of the fact that the "dimension" of these modes are much smaller than for quartz resonators and also it is consequence of the larger coupling coefficient (lower impedance level) which leads to much smaller attenuation in a 50 Ohm measurement system.

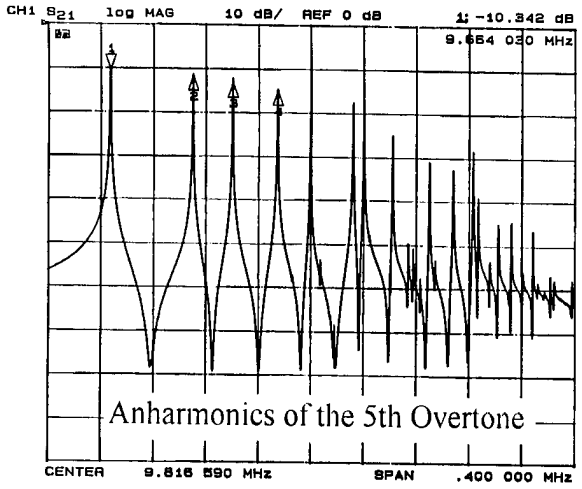


Figure 5-Anharmonics of the 5th overtone in a near 5MHz 5th overtone resonator.

Influence of the curvature radius.
A systematic study of the influence of the curvature on the Q factor was made for resonator having a thickness in the range 7 to .8mm, in all cases plates obtained from undoped crystals were used except for the 300mm radius of curvature where Nd doped material was used. The results are given for the 3rd, 5th and 7th overtones in figure 6 where it can be observed that the best results are obtained for Rc=200mm and Rc=250mm. The results obtained for Rc=300 may be some how pessimistic due to the smaller diameters (10mm) of the plates used.
During this study, for some curvatures, it was noticed that an interference can exist between the 9th overtone of the C mode and the weak anharmonics of a high overtone of another thickness mode (A or B) which are excited due to the non uniform field existing in plano-convexe resonators. We have previously observed, by X-ray topography, that a similar situation exists for the 7th overtone of certain plano-convexe AT quartz resonators. When there is no interference, very high Q factors can be obtained for the 9th overtone.

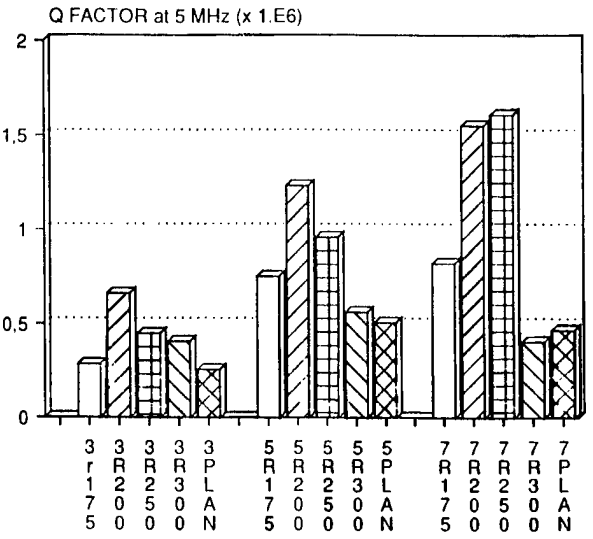


Figure 6-Influence of the curvature radius on the Q factors of different overtones.

TOPOGRAPHY

The first resonator which is presented is made with the same kind of material that the first sample (Fig. 1) and it has a similar crystalline quality. It is a Y cut plano convex resonator of 259 µm thick with a diameter equal to 10 mm, an electrode diameter equal to 3,5 mm and a curvature radius to 250 mm.
The topograph 7a shows the resonator without vibration, the second one (Fig. 7b) shows the fundamental mode and the third (Fig.7c) the third overtone with an elongation along the Z direction. The response curve with the fundamental mode and the first overtones is presented in Fig. 8.
On the Fig. 9 is given the curve showing the fifth overtone (the corresponding topograph is given on the Fig.10a) with an antisymmetrical anharmonic mode (4 on the curve) and the two first symmetrical ones (2 and 3 on the curve) which correspond to the topographs 10b and 10c.
The topograph 10d shows the seventh overtone (see Fig.8).

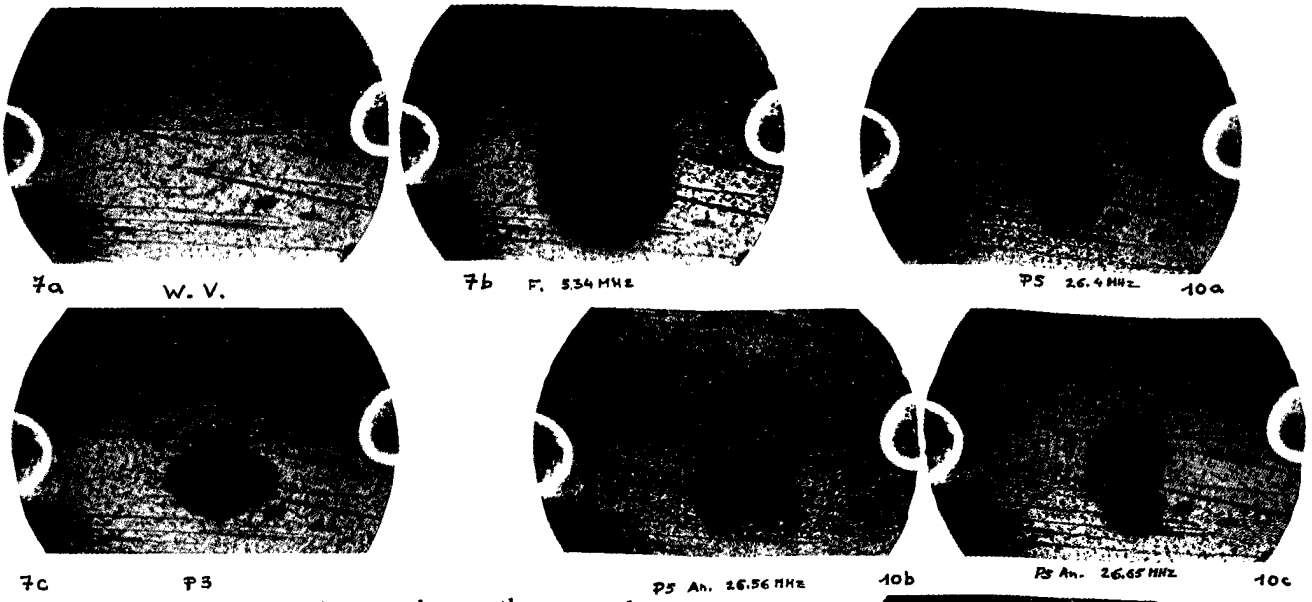


Figure 7-The topographs **a** shows the crystal without vibration In **b** is presented the fundamental mode and the third overtone is given in **c**.

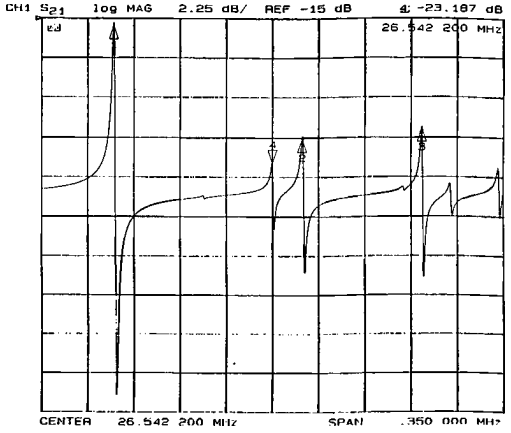


Figure 8-Response curve of the crystal of the previous figure.

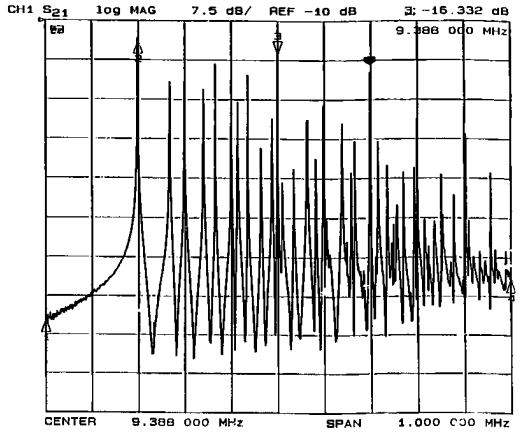


Figure 9-Response curve fifth

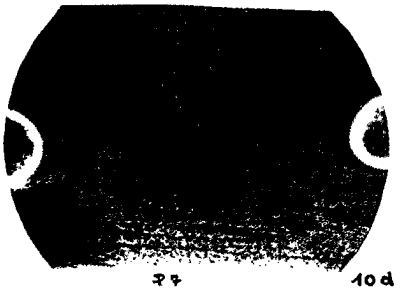
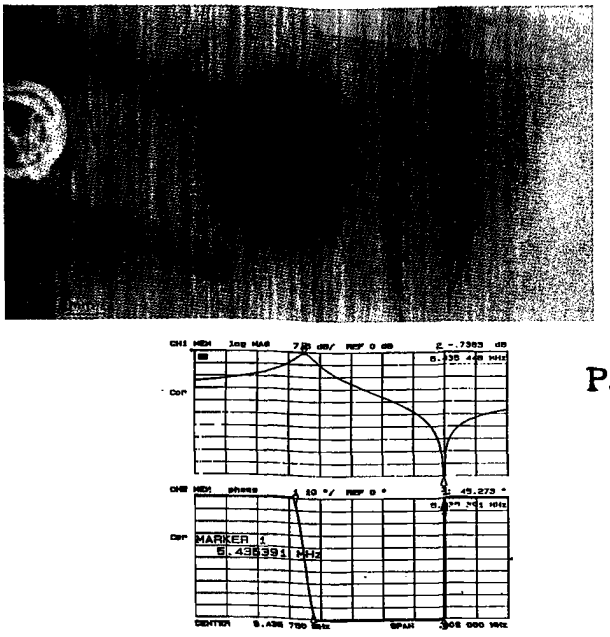


Figure 10-c and **b** correspond to the two first symmetrical anharmonics. In **d** is given the topograph corresponding to the seventh overtone.

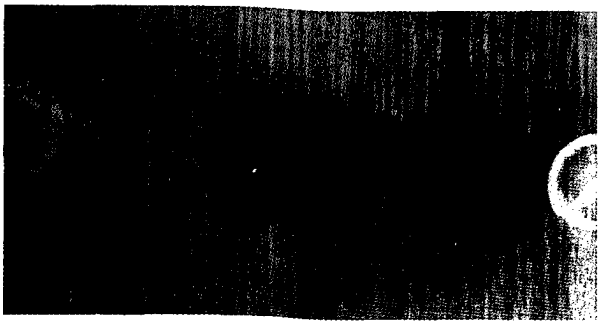
With the overtones (3, 5 and 7) coupled modes appear, in particular with the seventh one. These coupled modes are centered in a circle in the middle of the resonator It seems that these coupled modes disappear at some thickness corresponding with a cut-off frequency.

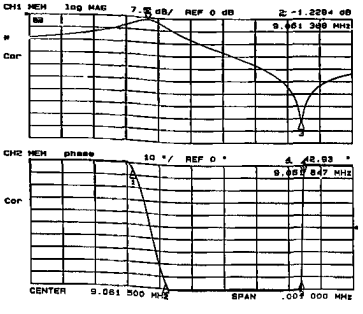
We have also studied the acoustic modes of another sample which is a Y cut plano convexe resonator of about 750 mm thick, it has a square form 15 by 15mm; the electrode diameter is equal to 8 mm and the curvature radius to 200 mm. The figure 11 presents the third overtone of this resonator. The excitation level is enough high but however the excited area is smaller than in an equivalent AT quartz resonator. Moreover it is elongated along the z direction and its edge is diffuse because of the dislocations. The response curve shows a low transmission losses. We can remark that the phase slope is steep at the resonance but still more at the antiresonance.



P3

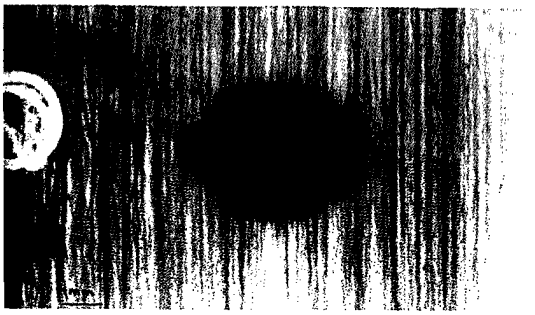
Figure 11-Topograph of the third overtone. The response curve shows low transmission losses

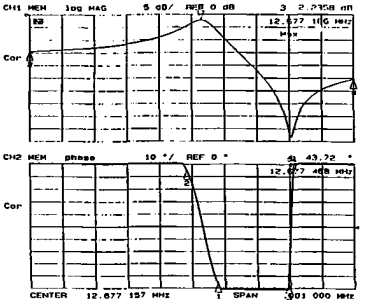




P5

Figure 12-Topograph of the fifth overtone.The response curve shows a Q factor equal to that of the third overtone but at a frequency higher.





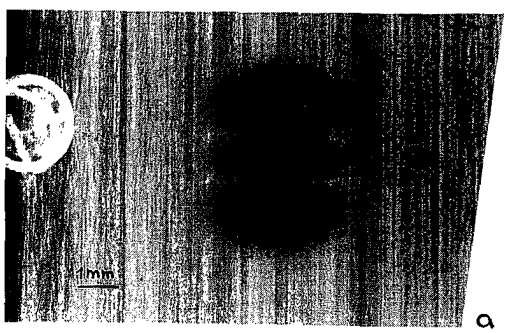
P7

Figure 13-This topograph, of the seventh overtone, is a stroboscopic topograph. The vibration is synchroneous with the pulsed synchrotron radiation, but there are no progressive components.

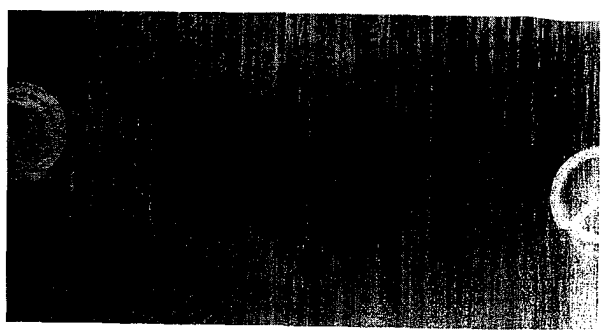
The fifth overtone (Fig.12) is elongated along the x direction like in quartz but its anisotropy is more important. The response curve shows a Q factor equal to that of the third overtone but at a frequency higher.

For the seventh overtone shown in the figure 13, the excitation level is higher that previously and the anisotropy is still important along the x direction. This topograph is a stroboscopic topograph, the vibration is synchroneous with the pulsed synchrotron radiation, but there are no progressive composent. Electrically these three modes are good with a Q factor, calculated for 5 MHz with the usual hypothesis of 1/f dependence, respectively equal to 0.6, 1.2 and 1.5 millions. These values are very good especially for our first langasite resonator with parameters which were not totally optimized.

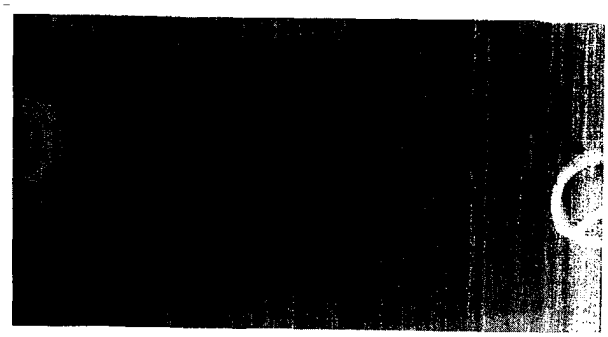
The response curve (Fig.5) shows the existence of a very great number of anharmonics of the fifth overtone. The topographs given in Fig.14a to 14e are the first anharmonics of the fifth overtone which appear on the response curve. The topograph, in the figure 14f presents an anharmonic mode with many nodal lines. The great number of un-attenuated modes results from the large dimensions of the resonator compared with those of the vibrating area. Then it is possible to make, with this material, miniaturized resonators.



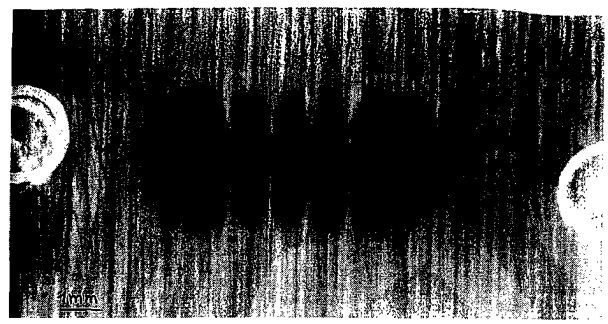
a



b



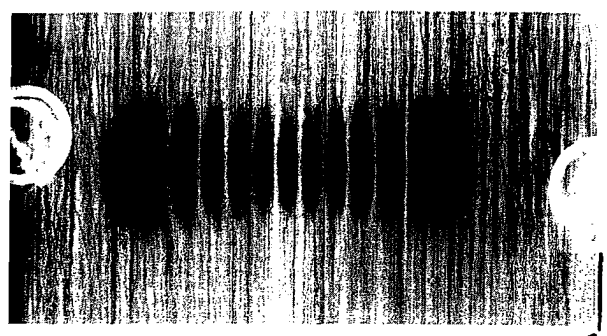
c



d



e



f

Figure 14-The topographs **a** to **e** are the five first anharmonics of the fifth overtone of the previous resonator (Fig. 13). In **f** is represented a topograph of a mode with many nodal lines.

The last example of langasite resonator corresponds also to a Y cut plate of 730 µm thick, but with horizontal Z axis. This sample has many dislocations which disturb the modes. The topographs show respectively the third, the fifth, the seventh and the ninth overtone (Fig. 15a, 15b, 15c & 15d).

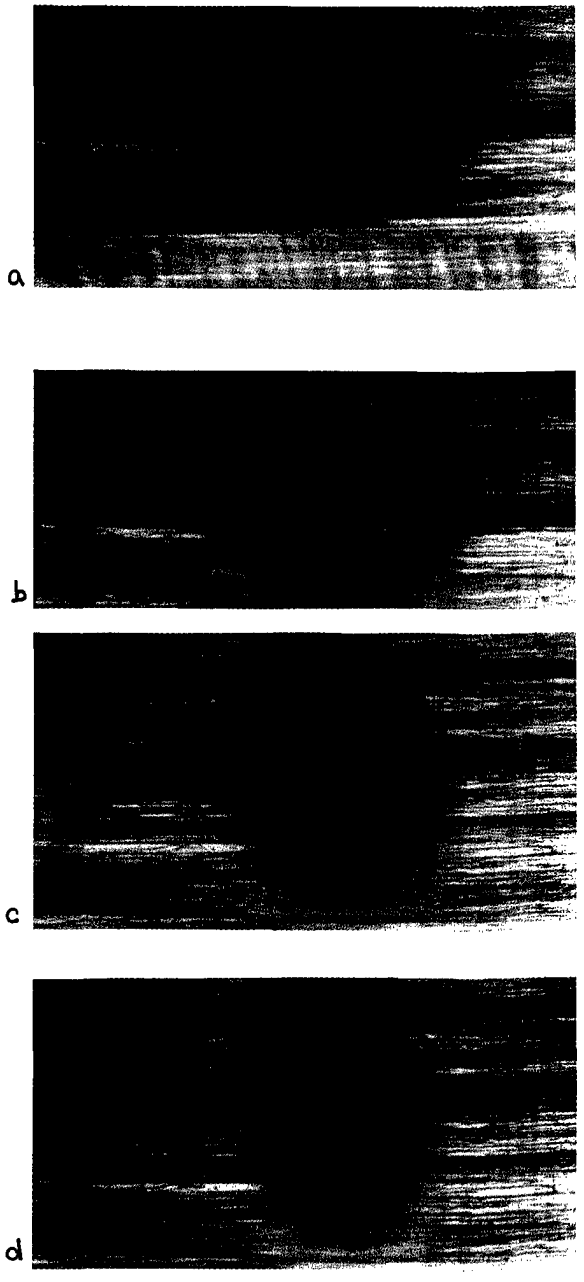


Figure 15-Another example of langasite crystal. This sample shows many dislocations which disturb the vibration modes. In **a**, **b**, **c** and **d** are respectively given the third, the fifth, the seventh and the ninth overtone.

CONCLUSION

Langasite (L.G.S.) is a new material with promising properties. Observations using X-ray transmission topography have permitted to verify the quality of the samples of different origins. Defects more generally incountered are dislocations and growth bands. Studies concerning the growth must be continued in spite of the good quality of the present samples. Vibration modes of different types of devices made with this material have been studied by X-ray topography. The results have shown many interesting features concerning the vibration modes of the resonators using the Y-cut of L.G.S. (shapes of the modes, number of anharmonics and interactions with the defects).The quality of the samples is now sufficient to permit studies of resonators and the associated vibration modes.

References

(1) J. SCHWARTZEL, B. CAPELLE, A. ZARKA, E. PHILIPPOT, A. IBANEZ & J. P. DENIS & J. DETAINT. 8th European Frequency and Time Forum. March . 9-11 1994. Munchen. Germany.
(2).B. CAPELLE, A. ZARKA, J. DETAINT, J. SCHWARTZEL, E. PHILIPPOT, A.IBANEZ & J. P. DENIS. Proc. of the Frequency Control Symposium. June 1-3 1994. Boston. U.S.A.
(3).A.N. GOTALSKAYA, D.I. DREZIN, V.V. BEZDELKIN & V. N. STASSEVICH. 1993 IEEE International Frequency Control Symposium.2-3 June 1993. Salt Lake City. p.339-347.
(4) I.M. SILVESTROVA, V.V. BEZDELKIN, P.A. SENYUSHENKOV & YU. V. PISAREVSKY, 1993 IEEE International Frequency Control Symposium.2-3 June 1993. Salt Lake City. p.348-350.
(5) A. ZARKA, J. DETAINT, B. CAPELLE, J. SCHWARTZEL. Proc.of theFrequency Control Symposium. 1987. pp. 236-240. Philadelphia.U.S.A.
(6) A.N. GOTALSKAYA, V.V. BEZDELKIN, D.I. DREZIN, & N. I SAVELEVA. Proc. of the Frequency Control Symposium.31 May -3 June 1995. San Francisco
(7) G.D.MANSFELD. Proc. 1994 IEEE Frequency Control Symposium p.35-39 (1994)

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THE NATURE OF LANGASITE CRYSTAL'S COLORATION

MIKHAIL F. DUBOVIK, KONSTANTIN A. KATRUNOV, TATYANA I.KORSHIKOVA

Institute for Single Crystals, National Academy of Sciences of Ukraine; Lenin Av. 60, Kharkov 310001, Ukraine; fax: 0572-320-273

Abstract- The $La_3Ga_5SiO_{14}$ (LGSO) crystals grown from stoichiometric starting material are characterizedby the absorption band edge in the UV-region at wavelengths beginning with λ =0.29 μm,the exitence of absorption bands with λ_{max} = 0.36, 2.9, 3.75 μm, as well as by $\rho > 10^{12}$ Ohm×cm, $\varepsilon_y < 25$, $\varepsilon_z < 58$ and tgδ <0.002. The presence of 10^{-3}-10^{-2} mass% of rodium in the crystals lead to the appearance of an absorption band with λ_{max} = 0.58 μm. Absorption at the said wavelength increases while heating the samples at T>623 K in electric field with E>400 V/cm. In this case ρ decreases by 100-1000 times, tgδ increases by an order. The band at λ= 0.58 μm disappears if the the colored samples are annealed in hydrogen at T >973 K and is restored at heating the colorless samples in oxigen or air above 1173K; the intensity of the absorption band of OH-group (λ = =2.9 μm) changes. The variation of crystal's composition (Ga or Si), as well as the introduction of Al or Ti ions in LGSO do not result in the emergence of absorption in the red region. However, in this case the character of absorption in the UV-region noticeably changes. Recommendations for obtaining transparent crystals are given.

I. INTRODUCTION

$La_3Ga_5SiO_{14}$ (LGSO, langasite) single crystals are a efficiency material for the piezo-engineering and quantum electronics [1]. They are used as resonators and filters of acoustic waves [1], [2]. However, coloration peculiar to the grown crystals may essentialy affect the parameters of wares. As shown in [3],[4], the coloration of these crystals is caused by either the inclusions of various impurities or the technological peculiarities of their growth. Nevertheless, a systematic study of coloration in langasite has not been performed so far.

The present research is aimed at studying the influence of a number of factors,such as the presence of rhodium impurities which enter the melt from the crucible or thermocouple, the heat treatment performed on colored LGSO crystals in hydrogen, oxygen or air at the temperature T > 973 K or in constant electric field with E > 400 V/cm at T > 623 K, the change of the content of gallium or silicon ions in the crystals, the substitution of aluminium and titanium ions for gallium and silicon, respectively, in the crystal lattice,on the value of transmission I, electrical resistivity ρ, dielectric permittivity ε and dielectric loss angle tangent tgδ of LGSO crystals.

II. EXPERIMENTS AND RESULTS

The LGSO single crystals were grown in the standard reproducible conditions [1] from the monophase stoichiometric starting material prepared by solid phase method at T > 1673 K. The deviation of the content of high-purity La_2O_3, Ga_2O_3 and SiO_2 oxides from the calculated quantities was not lager than 0.01 mass %. The content of each of the controlled impurities (Cu, Cr, Mn, V, Mg, Fe, Al, Ni, Pb) did not exceed 10^{-5}-10^{-4} mass %. The crystals were grown by means of the automate apparatus "SKIF-5" along the axes x [21 10] or y [10$\bar{1}$0] by the modified Czochralski method in air from platinum crucible with 99.99% purity. The average mass of the crystals was 900 g. The as-grown crystals were annealed in air at T = 1633 K during 12-14 hours. The additions of Al and Ti were introduced in the form of corresponding high-purity oxides into the starting material with the compositions $La_3Ga_{5-m}Al_mSiO_{14}$ (LGSO-mAl) and $La_3Ga_5Si_{1-n}Ti_nO_{14}$ (LGSO-nTi), where m(n) < 0.5. $La_3Ga_{4.975}Si_{1.0125}O_{14}$ (LGSO-NS) with an electrically neutral lattice was used as a non-stoichiometric composition.

The investigations were carried out on oriented and polished samples shaped eigther as plates of y- or z-cuts with a thikness h = 2-3 mm or as rectangular prisms with x = y = 5-13 mm and z = 5 - 8 mm. The optical transmission I of the samples was measured with respect to air in λ = 0.25-5 μm region using KSVU-23 and UR-10 spectrometers. The electrophysical parameters were determined by means of the unit E6- 13A and E8-4 type bridge (f = 1 kHz) with indium-gallium paste or silver conducting contacts precipitated on y- or z-faces.

Fig. 1 presents the optical transmission spectra of LGSO samples with different initial state.

Their electrophysical parameters are presented in Table I.

As seen from the optical spectra, the short-wavelength transmission boundary has λ = 0.29 μm (Fig. 1a), excluding the aluminium-doped samples with the concentration m = α_1 for which I is obseved at wavelengths starting from λ = 0.25 μm (curve 1). For the sample with the stoichiometry violated with

0-7803-2500-1/95 $4.00 © 1995 IEEE

respect to the content of Ga and Si, I is characterized by $\lambda = 0.36$ μm (curve 4). It should be noted that the absorption band with $\lambda_{max} = 0.36$ μm is also typical of both the samples grown from the stoichiometric starting material (curve 5) and those doped with Al ($m = \alpha_1$, curve 1) and Rh (curve 6).

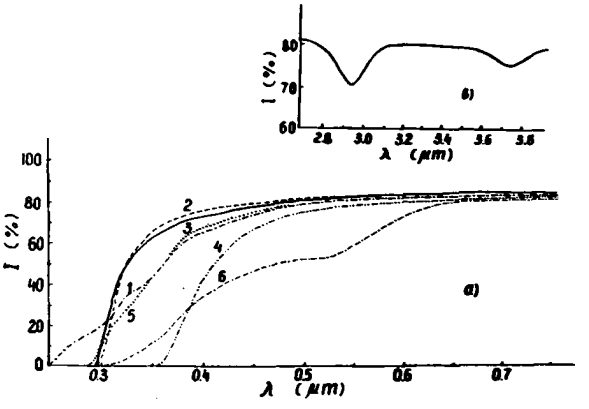


Fig. 1. Optical transmission spectra of LGSO crystals with different composition in the visible (a) and IR-region (b): 1-LGSO: α_1 Al; 2-LGSO: α_2 AL; 3-LGSO: γ_1 Ti; 4-LGSO with violated stoichiometry (LGSO-NS), 5-stoichiometric LGSO: 00, 6-LGSO: Rh (y-cut, 1-5- z-cuts).

At the increase of the content of Al up to $m = \alpha_2$, as well as at the introduction of Ti ($n = \gamma_1$) this band practically disappears (curves 2 and 3).

The doping of the crystal with Rh results in the appearance of a weak ($\lambda_{max} = 0.45$ μm) and a more intense ($\lambda_{max} = 0.58$ μm) absorption band in the visible region (curve 6); these bands make the color of the crystal pink. For the Ti-doped sample a weak absorption band is observed in the region of $\lambda_{max} = 0.41$ μm (curve 3). In the IR-region only two absorption bands with $\lambda_{max} = 2.9$ μm and $\lambda_{max} = 3.75$ μm are revealed (Fig. 1b). The intensities of these bands are practically the same in all the investigated samples.

Judging from Table 1, all the samples had a high resistivity ($\rho > 4.22 \cdot 10^{12}$ Ohm×cm) in the initial state, $\varepsilon_y < 24$, $\varepsilon_z < 58$, tg $\delta = 0.0003$-0.004. The minimal value of tgδ was chartacteristic of the crystals containing a large quantity of Al ($m = \alpha_2$) and Ti ($n = \gamma_1$).

Shown in Fig. 2 is the influence of heat treatment in different media on the transmission spectrum of "colored" Rh-doped LGSO sample.

Thus, annealing in oxygen stream (T = 1223 K, 4 h) resulted in a slight decrease of the intensity of

TABLE I
Some electrophysical parameters of LGSO crystals with different composition measured at room temperature

Crystal composition	$\rho, 10^{12}$ Ohm×cm	ε	tg δ
1 LGSO: α_1 Al	4.51	51.5	0.0018
2 LGSO: α_2 Al	5.76	58.0	0.0003
3 LGSO: γ_1 Ti	5.54	53.1	0.0006
4 LGSO: NS	5.1	56.0	0.0013
5 LGSO: 00	4.22	50.8	0.0012
6 LGSO: Rh	14.8	24.0	0.0045

Note: All the samples were annealed at T = 1623-1633 K, 14-16 h; the samples 1-5 had the z-cuts, 6-y-cut; the order of the samples correspond to those of the curves in Fig. 1.

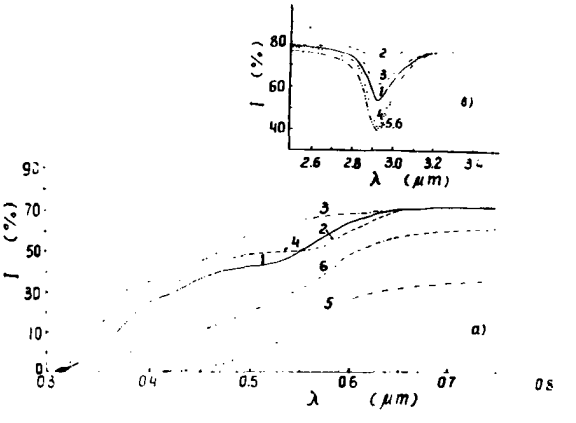


Fig. 2. Optical transmission spectra of the LGSO: Rh crystals depending on heat treatment type in the visible (a) and IR-region (b) for the sample (1) and the samples annealed in oxygen (2), hydrogen (3), air (4), electrical field (5,6). The annealing conditions are presented in Table II.

the absorption band with $\lambda_{max} = 0.58$ μm (curve 2); the band in the IR-region practically disappeared. At subsequent annealing in hydrogen stream (T = 973 K, 30 min) the red absorption band completely vanished, whereas the band with $\lambda_{max} = 0.45$ μm manifested itself more intensively (curve 3). After such an annealing the intensity of the absorption band with $\lambda_{max} = 2.9$ μm increased, too (Fig. 2b). The restoration of the band with $\lambda_{max} = 0.58$ μm took place at the annealing of the "colorless" crystals in oxygen or air (T>688 K, 20 min or longer,

curve 4). When annealed in constant electric field (E > 380 V/cm, T > 688 K, 20 min or longer), LGSO:Rh crystals acquired brown color, the absorption band with λ_{max} = 0.58 μm became more intense, the absorption edge shifted towards λ=0.4 μm (curve 5).

As seen from the values of the electrophysical parameters of LGSO:Rh samples subjected to different types of heat treatment (Table II), an essential decrease of ρ (by 100-1000 times) and the rise of tgδ by an order may be achieved only by annealing the samples in electric field.

TABLE II
Electrophysical parameters of LGSO:Rh samples subjected to different types of heat treatment

	State of sample	$\rho, 10^{12}$ Ohm×cm	ε	tgδ
1	Initial, z-cut (curve 1)	0.2	56.0	0.004
1a	Initial, y-cut	14.2	24.0	0.0045
2	Position 1, after annealing in oxygen stream (1223 K, 12 hrs, curve 2)	0.1	55.4	0.002
3	Pos. 1, annealing in hydrogen stream (973 K, 30 min, curve 3)	10.0	56.0	0.0025
4	Pos. 1, annealing in air (1633 K, 12 hrs, curve 4)	10.0	55.1	0.0024
5	Pos. 1, annealing at E = 380 V/cm, T= = 688 K, 20 min, curve 5)	0.074	22.0	0.01
6	Pos. 1, annealing at E = 620 V/cm, T= = 733 K, 20 min, curve 6)	0.0065	27.0	0.01

III. DISCUSSION

The character and location of absorption edges, the presence of absorption bands in the visible and IR-regions of the optical transmission spectra, as as well as the electrophysical parameters of the investigated samples allow to make an assumption on the nature of coloration of LGSO crystals and to propose a method for obtaining transparent crystals.

First of all, the change in the intensity of absorption in LGSO:Rh samples in the region of λ_{max} = = 0.58 μm and λ_{max} = 2.9 μm depending on heat treatment conditions points to the fact that this absorption may be caused by the formation of color centers by Rh ions and OH-group. This testifies to a deficiency of oxygen ions in the initial crystals and becomes apparent even at very small quantities of Rh (10^{-3}-10^{-2} mass %).

Moreover, on the base of the differences in the character of UV-absorption spectra for the samples with different composition we may suppose that the crystals grown from the "stoichiometric" starting material in fact are not stoichiometric from the viewpoint of the ion composition of the lattice (within the said accuracy of the control of the main components content). This is also characteristic of the crystals which have optical spectra similar to those of our crystals (see, e.g. [3]). The accuracy of chemical and X-ray analysis, i.e. 0.01 mass %, does not permit to reveal the presence of other phases ($LaGaO_3$ being the most probable [4]) in the crystals. Nevertheless, the character of the UV-absorption spectra of "stoichiometric" LGSO crystals and those containing small quantities of Al and Rh (Fig. 1a, curves 1, 4, 6), as well as the peculiarities of the absorption of the samples with Ti and a large content of Al (Fig. 1a, curves 2,3), testify to different degrees of the occupation of the crystal lattice sites with Ga^{+3} and Si^{+4} ions. Judging from the character of the absorption edge, the highest occupation degree will take place in the crystals containing Ti and an elevated quantity of Al. This correlates with the data of the paper [3] where an elevated volatility of gallium and silicon ions at the growth of LGSO single crystals is noted. The introduction of TiO_2 and Al_2O_3 oxides which have higher melting points seems to favour the occupation of gallium and silicon sublattice must contribute to the decrease of dielectric loss in LGSO:Al (Ti) crystals.

The most typical light yellow, "amber" color of the investigated crystals is caused by the presence of stable absorption bands in the 0.36-0.41 μm region of their spectra. These bands are absent in LGSO crystals with an elevated content of Al. Such crystals are characterized by less intensive coloration in comparison with other types of investigated crystals. Moreover, the additions of Al and Ti taken in the chosen concentrations improve the parameters of articles made on the base of these crystals [5]. For instance, piezoelements manufactured from LGSO:Al crystals compare favourably with those based on LGSO:00, since their inductance is by 7% lower, Q-factor is by 18 % higher and thermal frequency instability within T = 313-353 K range is by 25% lower. The use of LGSO:Ti crystals allows to decrease the inductance of resonators by 12% and to raise their Q-factor and resonance range by 12% and 10%, respectively.

VI. CONCLUSIONS

Rhodium seems to be a technological impurity which causes the coloration of langasite single crystals.

The introduction of titanium and aluminium ions into the crystal lattice results in the disappearance of coloration of the crystals, improves their stoichiometry and, consequently, the electrophysical parameters of articles based on this material.

REFERENCES

[1] M. F. Dubovik, I. A. Andreyev, Yu. S. Shmaly, "Langasite $La_3Ga_5SiO_{14}$, an optical piezoelectric: growth and properties," in Program of 1994 Freq. Contr. Symp., Boston, 1994, p. 16.

[2] V. B. Grousinenko and S. A. Sakharov, "Miniature crystal units and monolithic filters," in Program of 1994 Freq. Contr. Symp., Boston, 1994, p. 89.

[3] A. A. Kaminski, B. V. Mill, I. M. Silvestrova, "The nonlinear active material $(La_{1-x}Nd_x)_3Ga_5SiO_{14}$," in Izv. AN SSSR, ser. Fiz., vol. 47, pp. 1903-1908, 1983.

[4] M. F. Dubovik, G. M. Ivanova, S. A. Lebedev et al., "Optical quality and structure perfection of $La_3Ga_5SiO_{14}$ single crystals grown by the Czochralski method," in Izv. AN SSSR, ser. Neorg. mater., vol. 24, no. 3, pp. 517-519, 1988.

[5] I. A. Andreyev, M. F. Dubovik, "Modified langasite - a base of low frequency piezoelectronic devices," in Proc. 14th Conf. on Acoustoelectronics and Solid State Acoustics, Kishinev, 1989, p. 67.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Surface and volume defects in langasite crystals

S.A.Sakharov*, Yu.Pisarevsky**, A.V.Medvedev*,P.A.Senushencov**,V.Lider*

*FOCO Co , **Institute of Crystallography RAS

Abstract

Volume defects of langasite crystals were investigated by optical and x-ray methods. Procedure of mechanical and chemical treatment was developed. Study of surfaces defects were made by electronographical and optical methods. Samples of resonators with various volume defects and surface treatment were investigated. Based on chemical etching thecnology monolitic 80 MHz filter was designed and preparated.

Introduction

At present stage $La_3Ga_5SiO_{14}$ (langasite) crystals attract attention of many scientists of the world by their zero 1st order temperature coefficients, lower compared with quartz acoustic attenuation coefficient at hypersound frequency and by absense of phase transitions up to melting temperature(1).

Earlier LGS -crystals investigations were directed to the study of their applicability in different fields of piezoelectric engineering rather than aimed at crystal structure perfection, defects presents and impurity influence. Recently S.Laffey,M.Hendrickson and J.R.Vig investigated mechanicaly polish and chemically polish/etch this material (2).The purpose of this work is investigation of bulk and surface defects, development of procedures for evaluation of defects influence on product parameters. The investigated crystals have been grown by Czokhralsky metod along the optical Z-axes.

Observation internal stresses in polarized light.

IRL -425 universal conometer designed in ICRAS is meant for study of stractural defects in crytstals (mosaicity, slip traces, twins, foreign inclusions, residial stresses etc). Instrument makes possible examinations in convergent and parallel polarized light and in certain monochromatic wavelenths. Depending on the purpose of the study, universal conometer can be adjusted by simple substitution of components to give different versious of optical schemes. Instrument also have cuvets for study of unprocessed as-grown boules with immersion liquids.

In our case we used mercury PRC-4 lamp and green filter and three optical schems : convergent polarized light, parallel polarized light without and with quartz wedge.`On Fig. 1-3 are shown conometer pictures according 3 above optical scheme for crystals which were grown by A. Gotalskaya (Lantan Co).On Fig 1 conoscopic picture we can see rather big eliptical distorsion becouse of residial stresses . On Fig. 2 we can see internal stresses distribution in (001) plane in parallel light.

Area with big distortion go along all length (001) direction. With help third scheme (Fig 3) we could measure the distribution of internal stresses . Optical pass distortion would be proportional to the difference in the main stresses acting in plane perpendicular to the direction of observation

$$\Delta = C(\sigma 1 - \sigma 2)$$

0-7803-2500-1/95 $4.00 © 1995 IEEE

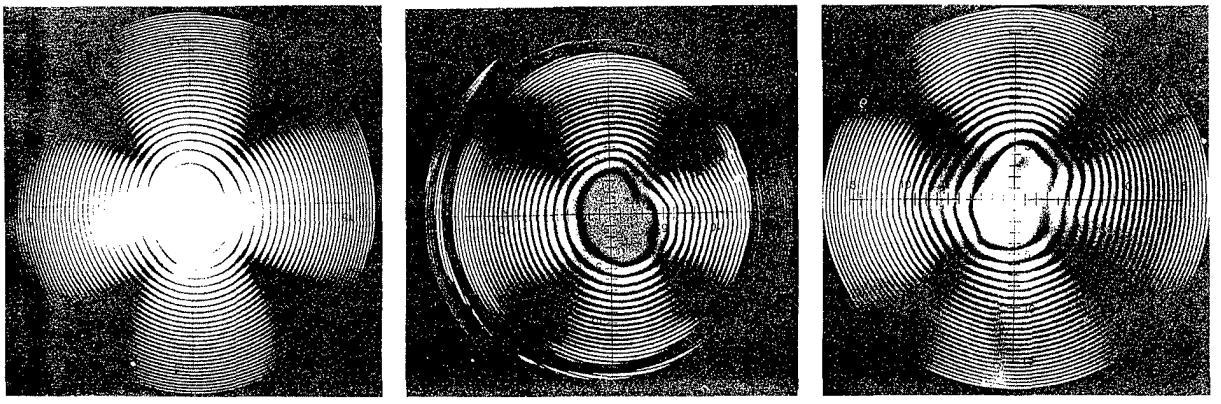


Fig. 1 Conoscopic picture of langasite samples. Top, middle and bottom parts

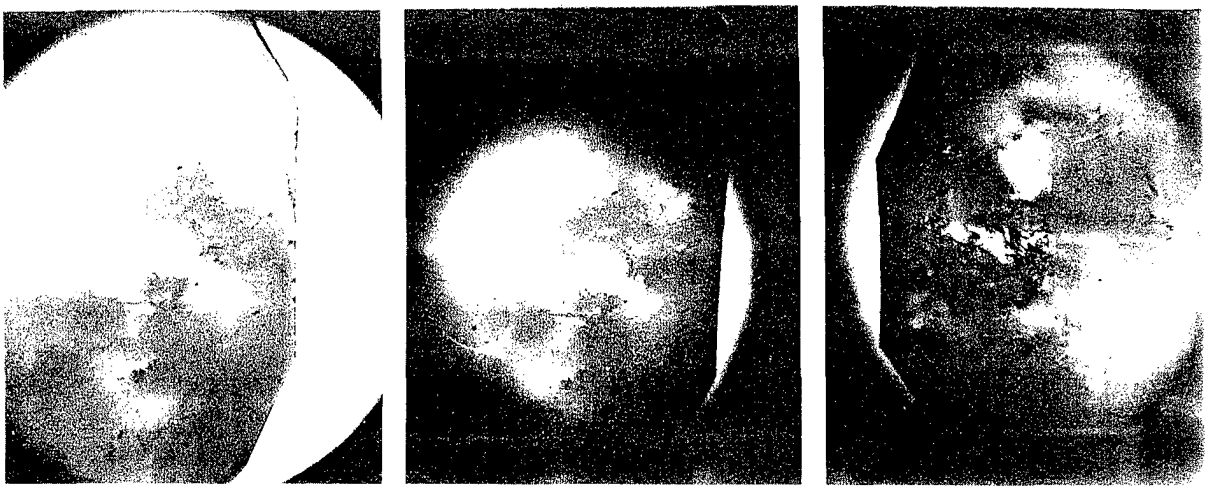


Fig. 2 Orthoscopic picture of langasite samples (same as on Fig. 1)

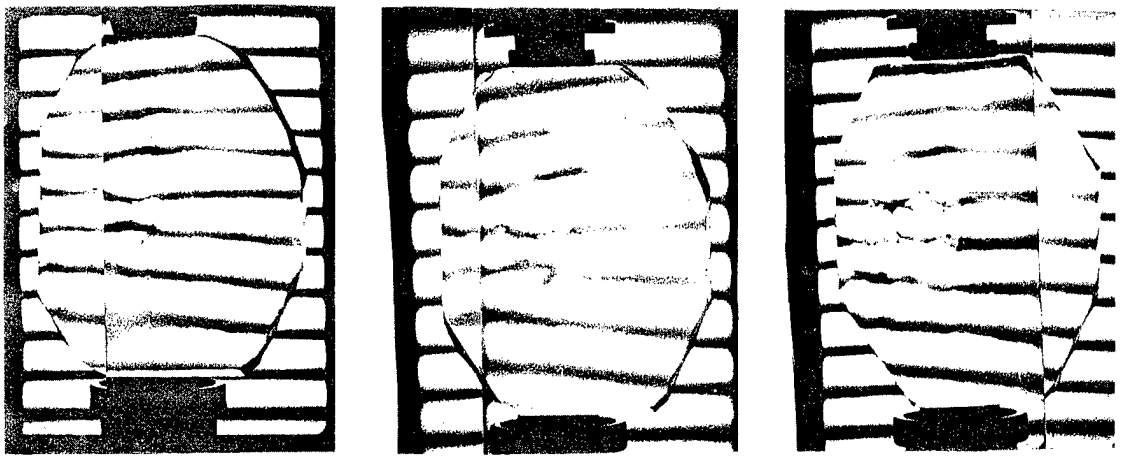


Fig. 3 Orthoscopic picture of langasite samples (same as on Fig. 1) with quartz wedge

C - effective piezooptic constant, which yet not measured. So on this stage we only could to receive qualitive distribution. only qualitive distribution. Crystals, which were grown in other than Lantan places have more uniform residial stresses distribution.

Observation of laser light scattering

To examine crystals for the presense of inclusions He-Ne laser was used to illuminate the crystal in the direction perpendicular to that of observation under such lighting bubles shine lightly due to scatter of light beam falling on the sample. Inclusions and other inhomogeneities can be seen with naked eye or with help of a microskop also we saw the static diffraction pattern which show on existanse of periodic layers in the direction perpendicular to that of growth.

OIbservation of X-ray diffraction.

The most exact results are obtained when using X -ray diffractometer methods. In particulary determining halfwidth of lines of X-ray reflections by means of a three -crystall X-ray spectrometer with extra-high resolution allow to evaluate both the extent of crystals perfectness and to measure deviations of lattice parameters. On Fig. 4 a halfwidth of line (330) Cu kα (FWHM) is shown . The period of growth ingomogenities 0.035 - 0,043 mm concide with figure obtained from laser difraction. Using the special method of determination of lattice constants we determined that such inhomogenities connected with periodic llattice parameters oscilations with amplitude $\Delta a/a = 1.5\ 10^{-4}$.

In course of investigations the most perfect crystals region have been chosen for further Y-cut crystals plate manufacturing.

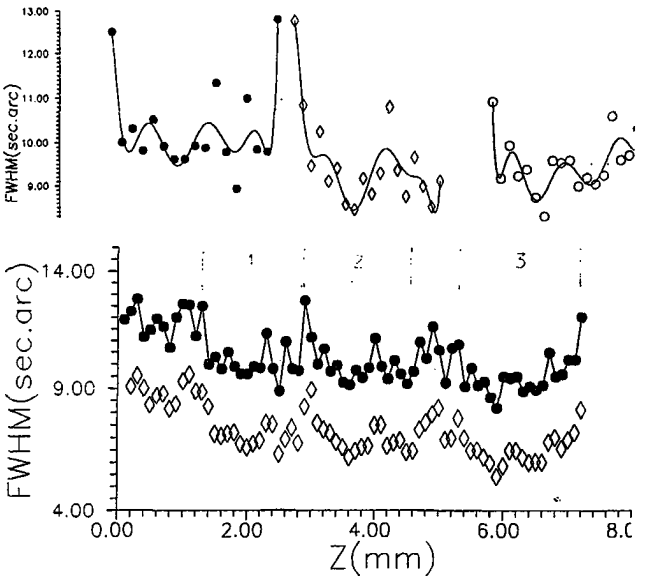


Fig. 4. Distribution of halfwidth of X-ray $Cu_{k\alpha}$ ine along (0001) LGS direction.

Surface treatment

While manufacturing these elementds fives different mechanical and chemical surface treatment have been used. In total 5 lots have been fabricated. Initialy a commonly used for quartz technology of treatment have been used. First lot don't undergo next treatment. 2nd and 3rd lots have been treated by means of a diamond slurry with different batch allowances for polishing . 4th and 5th lots have been subjested to an additional chemical-mechanical treatment, i.e. non deep etching has been performed after final polishing. This operation was aimed at removal of disturbed surface layer after polishing and at maintainence of the surface quality.

All the lots have been tested by measuring surface roughness Ra (Fig. 5) and disturbed layer depth. Then these plates were measured for relative Q evaluation plates. The measurement results are given the table.

Lot No	Disturbed layer, μm	Ra -value, μm	Quality factor Qx 10^{-3}	Treatment method
1	18.52	0.07	12-15	Quartz technology
2	18.05	0.017-0.019	20-25	Diamond slurry
3	18.14	0.019-0.021	20-25	Diamond slurry
4	17.81	0.016-0.018	27-32	Additional treatment
5	17.68	0.015-0.017	27-32	Additional treatment

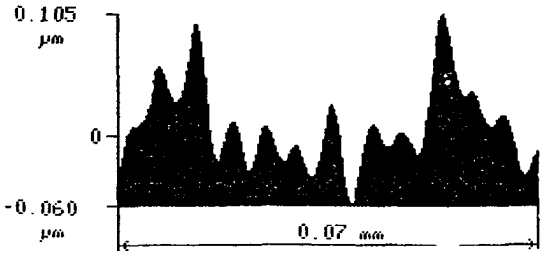


Fig. 5a. Profile of LGS surface roughness after quartz technology treatment

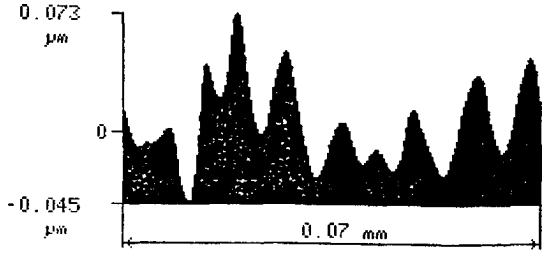


Fig. 5b. Profile of LGS surface roughness after diamond slurry treatment

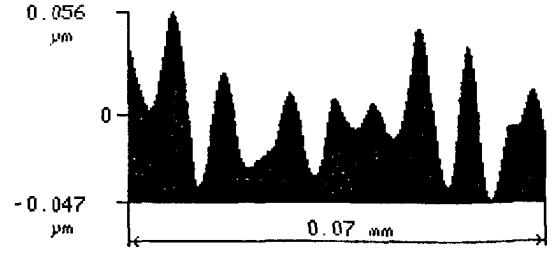


Fig. 5c. Profile of LGS surface rougness after additional treatment

Electron diffraction method was used for surface structure evaluation. Elecronogrammes have been takes from lots 2,3 and 4,5 which reveal a more perfect surface of the lots 4 and 5 ,though in all cases polycristaline layer have been seen sometimes covering surface partly.

Two lots of resonators were produced. One with diamond slurry and other with additional treatment.For 1st lot avarage Q value was about 40 $x10^3$ for 1st obertone and about $80x10^3$ for 3rd obertone. 2nd lot had avarage Q value $85x10^3$ and about $120x10^3$ for 3rd obertone. At low temperature (200 K) Q factor grow 4-5 times and at 75 K 10 - 12 times. Such dependence seems to connect with defect strusture of samples under investigation.

In spite on all above discribed defects state of art langasite allow to produce high quality piezoelectric elments. We designed two pole monolitic filter with central frequency 80 MHz .Fig 6 show frequency response of this filter. From this picture ones could see advantages in passbandwidth and stopbandwidth compare with quartz filter on same frequency range. It could be useful for the Group Spesial Mobile (GSM) digital cellar phones.

Conclusion

We studied pecularyties of volume defect structure of langasite crystals.Used optical and X-ray methods could be used for quality control under piezoelement production.

We developed methods surface treatment and chemical production of mesostruture of langasite pizoelements. Tested optical and electronnographical methods could be used for surface control of piezoelements.

With help above methods control and crystal surface treatment high frequency monolitic filters were preparated.

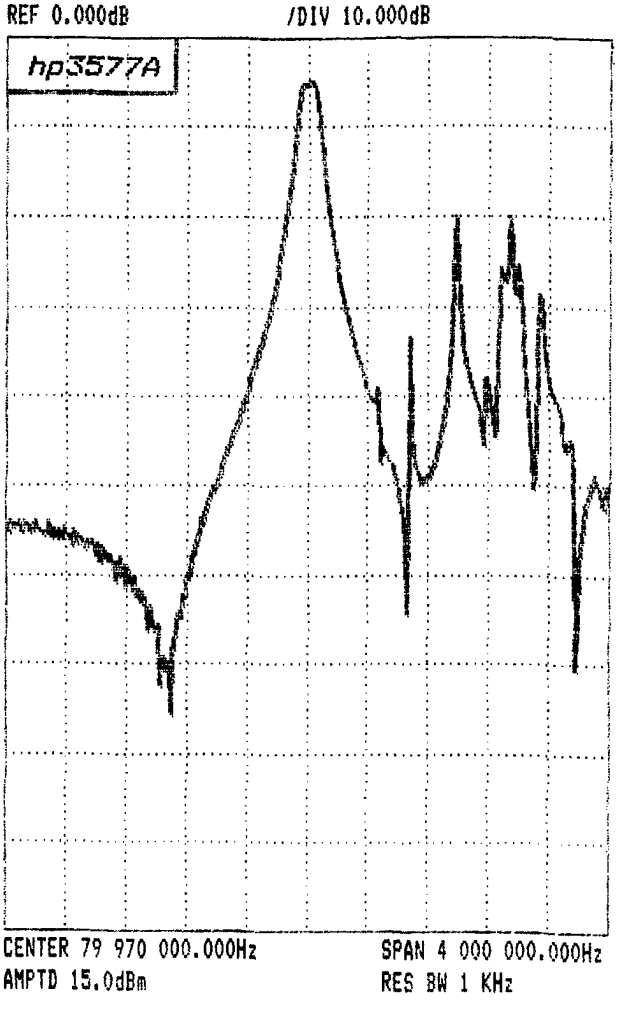


Fig. 6 Frequency response of monolitic LGS filter

Aknowlegement

The authors wish to thank A.N.Gotalskaya, S.Knyazev, B. Mill, and V.Lomonov for providing langasite samples and John R. Vig initiative and support this work.

References

1) S.A.Sakharov, I.M.Larionov and A.V.Medvedev. "Application of Langasite Crystals in Monolitic Filters Operating in Shear Modes" `1992 IEEE Frequency Control Simposium" pp713-723 , 1992.

2) Sally Laffey, Mary Hendrickson and John R.Vig
" Polishing and Etching Langasite and Quartz crystals", 1994 IEEE Frequency Control Simposium pp

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

New Data on Temperature Stabilyty and Acoustical Losses of Langasite Crystals

S.Sakharov*, P. Senushencov**, A.Medvedev*, Yu.Pisarevsky**

*FOCO Co., **Instute of Crystallography RAS

Abstract

Refined dates of temperature dependences elastic, dielectric , piezoelectric and thermal expantion coefficients were received at temperature range -100 - +100 °C.

Single and double rotated Z temperature cuts were calculated.

Q-factor of LGS resonators was measured in temperature range -200 - +100°C.

Introduction

Langasite crystals(LGS) are promising for FC applications due to high piezoelectric constants and low acoustic losses (1). Up to date several papers describing temperature dependences of elastic, piezoelectric and dielectric constants have been published (2 - 4), but considerable difference between calculated and experimental data have been noted. Also though acoustical losses of LGS on UHF less than quartz one quality factor of LGS notibly less than of quartz one.

In this work we made new measurements of temperature dependences of elastic, dielectric and piesoelectric constants in wide temperature range. Main difference from previous work in direct measurements of piezoelectric constants.Early dates were calculated from sound velocities.

By using experimental data of LGS material constants single and doubly-rotated zero temperature cuts have been calculated.

In accordance with this calculations resonators samples have been prepared and tested in wide temperature range.

Temperature measurements of LGS material constants

Temperature dependencies of elastic waves velocities in X,Y,Z and 45° to Y,Z directions were measured with phase-pulse method at 25 MHz frequency with accuracy 0.03%.Table 1 shows values of velocities and there temperature coefficients first and second order at room temperature with previous data.

Dielectric constants of LGS were obtained by measurenents of the capacity of X- and Z- cuts with automatic bridge at frequency 1 KHz with electric field on samples 150 V/cm.

Piesoelectric strain constants d_{ik} were measured by resonance-antiresonanse method.

Thermal expantion coefficients of LGS we measured using the modified capacitance method.

Elastic and piezoelectric stress constants

Dates obtained from above measurements allow us to calculate all independent elastic stiffness constants c_{ik} and piesoelectric stress constants e_{ik}. These are shown in Table 2 . We can see distinguished differense in temperature coefficients mainly piezoelectric constants.

0-7803-2500-1/95 $4.00 © 1995 IEEE

Table 1
Values of phase velocities and their temperature coefficients at room temperature.

	absolute quantities				temperature coefficients first order		second order	
	[*]	[1]	[2]	[4]	[*]	[1]	[*]	[1]
Waves	(m/s)				($10^{-6}K^{-1}$)		($10^{-9}K^{-2}$)	
VXL	5753	5746	5784	5748	-17	-24.9	-2.3	-24.7
VXSslow	2370	2394	2344		+50	+49.8	-110	-131
VXSfast	3317	3305	3344		-50	-53.2	-57	-39.1
VYQL	5762	5756	5792	5754	-32	-31.0	-64	-42.0
VYQS	3016	3003	2971	3015	-17	-18.8	-39	-57.5
VYS	2742	2750	2767	2745	+15	+6.6	-40	-89.0
VZL	6753	6826	6753	6750	-45	-50.0	-50	-29.2
VZS	3059	3043	3021	3061	-24	-26.5	-50	-64.9
VQL+45°Y,Z	5985	5949	5998		-29	-25.1	-17	-31.2
VQS+45°Y,Z	3228	3281	3141		-23	-52.2	-67	-87.0
VS +45°Y,Z	3291		3323		-50		-29	

Table 2
Values of elastic stiffness, piezoelectric stress constants, dielectric permittivity constants, thermal expansion coefficients and their temperature coefficients at room temperature.

Absolute quantities						Temperature coefficients First order ($10^{-6}K^{-1}$)				Second order ($10^{-9}K^{-2}$)		
Elastic stiffness ($10^{10}N/m^2$)	[*]	[1]	[2]	[3]	[4]	[*]	[1]	[3]	[4]	[*]	[1]	[4]
c_{11}^E	18.93	19.01	19.09	19.02	18.89	-53	-52.7	-47	-58.7	-4.2	-55.4	-113
c_{12}^E	10.50	10.30	10.63	10.63	10.46	-92	-107.4	-100	-218	+25	+48.1	+1085
c_{13}^E	9.528	9.524	10.42	9.19	9.68	-88	-14.4	-130	-101	-131	+156.8	-117
c_{14}^E	1.493	1.57	1.52	1.47	1.43	-205	-176	-370	-154	+870	-10.4	+452
c_{33}^E	26.24	26.83	26.19	26.21	26.22	-104	-105	-94	-135	-109	-62.8	+ 65
c_{44}^E	5.384	5.331	5.24	5.382	5.39	-62	-57.6	-30	-79.7	-111	-133	+102
c_{66}^E	4.216	4.35	4.32	4.20	4.22	-4.7	+12	+36	+139	-40.7	-178	-1564
Piezoelectric stress constants (C/m^2)												
e_{11}	-0.431	-0.41	-0.45		-0.44	456			-2400	1032		
e_{14}	0.108	0.08	0.077		0.07	-628			-5000	1480		
Dielectric permitivity constants												
$\varepsilon_{11}^T/\varepsilon_o$	18.97	18.89	18.99	18.86	18.86	137	153		447	82		-2490
$\varepsilon_{33}^T/\varepsilon_o$	52.00	49.0	49.32	48.7	49.10	-795	-760		-618	1076		-66.8
Thermal expantion coefficients												
α_{11}	5.07	5.11			5.84							
α_{33}	3.60	3.61			4.03							

Zero temperature cuts calculations

It is known, that calculations of 2nd and 3rd-order temperature coefficients of frequency (TCF) are connected with calculation of corresponding derivatives over temperature of a barge number of material constants.

Data on high-order temperature coefficients of these constants are often absent in the literature. Moveover, we consider that during calculation of zero TCFs with minimum values of higher-order TCFs for afixed operating temperature range, it is reasonable to take temperature coefficients of constants of corresponding order, calculated for a difinite material. As it is known, a great number of such ranges exist, and obtaining necessary data on temperature coefficients of constants is, apparently, very difficult.

However, data available to authors on absolute values of langasite material costants at different temperatures over the range from -100 °C top +100 °C, enable to consider the task of finding thermostable cuts from other positions.

Let us interest the frequency deviation at an arbitrary temperature with respect to a certain fixed value, spesified at detailed temperature. Then it is possible to write, that

$$\Delta f/f = (f_T - f_{27})/f_T \quad (1)$$

where f_T, f_{27} - are frequencies at temperatures T and T=27°, respectively.

For resonance frequency definition we use the transcendental equation:

$$tg(2\pi f^{(i)}t(\rho/c^{(i)})^{1/2} = 2\pi f^{(i)}t(\rho/c^{(i)})^{1/2}/K^2 \quad (2)$$

where $f^{(i)}$ - is the resonance frequency of the given mode;

t - is half thickness of a crystal element;

ρ - is LGS density;

K^2 - is the electromechanical coupling coefficient of the given mode;

$c^{(i)}$ - is the eigenvalue of Kristoffel matrix of a given mode.

Let us designate $2\pi f^{(i)} t (\rho/c^{(i)})$ through α and resolve (2) with respect to it.

By substituting the obtained value a in (1) we obtain:

$$\Delta f/f = \alpha_T/\alpha_{27} * t_{27}/t_T * * \{(c^{(i)}_T/\rho_T)/(c^{(i)}_{27}/\rho_{27})\}^{1/2} - 1 \quad (3)$$

where α_T ,α_{27} - are solution of trancedental equation (2) at instantaneous temperature T and T=27°C, respectively;

t_T ,t_{27} - is half thickness of a crystal element instantaneous temperature T and T=27°C, respectively.

The relation obtained in such a way anables to find the relative frequency deviation at a specifiad tewmperature without the use of any approximations connected with expanding temperature dependent of material constants in power series. The terms in (3) can be easile calculated for any angele of cut and taken directly from the experimental data.

Relative values of frequency diviations obtained for a given angle of cut over the temperature range from -100 C to+100 C can be represented in the form of a graph of $\Delta f/f$ versus temperature dependence .

For analysis of a larger number of data it is more convenient to use a traditionnal expansion of resonance frequency dependence vs resonance frequency in a power.

On Fig 1-3 represented dependences of electromechanical coupling coefficient, 1st TFC and 2nd TFC vs two angles rotation α and β. α is angle rotation in XY plane and

β is angle rotation in YZ plane.

Temperature characteristic of LGS resonators

Based on calculated data LGS resonators were preparated. Temperature dependences resonant frequency and quality factor were measured. Main parameters at room temperature represented in Table 3 for 1st and 3rd obertones. On Fig.4 temperature dependence of resonanse frequencies are given. Character dependency for 1st obertone are different compare with 3rd and 5th ones. Similiar behevior ones could see for quartz resonators. Frequency deviation for 1st obertone in temperature range - 20°C - +80°C about 1.6 ppm/° C. Experimental dates are in good agrement with calculated ones.

On Fig. 5 temperature dependeces of LGS resonators quality factor are represented for 3rd and 5th obertone. Ones could see fast growth of Q-factor in low temperature range. Q-factor for 5 th obertone grow up more than 5times to value 0.5×10^6 when temperature fall from +20 till - 80 °C. For quartz resonators q-factor changes much more weak. We suppose the reason such LGS behavior in it defects structure.

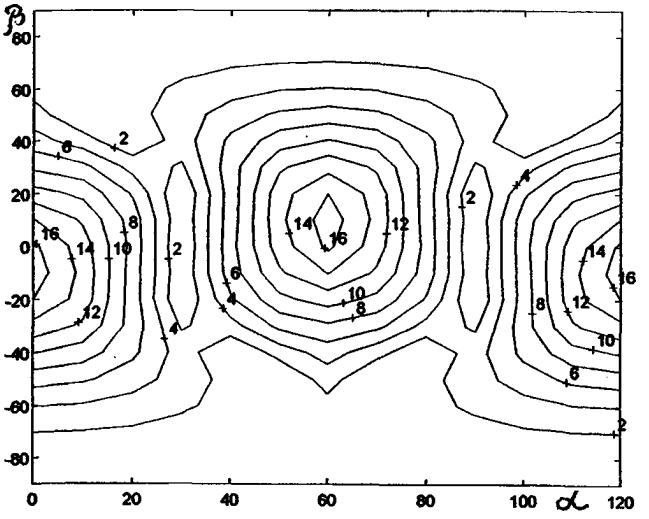


Fig. 1 Dependence of electromechanical coupling coeffcient vs angels α, β

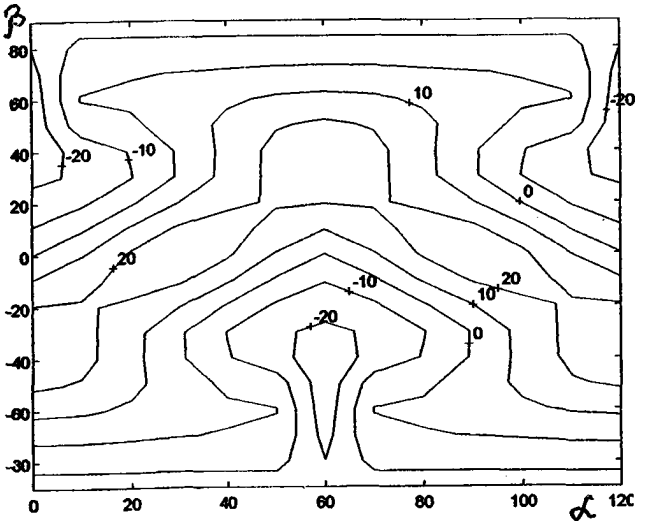


Fig. 2 Dependence of 1-st order TCF vs angels α, β

Conclusion

Refinment of elastic,dielectric and piezoelctric constants allowed calculated zero temperature compensated cuts which have good agrement with experimental dates. Such results connected not only with more precise measurement but with fact that sample for constants and resonators were produced from same crystall boule.

Dates of Q-factor temperature dependences show on reserve of improvement LGS resonators parameter.

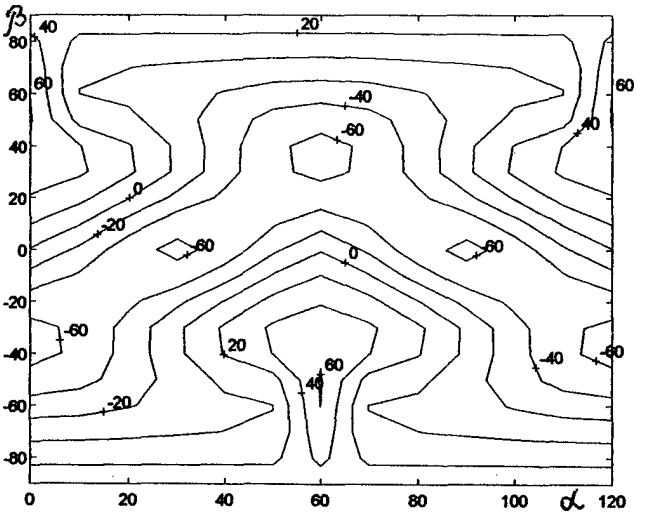


Fig. 3 Dependence of 3-th order TCF vs angels α, β

Resonator: LGS N Note: + 27

Num.	Fs, Hz	dFar, %	Q	R1, ohms	L1, H	C1, pF	C0,pF	Pr, mcW
1	3 002 404.41	0.5362	62 237	18.67	6.158E-2	4.563E-2	4.24	4.449
5	3 000 533.73	0.5288	74 227	15.57	6.129E-2	4.590E-2	4.33	4.045
7	3 000 433.12	0.5313	62 089	18.70	6.158E-2	4.569E-2	4.29	4.463
10	3 000 642.89	0.5296	75 776	15.65	6.290E-2	4.473E-2	4.21	4.056
11	2 999 210.77	0.5548	75 097	14.79	5.893E-2	4.779E-2	4.30	3.926
12	3 000 508.56	0.5351	51 587	22.63	6.193E-2	4.543E-2	4.23	4.873
14	3 002 656.40	0.5217	32 161	37.08	6.320E-2	4.445E-2	4.25	5.695
15	3 000 851.31	0.5472	47 496	23.51	5.923E-2	4.749E-2	4.33	4.947
20	3 000 259.30	0.5406	80 935	14.16	6.079E-2	4.629E-2	4.27	3.832
21	3 002 175.44	0.5103	108 325	10.47	6.013E-2	4.674E-2	4.57	3.155

Resonator: LGS N Note: + 27

Num.	Fs, Hz	dFar, %	Q	R1, ohms	L1, H	C1, pF	C0,pF	Pr, mcW
1	8 996 084.07	8.522E-2	136 150	17.96	4.325E-2	7.237E-3	4.24	4.355
5	8 992 114.48	8.351E-2	117 204	20.86	4.327E-2	7.239E-3	4.33	4.677
7	8 991 019.99	8.420E-2	109 226	22.45	4.340E-2	7.219E-3	4.29	4.818
10	8 992 306.52	8.385E-2	137 128	18.27	4.433E-2	7.066E-3	4.21	4.393
11	8 987 109.85	8.093E-2	127 366	20.00	4.511E-2	6.952E-3	4.29	4.587
12	8 990 219.26	0.1030	85 803	23.59	3.583E-2	8.748E-3	4.24	4.909
14	8 995 320.05	7.639E-2	110 322	24.70	4.821E-2	6.494E-3	4.25	5.004
15	8 993 239.25	8.181E-2	118 474	21.06	4.416E-2	7.092E-3	4.33	4.695
20	8 991 743.62	8.480E-2	119 850	20.42	4.332E-2	7.231E-3	4.26	4.634
21	8 994 143.69	7.437E-2	129 485	19.93	4.568E-2	6.856E-3	4.61	4.580

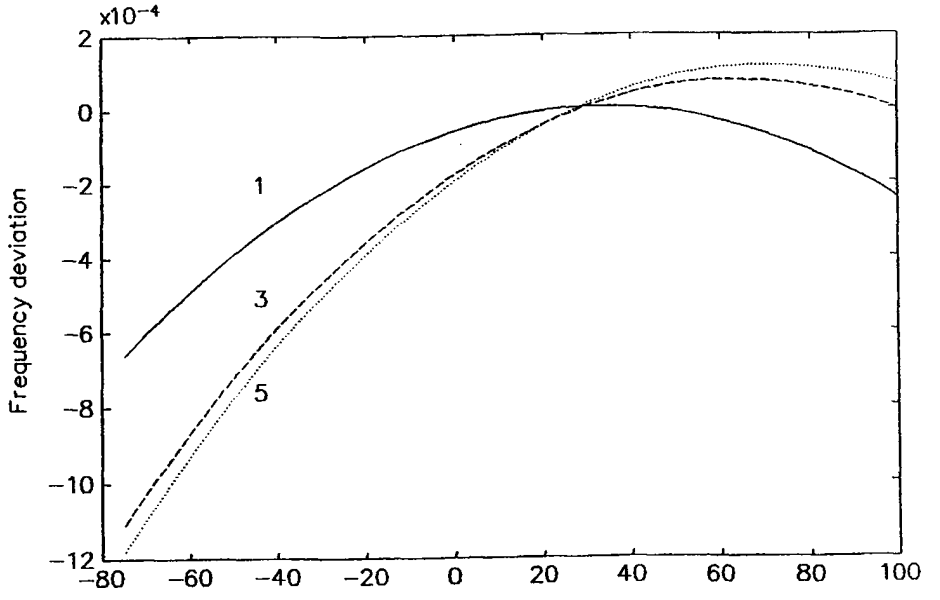


Fig. 4 Dependence of frequency deviation vs temperature

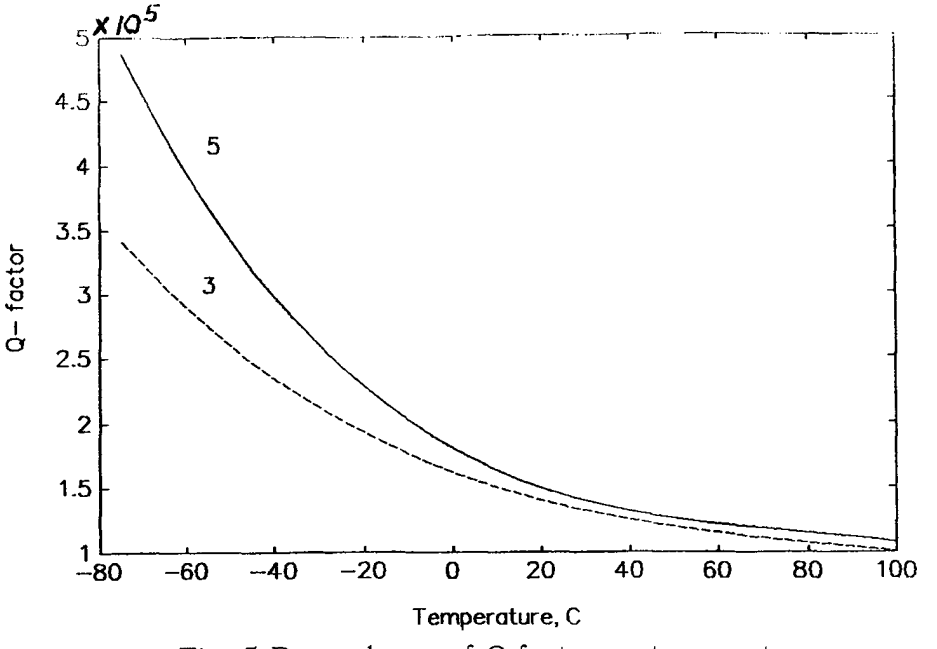


Fig. 5 Dependence of Q-factor vs temperature

References

1.A.A.Kaminskii,I.M.Silvestrova ,S.E.Sarkisov,G.A.Denisenko.
Phis. Stat. Sol. (a) v80, 1983 p 607

2.I.M.Silvestrova,P.A.Senushenkov,
Yu.V.Pisarevsky, A.I.Krupnii
Sov.Phys. Sol. state v28, 1986 N9 ,p2875

3.I.M.Silvestrova,P.A.Senushekov,V.Bezdelkin,Yu.V.Pisarevsky
Proc. of the 1983 IEEE Frequency Control Simposium 1993 p.348

4.A.B.Ilyev,B.S.Umarov,L.A.Shabanova,
M.F.Dubovic
Phis. Stat. Sol. (a) 1986 K109 p.98

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

NEW STRONG PIEZOELECTRIC $La_3Ga_{5.5}Nb_{0.5}O_{14}$ WITH TEMPERATURE COMPENSATION CUTS

Yu.V.Pisarevsky, P.A.Senushencov, P.A.Popov B.V. Mill,

Institute of Crystallography RAS, 59 Leninsky prosp. 117333 Moscow Russia

Abstract

Temperature dependences of various important for FC applications parameters of $La_3Ga_{5.5}Nb_{0.5}O_{14}$ (LGN) single crystals were measured. These are sound velocities within range 77-370°K, dielectric permittivities within 4.2-400°K, thermal conductivity within 50-300°K. From above dates temperature dependences full set of elastic and piezoelectric constants were calculated. Results shows on existence of cuts with zero temperature coefficients of resonant frequencies and sound velocities. We found anomaly in derivative of ε_{33} near 50°K also. Thermal conductivity behavior shows on disorder in crystalic structure.

Introduction

There are growing interest to $La_3Ga_5SiO_{14}$ (LGS) single crystals for FC applications, because of high piezoelectric properties, zero temperature cuts existences, low acoustical losses. Up to date were synthesized above 30 compounds with similar structure [1]. For some of them were measured elastic and piezoelectric properties at room temperature (Silvestrova et.al. [2,3]). But temperature dependences of material constants were studied only for LGS. We started such investigations from LGN single crystals because of piezoelectric constants LGN larger than LGS ones and LGS single crystals could be grown with high quality. First of all were investigated temperature dependences of elastic , piezoelectric, dielectric and thermal expansion coefficients. All of them needed for zero cuts of resonant frequencies and surface acoustic waves calculations.

High frequency applications restricted by acoustical wave attenuation. Nature of these losses depend on thermal fonon relaxation. So we measured temperature dependences of thermal conductivity of LGN .

Crystals were grown by Czochralski technique in flow of 3% O_2 - 97% N_2 mixture at the temperature 1470°C. Crystals had 25 mm diameter and 50 mm length.

Table 1.
Values of phase velocities and their temperature coefficients at room temperature.

Waves	absolute quantities (m/s)		temperature coefficients first order ($10^{-6}K^{-1}$)		second order ($10^{-9}K^{-2}$)	
	NGL	SGL	NGL	SGL	NGL	SGL
VXL	5684	5753	-36	-17	-51	-2.3
VXSslow	2293	2370	83	+50	-39	-110
VXSfast	3178	3317	-64	-50	-340	-57
VYQL	5684	5762	-36	-32	-36	-64
VYQS	2882	3016	1.6	-17	-237	-39
VYS	2665	2742	17	+15	-48	-40
VZL	6625	6753	-47	-45	-6.6	-50
VZS	2916	3059	-18	-24	-466	-50
VQL+45°Y,Z	5850	5985	-24	-29	-47	-17
VQS+45°Y,Z	3154	3228	-41	-23	-64	-67
VS +45°Y,Z	3155	3291	-40	-50	-74	-29

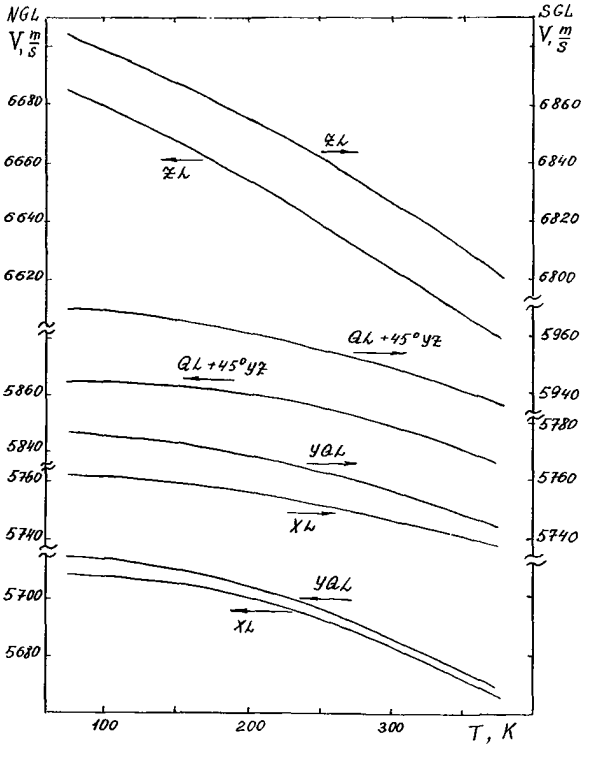


Fig.1 Temprtature dependencies of velocities of longitudinal and quasi-longitudinal acoustic waves of LGN and LGS crystals.

0-7803-2500-1/95 $4.00 © 1995 IEEE

Sound velocities

Temperature dependencies of elastic waves velocities in X, Y, Z, and along 45° to Y,Z crystallographic directions were measured with phase-pulse method at 25MHz frequency with accuracy 0.03%. A system of regulation and automatic maintenance of temperature provided a temperature interval between 77 and 375°K with an

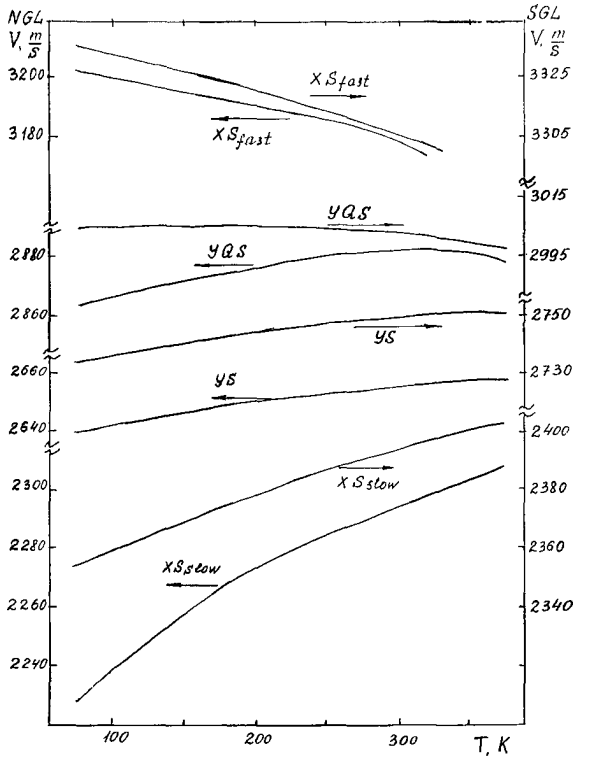


Fig.2 Temprtature dependencies of velocities of shear and quasi-shear acoustic waves of LGN and LGS crystals along X- and Y-axis directions.

accuracy and uniformity of 0.01°. Y-cut and X-cut quartz plates were used as transducers of shear and longitudinal waves respectively. For investigating of elastic properties several samples ranging from 7 to 12 mm in length were used. Samples with small internal stress were chosen. Flat parallelism of the opposite faces was brought up to 0.001 mm by fine lapping or polishing. The accuracy of the face orientation was 20' checked by X-ray diffraction.

The temperature dependences of the elastic waves velocities are shown in Fig.1,2,3 together with dates for LGS. Table 1 shows values of velocities and there temperature coefficients first and second order at room temperature with same data for LGS. We take note of existence of some elastic waves with positive temperature coefficients. Two of them similiar LGS. One is piezoactive pure shear Y-mode, another is slow shear mode along X-axis. Above these is arrived third . It is quasi-shear mode along Y-axis. So the cuts with zero temperature coefficients of some acoustic waves and resonant frequencies are exist.

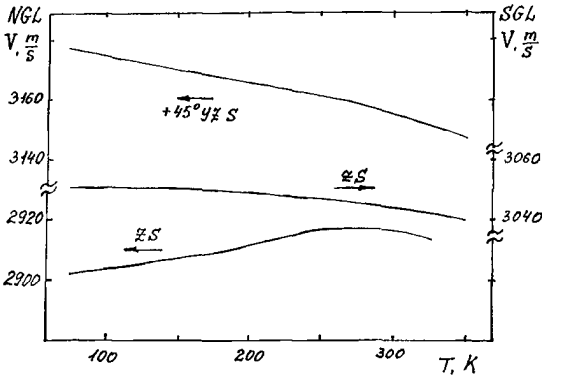


Fig.3 Temprtature dependencies of velocities of shear and quasi-shear acoustic waves of LGN and LGS crystals along Z- and 45° to Y and Z-axis directions..

Dielectric permittivity

Dielectric constants of LGN were obtained by measurements of the capacity of plates of X- and Z-cuts with an automatic AC bridge at frequency of 1 kHz with an electric field in specimen of 100 to 200 V/cm. In the experiment the dielectric loss tangent was also determined. Temperature in dielectric experiments changes in interval between 4.2 and 400°K. Temperature dependences of dielectric constants are given in Fig.4,5. Values of dielectric constants are larger the same of LGS (Tab 2.), particularly ε_{33}. We take note rather big value

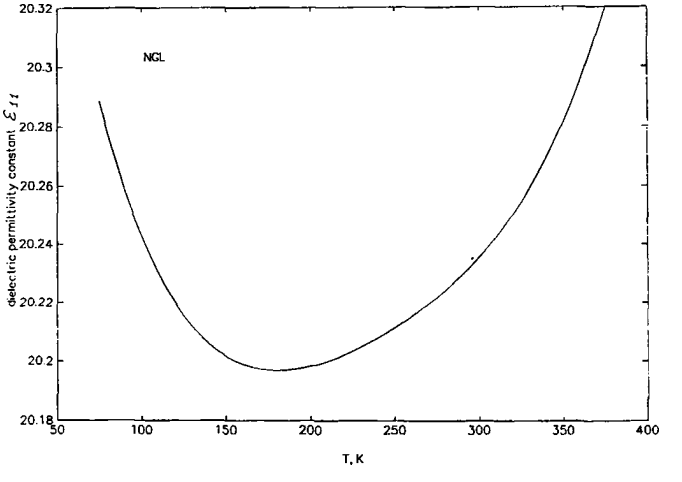


Fig.4 Temprtature dependence of dielectric permitivity ε_{11} of LGN crystals.

and quick increasing till 50°K of e_{33} with temperature decrees and very week temperature dependence below 20°K.

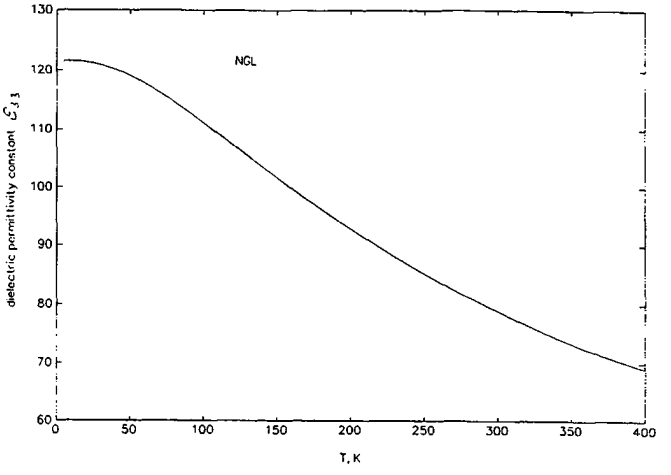


Fig.5 Temprtature dependence of dielectric permitivity ε_{33} of LGN crystals.

Thermal expansion

We measured thermal expansion coefficients of LGN using the modified capacitance method in temperature range 77-375°K. Dates obtained are shown in Fig.6,7. Thermal expansion of LGN have no any anomalies.

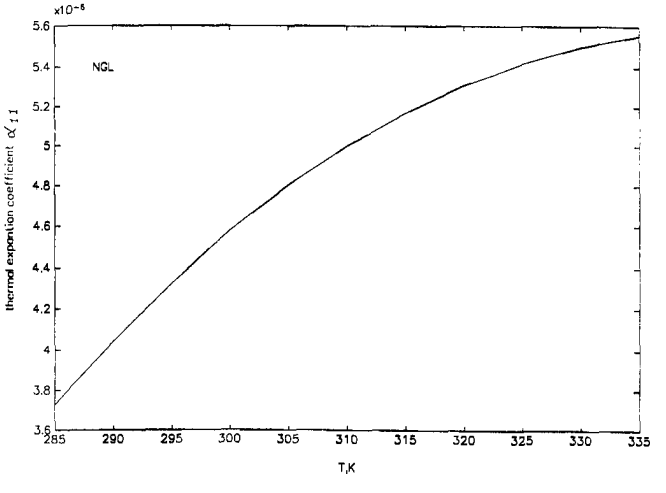


Fig.6 Temprtature dependence of thermal expansion coefficient α_{11} of LGN crystals.

Temperature dependences of thermal expansion coefficients are rather week. Absolute values are smaller then for LGS.

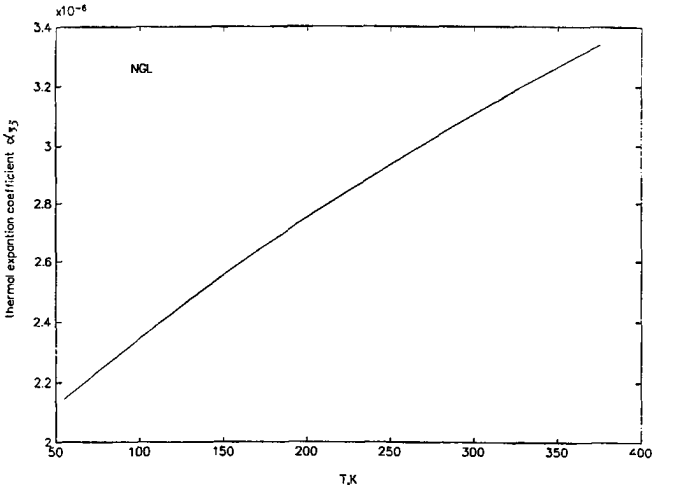


Fig.7 Temprtature dependence of thermal expansion coefficient α_{33} of LGN crystals.

Table 2.
Values of elastic stiffness, piezoelectric stress constants, dielectric permittivity constants, thermal expansion coefficients and their temperature coefficients at room temperature.

Absolute quantities			Temperature coefficients			
			First order		Second order	
Elastic stiffness (10^{10}N/m^2)			(10^{-6}K^{-1})		(10^{-9}K^{-2})	
	NGL	SGL	NGL	SGL	NGL	SGL
c_{11}^E	19.07	18.93	-77	-53	41	-4.2
c_{12}^E	10.93	10.50	-112	-92	6.4	25
c_{13}^E	9.82	9.528	-84	-88	348	-131
c_{14}^E	1.35	1.493	-502	-205	-850	870
c_{33}^E	26.05	26.24	-108	-104	-22	-109
c_{44}^E	5.04	5.384	-50	-62	-942	-111
c_{66}^E	4.07	4.216	-29	-4.7	885	-40.7
Piezoelectric stress constants (C/m^2)						
e_{11}	-0.44	-0.431	750	456	-4500	1032
e_{14}	0.05	0.108	39500	-628	-83900	1480
Dielectric permitivity constants						
$\varepsilon_{11}^T/\varepsilon_0$	20.2	18.97	30	137	180	82
$\varepsilon_{33}^T/\varepsilon_0$	79.4	52.00	-1500	-795	2360	1076
Thermal expantion coefficients (10^{-6}K^{-1})						
α_{11}	4.32	5.07				
α_{33}	3.09	3.60				

Elastic constants

Dates obtained allow us to calculate all independent elastic stiffness c_{ik} and piezoelectric stress e_{ik} , represented on Fig.8, of LGN. These results are shown in Table 2. Calculated temperature coefficients of first and second order for all obtained constants are also represented in this table. Table 2 allow us to compare elastic, piezoelectric, dielectric and thermal expansion properties LGN and LGS.

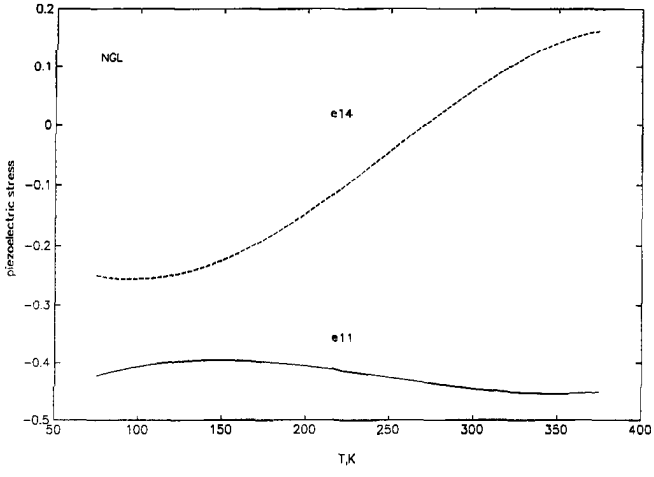


Fig.8 Temprtature dependencies of piezoelectric stress e_{ik} of LGN crystals.

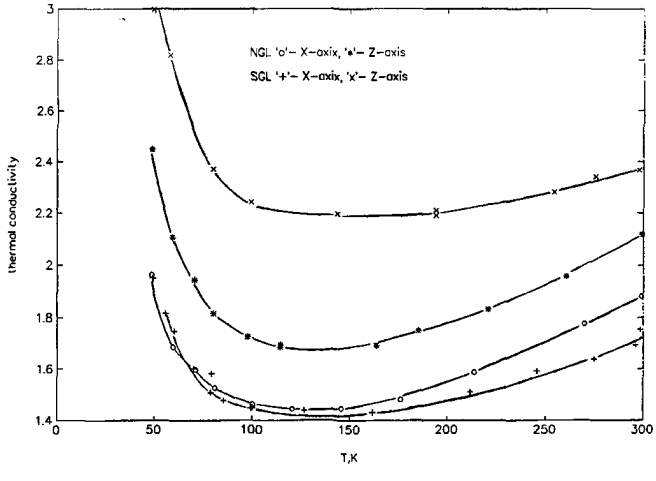


Fig.9 Tempertature dependencies of thermal conductivity coefficients κ_{ik} of LGN crystals.

Thermal conductivity

To explain the Nature of acoustic losses in LGN and LGS crystals and for all the same structure crystals family also, we need to study thermal fonon relaxation. For this purpose we measured thermal conductivity coefficients. Temperature range was from 50 to 300°K. We used method of stationary heat flow. Specimens in form of bar with length along X- ore Z-axis has cross section 3.7x3.7mm^2. The space between temperature sensor was 8 mm. Accuracy of measurements was nearly 7%. Results obtained are shown on Fig.9. We note very low (for example in comparison with SiO_2) value of thermal conductivity, which is equal 1.9 $Wm^{-1}K^{-1}$ along X-axis and 2.1 $Wm^{-1}K^{-1}$ along Z-axis for LGN. Temperature dependence of thermal conductivity both crystals is rather week. Anizotropy coefficient of thermal conductivity of LGN inside investigated temperature range changes from 1.3 at 50°K to 1.1 at 300°K.

We note behavior of temperature dependence of thermal conductivity iboth crystals is common to disorder structures but not for crystals.

Conclusion

Above dates shows that LGN have good perspective for frequency control applications because of high quality crystals could be grown , first order zero temperature cuts are available. If one compare with LGS crystals he find that LGN properties slightly favorable because of more piezoelectric constants , less thermal expansion coefficients.

LGN is the second compound from langasite family with thermocompensated cuts. So we hope that other compounds from this family also will have such usefully properties. But tendency of properties variations ones would understand after additional dates about other compounds from this family. We suppose at near future to receive new dates about temperature properties other LGS family crystals

Acknowledgement

Authors wish to thank International Soros Foundation for partly support of this work.

References

1.B.V.Mill, A.V.Butashin, G.G.Khodzhabagyan, E.L.Belokoneva, N.V.Belov. "Modifyed Rare-earth Galates with $Ca_3Ga_2GeO_4$ Structure" Sov.Phis. Doklady, 1982, V.264, N6 , pp. 1385 -1389.

2. I.M.Silvestrova, Yu.V.Pisarevsky, B.V.Mill, A.A.Kaminsky, "Acoustic and Electromechanical Properties of Piezoelectric Crystals with Ca-galogermanate Structure" Sov. Phys. Doclady 1984, V. 266 N5 pp. 574 - 578.

3. I.M.Silvestrova, Yu.V.Pisarevsky, B.V.Mill, "Elastic, Piezoelectric and Dielectric Properties of $La_3Ga_{5.5}Nb_{0.5}O_{14}$ Crystals". Sov.Phys. Solid State, 1987 V.29, N5, pp. 1520 - 1522.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

LANGASITE CRYSTAL QUALITY IMPROVEMENT AIMED AT HIGH-Q RESONATORS FABRICATION

Gotalskaja A.N., Dresin D.I., Schegolkova S.N.,
Saveleva N.I., Bezdelkin V.V., Cherpoukhina G.N.

"LANTAN" Co. Ltd., 44, Krasnobogatyrskaya, Moscow, 105023, Russia

ABSTRACT

Processes of langasite crystal growth are considered. Heat assembly design, pulling and rotation rates are optimized. Composition of charge, impurity content and technology of its manufacture have been studied. Influence of annealing regimes on crystal quality are discussed. Design and technology investigations of langasite resonators for crystal material quality factor evaluation is reported. Langasite resonator at 5 MHz, 3rd overtone has been developed.

INTRODUCTION

Gallosilicate of lanthanum ($La_3Ga_5SiO_{14}$) is a promising material for piezoelectric engineering, because it possesses unique piezoelectric parameters.

In connection with new requirements to LGS crystal quality improsed by their application not only in monolithic filters, but also in high-quality resonators, we have investigated the influence of various technological parameters on LGS crystal quality.

INVESTIGATION OF TECHNOLOGICAL REGIMES INFLUENCE ON CRYSTAL QUALITY

For optimization of the heat assembly design we investigated the influence of temperature field distributions on the crystal quality. Crystals have been grown up at different values of axial temperature gradient above the melt ranging from 8 to 40 °C/cm, with constant values of rotation and pulling rates.

Axial gradient variation was attained by design variations of upper heat assembly part, consisting of platinum and ceramic screens. Crystals grown with axial temperature gradients over the melt above 40 °C/cm contained macroblocs, within 25-40 °C/cm they cracked during cutting. Langasite suitable gradients values are below 25 °C/cm.

Initially, we have grown langasite crystals with the pulling rate of V=3-3.5 mm/h. However, under these conditions "as grown" crystal were stressed, had a considerable impurity content and a pronounced microblock structure. Therefore, we have lowered the pulling rate up to 1,5-2,5 mm/h. With lower rates the crystal quality did not change, and the process became less productive.

Pulling rate change in crystals caused the necessity of rotation rate change for keeping crystallization front shape. With these pulling rates, the permissible rotation rate range turned out to be narrower than with V=3 mm/h. For example, rotation rate increase for one rotation changes crystallization front shape from a strongly convex to a strongly concave one.

Hence, pulling rate and, subsequently, rotation rate reducing allowed us to limit impurity content in crystals by 1.5 times (see table 1) and to reduce both microblock structure and elementary cell parameter spread. However, this did not help to remove the central non-homogeneous stressed region, which is present in crystals.

For improving conical crystal part we have also selected rotation rates at the cone. The cone grows in the preinversion regime and has a large number of bubble inclusions in it, which can be inherited by the growing cylindrical part of crystal. For suppression of natural convection the growth of crystal began with higher rotation rate, which then gradually lowered along the cone length. This enabled us to remove bubble inclusions (knags) from the cone, while the growth column remained within the area of the crystal conical part All this had no significant influence on non-homogeneously stressed region presence.

The effect of additional crystal annealing on crystal quality was studied. At the initial stage of investigations "as grown" crystals were annealed in a crystallization chamber after growth process during one day. Crystals, additionally annealed under 1100-1400°C, have shown a more pronounced pattern in crossed polaroids and even crystals with strong stresses did not crack during cutting. Crystals annealed at $t>1100$ °C changed their colour (rose shade disappeared) and became yellower. This fact has been fixed during absorption spectrum

0-7803-2500-1/95 $4.00 © 1995 IEEE

measurement by using spectrophotometer over the wavelength range of 350-2400 nm (Fig. 1, 2). As one can see from the figure, absorption over the range of 800-2400 nm is absent. The crystal color is caused by absorption in the visible region of the spectrum (400-700 nm before annealing and 400-600 nm after annealing). However, the high-temperature annealing did not allow us to reduce significantly non-homogeneous stressed region available in crystals.

"As grown" crystals were tested for the impurity content. For charge obtaining we have used the following reactives as initial materials: lanthanum oxide (La_2O_3) with the purity of 99.9%, silicium oxide (SiO_2) - 99.9% and gallium oxide (Ga_2O_3) - 99.9%. The nature of impurity distribution in the crystal was investigated. Impurity quantity was determined by direct atomic-emission analysis and spectrum excitation in d.c. arc with recording by DCC-8 spectrometer with the reticule of 600 lines/mm. Table 1 gives the data on impurity content in the melt before growth (Sample № 1), after the growth (Sample № 2) and in two crystal samples grown up from this melt (Samples № 3, 4)

As one can see from the data listed in the table, the impurities are being concentrated in the melt during the growth process (ferrum oxide content increases two times, especially). Apparently, due to this the rests of the melt used have a dark brown color. Samples № 3 and № 4 differed by the pulling rate: 3 and 1.5 mm/h, respectively. As it is evident, sample № 4 has a lesser impurity content. Judging by the results obtained, the impurity distribution coefficients for sample № 4 are the following:

$K_{Al_2O_3} = 0.196$ $K_{TiO_2} = 0.121$ $K_{Fe_2O_3} = 0.262$

5 MHz resonators have been fabricated from samples № 3 and № 4 by using technological process and design described below. The quality factor Q of resonators fabricated from sample № 3 did not exceed $(12\text{-}15)\times10^3$, and the quality factor Q of resonator fabricated from sample № 4 was $(25\text{-}30)\times10^3$.

Crystals grown during the last five runs were tested for their general composition by using spectrophotometer Link 10000 and for impurity content by using Camebax SX-50. The following results have been obtained:

A. The impurity content in crystals was the following:

Al < 0.015 % weight
Zr < 0.01%
Fe < 0.01 - 0.0005 %
Ce was not observed
Nd < 0.005 %
Pt < 0.003 %

B. Crystals grown up from the charge of stoichiometric composition can differ considerably from the stoichiometry of the composition. It has been discovered that langasite has a sufficiently wide region of homogeneity. For example, a single crystal of composition $La_{3.27}Ga_{4.9}Si_{0.87}O_{14}$ has been grown up. This fact explains the existence of non-homogeneous region in crystals. It is caused by impurity phases present in the melt, which are trapped by a growing crystal. This trapping of impurity phases increases at poor melt mixing. The presence of chemical composition spread in crystals has been confimed by optical analysis methods by measuring specifi rotation (optical activity) and residual light flux. It was observed that during crystal growing from a non-stoichiometric composition (deviation of 3 weight %) concentration of impurity phases in the melt and deterioration of crystal quality, can occur (Table 2).

C. A new additional crystal phase with greater Ga content has been discovered for the first time. It appears sometimes as a small white thin coating at the crystal surface at a place of tearing off crystal from the melt and is presumably present in the melt. By using Link 10000 instrument, this crystal phase has been identified as $La_{2.56}$ $Ga_{5.8}$ $Si_{0.72}$ O_{14}. Up to now langasite was considered to be a unique triple compound in the phase system La_2O_3-Ga_2O_3-SiO_2 (Fig. 3) [1-7]. Therefore, LGS-phase diagramme needs to be investigated in more detail.

We have also tested the charge produced at the plant GIREDMET (Podolsk), obtained by using the method of self-distributing high-temperature synthesis (SHS). The charge has been synthetisized at 15% of metal gallium ignition, contained in the mixture of gallium, silicium and lanthanum oxides. However, it also differed from the stoichiometry in Ga content, its general composition corresponded to the formula $La_{3.05}$ $Ga_{4.75}$ $Si_{1.18}$ O_{14} and it contained a large number of impurities (Fig. 4, 5):

Fe ≈ 0.1 % weight
Ca ≈ 0.02
Zr ≈ 0.01
Y ≈ 0.005

Therefore we had refused to use it and began develoment of our technology with testing at all stages up to growing crystal. We have also refused to use a charge obtained as a result of recrystallization because it is difficult to control stoichiometric composition.

For crystal quality evalution the chemical etching the solvents described below were used. Real structure elements (macro- and microdefects) have been identified which can be inherent to langasite

Table 1
Investigation results for impurity content in the melt and in LGS-crystals

Impurity	Sample № 1	Sample № 2	Sample № 3 -	Sample № 4
Al_2O_3	$2.8x10^{-2}$	$5.7x10^{-2}$	$2x10^{-2}$	$5.5x10^{-3}$
TiO_2	$1.4x10^{-3}$	$2.2x10^{-3}$	$1.4x10^{-4}$	$1.7x10^{-4}$
Fe_2O_3	$8.4x10^{-3}$	$1.7x10^{-2}$	$3.2x10^{-3}$	$2.2x10^{-3}$
CaO	$5.6x10^{-3}$	$8.4x10^{-3}$	$1.4x10^{-3}$	$< 1.4x10^{-3}$
CeO_2	$< 1x10^{-2}$	$< 1x10^{-2}$	$< 1x10^{-2}$	$< 1x10^{-2}$
PtO_2	$< 3x10^{-4}$	$< 3x10^{-4}$	$< 3x10^{-4}$	$< 3x10^{-4}$

Table 2
Homogeneity region test results at 1400°C

Composition, mol % La_2O_3 : Ga_2O_3 : SiO_2			Impurity phases
30	50	20	-
28	51.5	20,5	$Ga_2O_3 + La_2Si_2O_7$
29	49	22	$Ga_2O_3 + La_2Si_2O_7$
30	48	22	$La_2Si_2O_7$
31.7	48.3	20	$LaGaO_3 + La_{14}Si_9O_{39}$
30	52	18	$LaGaO_3$

Table 3
Results of surface roughness Ra measurements for langasite and quartz after identical mechanical treatment at the flat cast iron disk.

Abrasive powder type	Mean grain size of micropowder μm	Mean surface roughness R_a ($R_a \approx 0.2R_z$) Langasite	AT-cut quartz
M28	28-20	0.69	0.28
M14	14-10	0.37	0.18
M10	10-7	0.31	0.14
M7	7-5	0.26	0.12
M5	5-3	0.18	0.10

crystal [8]. The main microdefects are:

— dislocation slipbands,
— places of dislocation exit at the surface of Y- and Z-cut plates,
— inclined dislocations,
— screw dislocations,
— platinum inclusions,
— point defects.

Measurements of elementary cell parameters have been made by X-ray goniometer "DRON-3" from crystal grown up under different technological regimes Z-cut plates have been sawn with the thickness of 10 mm and Y- and Z-planes oriented with the accuracy of 10'. Crystal grown with pulling rates above 3 mm/h had a-parameter — spread of $1x10^{-4}$Å, but greater spread in c-parameter $1x10^{-3}$Å. These c-parameters values increased along the crystal length. Plates cut from less defective regions of crystal grown with pulling rate 1.5-2 mm/h had the following cell parameters:

$a=8.1684\pm1x10^{-4}$Å and $c=5.0975\pm1x10^{-4}$Å.

Resonators have been later manufactured of these plates.

DESIGN AND TECHNOLOGY INVESTIGATIONS OF LANGASITE RESONATORS FOR CRYSTAL MATERIAL QUALITY FACTOR EVALUATION

The parameter most completely characterizing resonator quality is its quality factor Q. We have manufactured evacuated langasite resonators with plano-convex piezoelectric elements of Y-cut operating at 5 MHz, 3rd overtone.

Due to drastical difference of langasite from quartz in major physical-chemical properties: hardness, chemical resistance to acids and alkalis, chemical activity of the surface etc., manufacturing processes commonly used for quartz resonators [9] required improvement for langasite resonators taking into account these properties.

Electrical parameters and quality factor Q, in particular, of piezoelectric resonators depend to a considerable extent on the quality of mechanical surface treatment of crystal elements (CE). Experiments have shown that langasite is more sensitive to mechanical treatment compared to quartz, for example, the rate of langasite material removal and the mean surface roughness after lapping is 2-3 times greater compared to quartz. Table 3 gives results of surface roughness Ra measurements for langasite and quartz after identical mechanical treatment at the flat cast iron disk.

For removal of layer, disturbed by mechanical treatment, a method of deep chemical etching to the depth of about 15 μm was used. Chemical etching process of langasite crystal elements is obligatory between lapping and polishing processes. However, the application of quartz etching technology in HF solutions to langasite leads to formation of a white loose deposit at its surface, necessitates the use of new etchants.

It is also known that lanthanum and gallium salts: fluorides, chlorides, carbonates, phosphates, oxalates etc. are poorly water-soluble and they can form deposits at langasite surface treated by corresponding acids under ordinary conditions. The influence of the these compositions on deposit formation at the plate surface, on etch rate and surface roughness was tested. The use of HF, H_2SO_4, NH_4F + HF and also of the solution $K_2Cr_2O_7$ in H_2SO_4 for langasite etching causes chemical reaction with lanthanum and gallium and formation of a white deposit at langasite surface. This deposit is easily scraped, by uncovering smooth surface of a langasite plate. For example, langasite etching at the room temperature during 10 min causes formation of non-solid thin deposit layer at the surface of plates. Heating of solution up to 60-80°C causes formation of a thick loose deposit layer at the surface of a plate during the same period of time. These reactions were particularly well observed at the surface of polished plates.

Langasite dissolves in solutions on the basis of HCl, HF and HNO_3 without deposits. This is explained by a good solubility of La- and Ga-fluorides in these acids. Etch rate, surface quality (roughness after etching) and shape of etch figures change depending on the composition and concentration of such solutions.

Most quickly, approximately 2-3 μm/min, langasite dissolves in hot solutions of HCl and in solutions with HF content not over 6-8%. Surface quality of plates becomes worse and its microrelief reminds "cobble-stone" structure, which is characteristic for solutions, containing HCl.

Langasite etch rate in HNO_3 is approximately three times less, than in HCl the surface quality degrades considerably less and microrelief reminds narrow long wedges, stretched along the Z-axis. Addition to HNO_3 of even a small quantity of HF leads to formation of a deposit at the crystal plate surface. Addition of HF to HNO_3 causes etch rate increase, but deteriorates the surface quality (Fig. 6). Addition of H_2SO_4 to the compositions described above causes usually microrelief deterioration of the

Table 4

Etch rates and surface quality evaluations after etching in several solution in the the vicinity of their boiling point

Solution composition	Mean etch rate, μm/min	Microrelief better (+) worse (-)
$HCl : H_2O = 1 : 2$	2-3	- -
HCl + 5%HF	5-6	-
$HNO_3 : H_2O = 1:2$	0.6-0.7	+
$HNO_3 : HCL = 1:3$	1-2	-
$HNO_3 : CH_3COOH = 2:1$	0.1-0.2	+ +

Table 5

Langasite resonator parameters

Resonance frequency f_r (Hz)	5002838
Quality factor Q, 10^3	115
Motional resistance R_1 (Ω)	77.8
Motional inductance L_1 (H)	2.83×10^{-1}
Motional capacitance C_1 (pF)	3.57×10^{-3}
Static capacitance C_o (pF)	6.35
Capacitance ratio C_o/C_1	1.779×10^3
Dissipation power P_{diss} (μW)	9.92

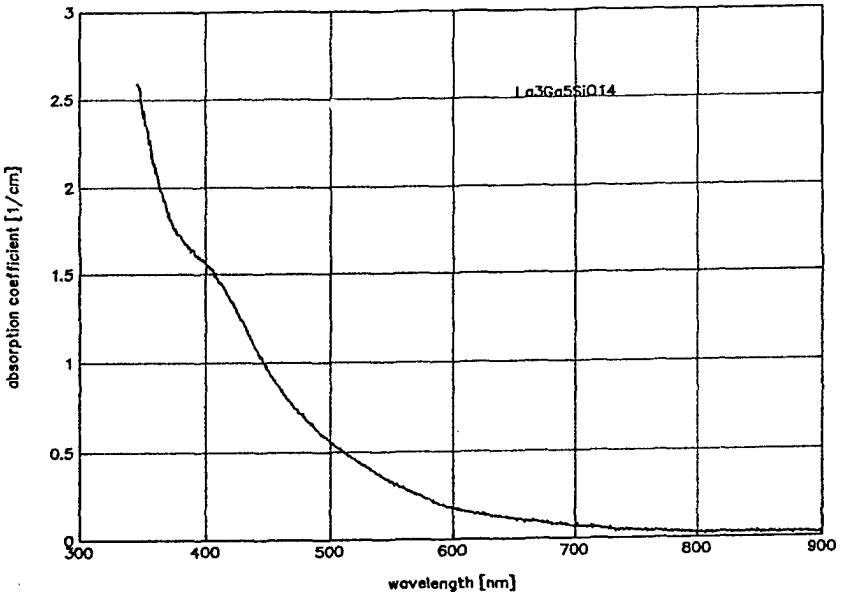


Figure 1. Absorption spectrum measurement over the wavelenght range of 350-2400 nm before annealing

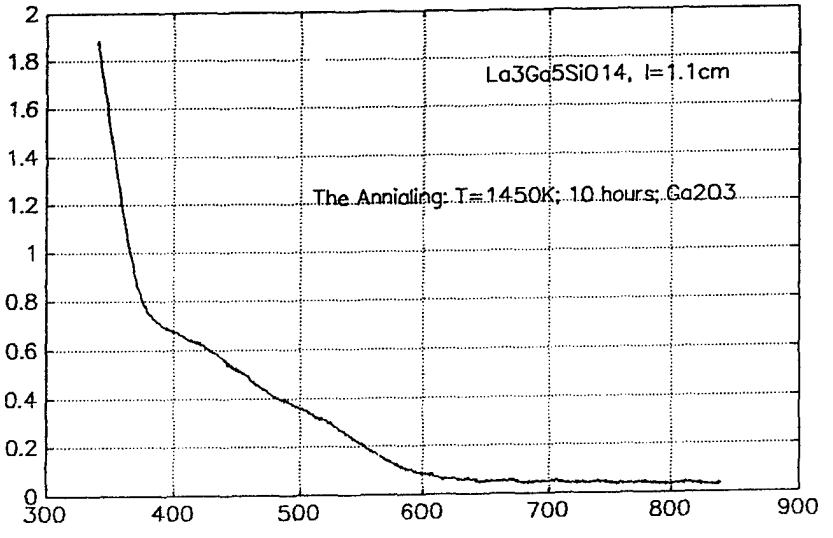


Figure 2. Absorption spectrum measurement over the wavelenght range of 350-2400 nm after annealing

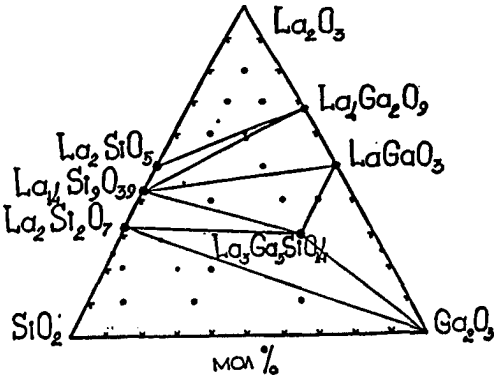
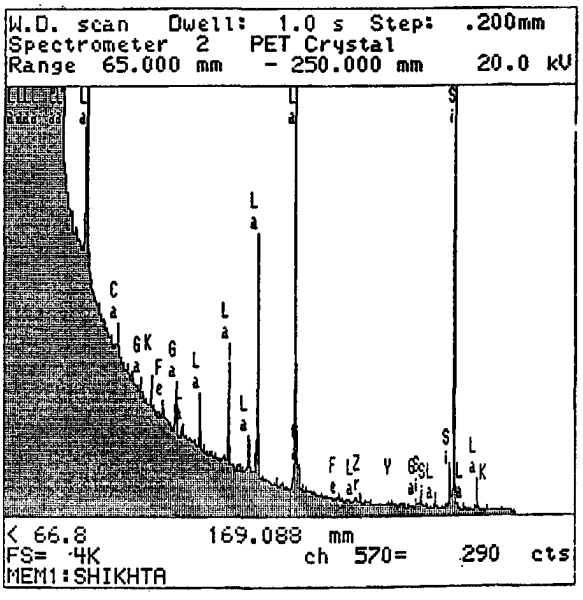
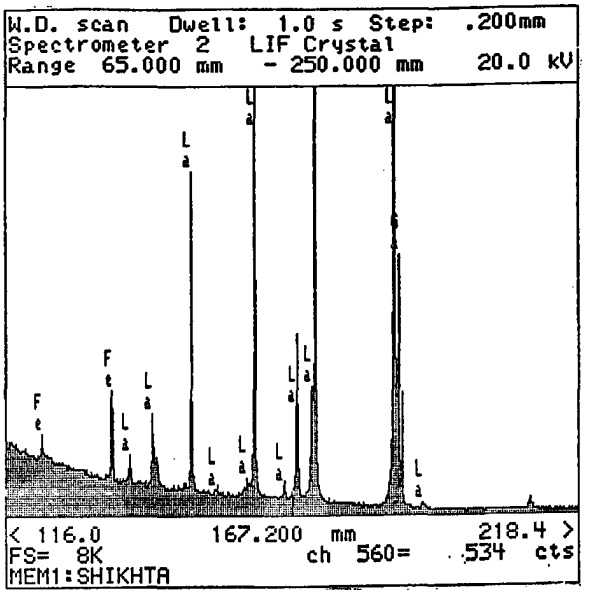
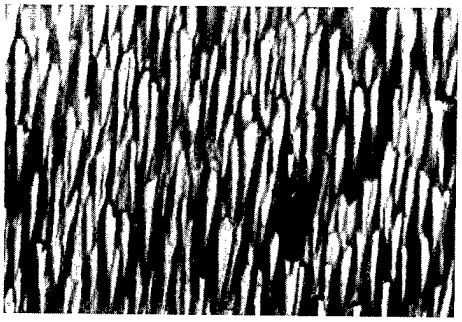


Figure 3. Phase diagram in the system La_2O_3-Ga_2O_3-SiO_2

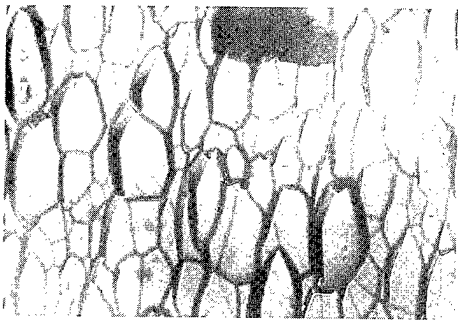




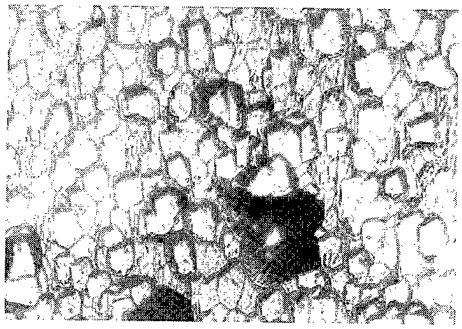
Fugures 4, 5. Test results the charge obtained by self - distributing high temperature synthesis



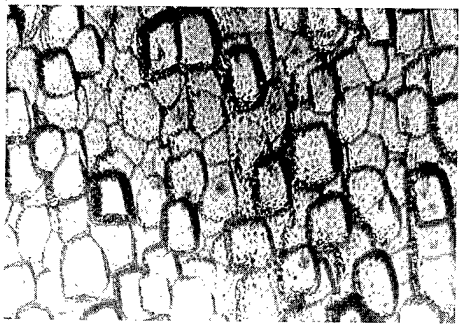
$HNO_3 : H_2O = 1: 2$



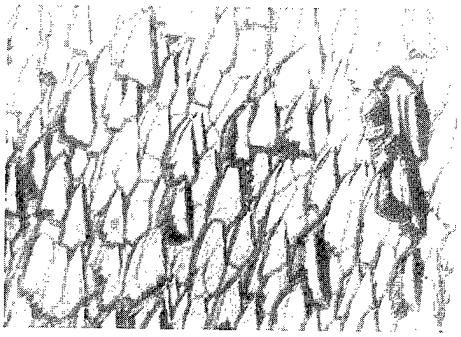
$HCl : H_2O = 1 : 2$



$HCl : HNO_3 : H_2SO_4 = 1 : 1: 1$



$HNO_3 : CH_3COOH = 2 : 3$



$HCl : HNO_3 = 3 : 1$



$HNO_3 : H_2SO_4 : H_2O = 1 : 1 : 1$

Figure 6. Chemical etch figures of langasite plates obtained in different solutions at temperatures near to boiling point (magnification x 300)

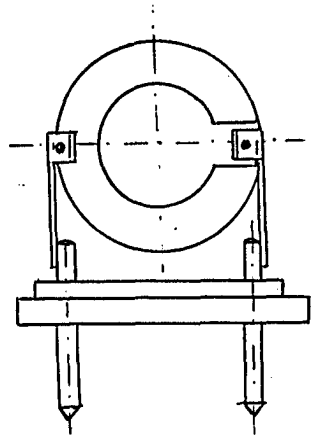


Figure 7. Langasite resonator assembly

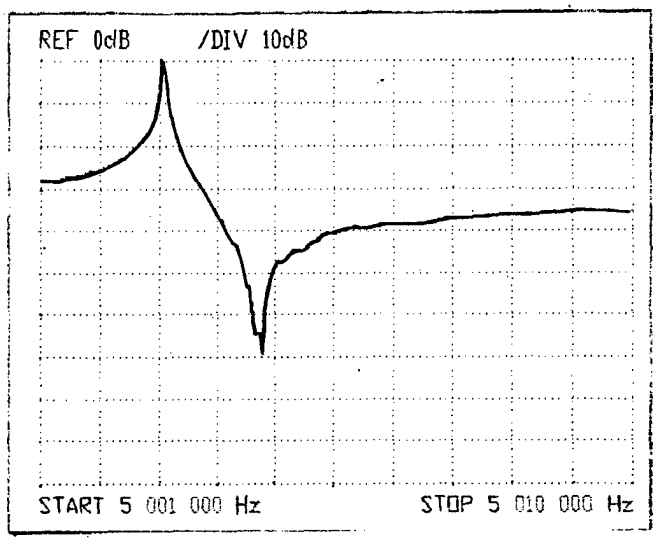


Figure 8. Amplitude-frequency characterictic of a langasite resonator

surface. Mixtures of nitric and acetic acids possess polishing properties.

In Table 4 etch rates and surface quality evaluations are given after etching in several solutions in the vicinity of their boiling point. Silver film electrodes were vacuum evaporaed at a cold substrate.

Langasite piezoelectric elements were mounted into package by two methods: by using conducting glue with temperature stability of 300°C and by sealing. Resonators were hermetically sealed in glass enclosure by high-frequency welding method in vacuum with a residual pressure $\approx 5x10^{-3}$ mm Hg. In accordance with the calculated data and frequency coefficient of langasite for PE diameter of 15 mm and the edge height of 0.2-0.3 mm, optimum radius of sphere is 100 mm. Optimum electrode diameters calculated were taken 5-8 mm (Fig. 7). Due to the fact that no reliable theory exists on calculation of single frequency resonators with suppressed anharmonics, samples with different electrode dimensions have been fabricated. Parameter measurements of these resonators allowed us to find out that resonators with electrode diameter of 5 mm have maximum quality factor Q. The main resonator parameters are represented in Table 5 and Fig. 8.

In spite of the fact, that it has not been managed to obtain desirable crystal quality at this stage of the work, rather high Q-value of langasite resonators is achieved, which evidences of perspectiveness of this material and expediency of its further investigation.

CONCLUSIONS

1. Optimization of heat assembly and pulling rates allowed us to reduce impurity content in crystals by 1.5 times and to reduce c-parameter spread of the crystal lattice.

2. Investigations of charge and crystal composition have showed that langasite has a wide range of homogeneity. This can cause strong discrepancy from a stoichiometric composition, which requires rigid composition tests at all stages of charge preparing.

3. A new crystal phase of triple composition at the crystal surface has been discovered, which evidences of the necessity to make phase diagram more exact.

4. For eliminating tempering stresses and more regular stress distribution along the crystal radius, this additional annealing process at 1100-1300°C is sufficient.

5. Langasite resonator design at 5 MHz, 3rd overtone has been developed. Maximum quality factor Q obtained at this stage of investigation was $1.15x10^5$, but this Q-value is not a limiting one, which evidences of langasite crystals perspectiveness.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Dr. J. Vig for his general assistance and fruitful discussions. This work has been supported by the U.S. Army c/o NAVAL REGIONAL CONTRACTING CENTER under the Contract № 68171-94-C-9004.

BIBLIOGRAPHY

[1] Stolin Yu.I., Tkachev S.P., Determination of orthosilicate gadolinium Gd_2O_3-SiO_2, Kristallography, 1969, v.14, p.p. 22-25.

[2] Kousmin E.A., Belov N.A., Crystal structure of simplest silicates La and Sm, Report of the Academy of Science of U.S.S.R., 1965, № 1, p.p.88-90.

[3] Mizuno M., Yamada T., Ohtoke T., Phase Diagram of the System Ga_2O_3-La_2O_3 at High Temperature. -Ceram. Soc. Jap., v. 93, № 6, p.p. 295-300, 1985.

[4] Brandle C.D., Steinfink H., Crystal Structure $Eu_4Al_2O_9$, Inorg. Chem., 1969, v. 8, № 6, p.p. 1320-1324.

[5] Scheneider S.F., Roth R.S., Waring T.L., Solid State Relations Involving Oxides of Trivalent Cations - F.Res. NBS, Ser. A, 1961, v. 65, p.p. 345-374

[6] Nicolas F., Coufures F.P. Etude du Digram de Phase Nd_2O_3-Ga_2O_3., Rev. Intern. Hautes Temper. Refract. Fr., 1983, v. 20, p.p. 129-133.

[7] Toropov N.A., Lin Tsu-Cjan, Study of a binary system Ga_2O_3-SiO_2, Journal of Nonorganic Chemistry, 1960, v. 5, № 11, p.p. 62-65.

[8] A.N. Gotalskaja, D.I. Dresin , G.N. Cherpoukhina, V.V. Bezdelkin, "Aspects of growing langasite crystal and properties", Journal de physique IV, Colloque C2, supplBment au Journal de Physique III, Volume 4, fBvrier 1994

[9] Smagin A.G., Jaroslavsky M.I., Piezoelectricity of Quartz and Quartz Crystal Units, Publ. House "Energija", 1970, p.p. 487

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

BRIDGMAN GROWTH OF $Li_2B_4O_7$ CRYSTAL OF 4-INCH DIAMETER

Shi-ji Fan, Jue Wang and Yi-bin Xu*

Shanghai Institute of Ceramics, Chinese Academy of Sciences
1295 Dingxi Road,Shanghai 200050, P.R.China

Lithium tetraborate $Li_2B_4O_7$ (LBO) piezoelectric crystal of 4-inch diameter by 3-inch length was grown by the modified Bridgman method. The most important factor in large diameter of LBO growth is growth rate of the crystal. In order to enssure perfection of LBO material in core-free, twin-free, no crack and no striation, the growth rate V must be reduced with increasing of the diameter D of LBO boule. The relationship between V and D obeys the fitting equation of experimental curve: $V=(0.14+0.34D)^{-1}$. The irregularity of V and its influenced factors, such as the fluctuation of the furnace temperature, the shape and location of solid-liquid (S-L) interface in the furnace as well as effects of V on perfection of the crystal have been also reported.

1. Introduction

Lithium tetraborate (LBO) piezoelectric crystal substrate material has attracted more and more attention due to its fairly large coupling coefficient for surface acoustic wave (SAW) and bulk acoustic wave (BAW), zero temperature coefficient of delay (TCD) [1] and high reflectivity of SAW for aluminum strips on LBO substrate [2].The crystal growth technique of LBO [3-6] and the fabrication process of SAW devices of high frequency, moderate bandwidth, low insertion loss, high rejection, high stability of central frequency and small size on LBO[7] have been developed. LBO crystal in 3-inch diameter has been produced by the modified Bridgman technique and more intensive research on the properties of BAW, leaky SAW and Bleustein-Gulyaev wave of LBO crystal[8-11] are investigated recently.

In this paper, the experimental results of 4-inch diameter LBO crystal grown by the modified Bridgman technique are reported for the first time. The purpose of this research is to establish a more substantial and economic LBO material production base.

* Jia-yue Xu, Guang-yu Zhang, Jin-Di Wu, Ren-ying Sun, Ya-fang Lin, Xue-wu Xu, Wen Wang worked in this research.

2. Experimental

The procedure of 4-inch diameter LBO crystal growth is about the same as that of 3-inch diameter LBO . The scheme of LBO crystal growth by the modified Bridgman method and its temperature control is shown in Fig.1.The composition of raw material LBO powder with a purity of 4N or 3N corresponds nearly to stoichiometry in the LBO in which the weight ratio of Li_2O and B_2O_3 is 0.2146. The raw material is pressured to form a round block of near 4-inch diameter with density of 1--1.4 g/cm^3 and placed in Pt cylindric crucible. LBO seed takes usually <110> orientation which is very useful for SAW device and is put at the bottom of the crucible in advance. The assembled crucible is supported in the Bridgman furnace by a refractory tube filled with Al_2O_3 powder. The feed material blocks and the top of the seed are melted in the furnace and then the Pt crucible is lowered down by a mechanism driven by a motor for LBO crystal growth. The temperature of the furnace is controlled by a DWT-702 model controller with an accuracy of $\pm 0.5\,^{\circ}C$, shown in Fig.1. Pt/Pt-10% Rh thermocouples are used for the control and measurement of temperature in both furnace and refractory tube.

3.Results and discussion

3.1. The growth rate V of LBO crystal and the diameter D of LBO boule

Growth rate of crystal from melt is a key factor for perfection of crystal. Generally speaking, based on mass and heat transfer during crystal growth,with increasing of diameter of crystal boule growth rate must be reduced in order to ensure perfection of crystal. The LBO crystals in diameter of 1 to 4 inches have been grown by us to find the relationship between growth rate and diameter and a suitable rate for large diameter. The results are shown in Fig.2. The experimental results indicate obviously that the growth rate V of LBO crystal decreases rapidly with increasing of the diameter D of the boule from 1 to 3 inches and then V decreases slowly with increasing of D

0-7803-2500-1/95 $4.00 © 1995 IEEE

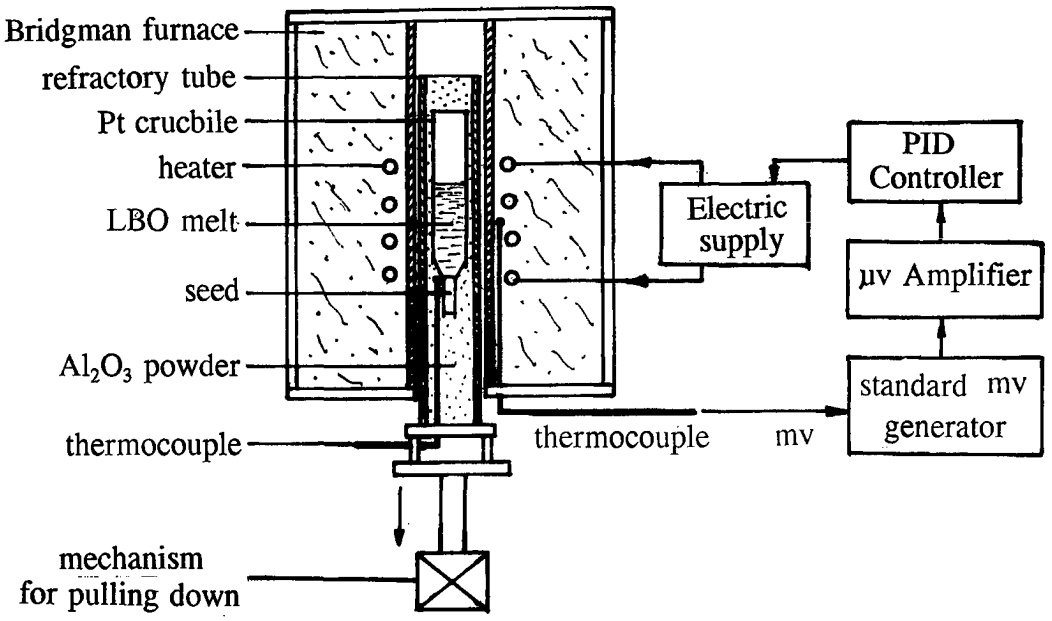


Fig.1 Scheme of Bridgman growth of LBO and temperature control

from 3 to 4 inches, which is much beneficial to production of the material of large diameter. The fitting equation of the relationship between V and D simulated by computer is:

$$V=(0.14+0.34D)^{-1}$$

In Fig.2. the fitting curve of the equation accords well with the experimental data.

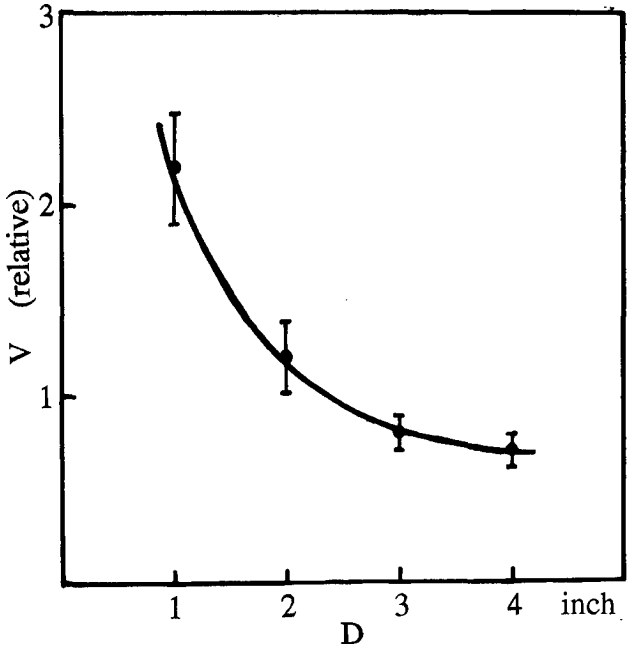


Fig.2 The experimental data and the fitting curve of V and D for LBO growth

3.2. The irregularity of the growth rate V

In the Bridgman growth of LBO, the growth rate V is controlled by the lowering rate of the crucible with a mechanism. The lowering rate of the crucible is constant during whole period of the growth and the V was designed to equal the lowering rate of the crucible. However, the growth rate V is changed generally due to fluctuation of furnace temperature, change of temperature field profile in the crucible during lowering down. When the furnace temperature decreases abruptly , the V must increase irregularly. The fluctuation of the furnace temperature is caused mainly by the instability of electric supply and the temperature in circumstance located the furnace. So stabilizing the electric supply and the circumstance temperature as well as choosing suitable PID parameters of the controller must be made to improve the stability of the furnace temperature.

3.3. The solid-liquid(S-L) interface and the growth rate V

When the top of the LBO seed and the feed material are melted, the S-L interface is formed in the crucible. The shape of the S-L interface and the location of it in the furnace influence the V also. Fig.3. shows the shapes of S-L interface in LBO growth.

When the S-L interface is convex (Fig.3,a) or flat (b), the V accords normally with the lowering rate of the crucible however if the S-L interface is located at the position lower than the normal position at which the S-L

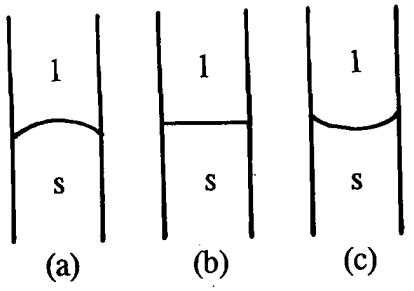


Fig.3 The shapes of S-L interface in LBO growth

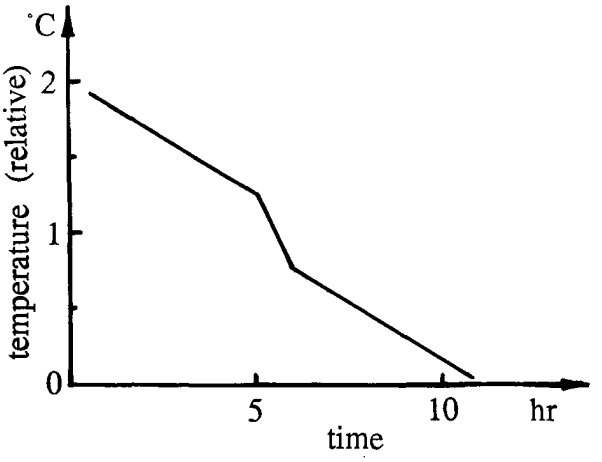


Fig.4 The abrupt decrease of the temperature on the S-L interface

interface is convex or flat in the furnace, the S-L interface becomes concave (Fig.3,c) while the V increases during lowering of the crucible, especially at the centre of the boule i.e. the bottom of the concave interface so that coring occurs at the centre of the boule. With lowering of the crucible, the temperature field in the crucible changes,concave S-L interface can become convex or flat and then coring disappears, V keeps stable.

3.4. The growth rate V and imperfection of LBO crystal

Perfection of LBO crystal of large diameter depends on the V mainly. When V is about 5 mm per day, the dendrite growth and coring of LBO can take place. During the growth, abrupt fluctuation of temperature in the furnace happens sometimes, accordingly, the decreasing rate of temperature on the S-L interface becomes fast, as shown in Fig.4, that causes the V increasing to form the inclusions of constitutional supercooling, a white striation throughout the whole or partial cross-section of LBO boule.

3.5. The growth rate V and cracking of the boule

Cracking of the 4-inch diameter boule depends on completeness of as-grown boule, i.e. the total crystallization of the melt of LBO. If the melt remained on top of the as-grown boule is freezed into polycrystalline or glass phase during finishing of the growth, partial cracks in the top of as-grown boule occur often because of large difference of thermal expansion. On the other hand, the temperature of the melt on the top region decreases much slowly because of the thermal radiation from high temperature region of the crucible so that the V in the top region becomes much slower than the lowering rate of the crucible. In order to obtain total crystallization of the melt to speed up the lowering rate of the crucible or to decrease the temperature of the furnace during finishing of the growth are the appropriate measures.

3.6. Perfect growth of 4-inch diameter of LBO crystals

Based on the above results, choosing the shape and location of S-L interface, we have grown stably the <110> oriented LBO boules (Fig.5) with core-free, twin-free, no crack and no striation in 4-inch diameter by 3-inch length under the temperature of the Bridgman furnace 950-1050 °C and the temperature gradient 10-20 °C/cm near the S-L interface and at the lowering rate of the crucible less than 3 mm per day. The results of 4-inch diameter of LBO growth have established a more substantial and economic LBO piezoelectric material production base.

References

[1] R. W. Whatmore, N.M.Shorrocks,C.O'Hara, F. W. Ainger and I. M. Young, "Lithium Tetraborate: A New Temperature-Compensated SAW Substrate Material", Electronics Letter, vol, 17, 1981, pp.11-12.

[2] Y. Ebata and M. Koshino,"SAW Resonator and Resonator Filter on $Li_2B_4O_7$ substrate",Jpn. J. Appl. Phys. vol.26, Suppl. 26-1, 1987, pp.123-125.

[3] S-J. Fan, G-S. shen, W.Wang, J-L. Li and X-H. Le,"Bridgman Growth of $Li_2B_4O_7$ Crystals", J. Crystal Growth, vol.99, 1990, pp. 811-814.

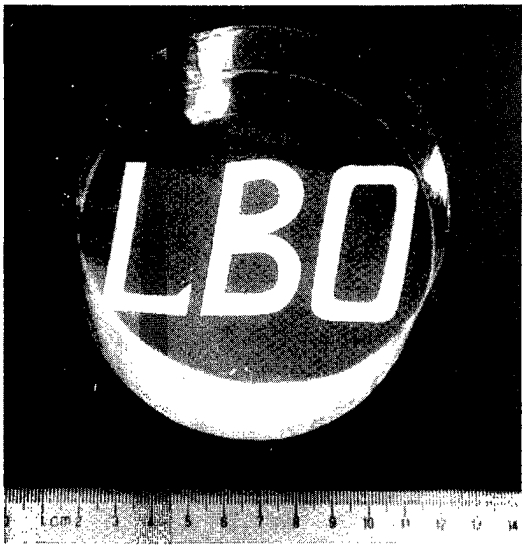


Fig.5 The LBO crysatl boule of 4-inch diameter

[4] S-J.Fan, W. Wang, J-J. Xiang and X-H. Le, "Core-free Growth of Three inch Diameter $Li_2B_4O_7$ Crystal for SAW Devices", Crystal Properties and Preparation, vols 36-38, 1991, pp. 46-51.

[5] S-J. Fan, W. Wang, J-J. Xiang and J-Y. Xu, "The Development of Industrial Growth Technology of New Piezoelectric Crystal $Li_2B_4O_7$ Material", Proc. of International Conference on Electronic Components and Materials, 1992, pp. 206-209.

[6] S-J. Fan, " Properties,Production and Application of New Piezoelectric Crystal Lithium Tetraborate $Li_2B_4O_7$", Proc. 1993 IEEE Freq. Contr. Symp. 1993, pp. 353-358.

[7] H.Abe, M.Ohmura and H. Saitou, "SAW Devices on Lithium Tetraborate ($Li_2B_4O_7$)", Proc. 1994 IEEE Freq. Contr. Symp., 1994, pp.289-295.

[8] J.Kosinski ,A. Ballato, and Yicheng Lu "Measured Properties of Doubly rotated Dilithium Tetraborate ($Li_2B_4O_7$) Resonator and Transducers", IEEE Trans. Ultranson., Ferroelec., Freq. Contr., vol. 40, No.1, 1993, pp.154-161.

[9] J.A. Kosinski, A, Ballato and Yicheng Lu ,"Pure-mode Measurements of Dilithium Tetraborate Material Properties" Proc. 1993 IEEE Freq. Contr. Symp. 1993, pp.359-370

[10] M. Adachi, T. Shiosaki, K. Ohtsuka and A. Kawabata, "Leaky SAW Propagation Properties on $Li_2B_4O_7$ Substrates", Proc. 1994 IEEE Freq. Contr.Symp., 1994,pp. 296-300.

[11] W. Soluch, "Properties of Bleustein-Gulyaev Wave in $Li_2B_4O_7$ Crystal", Proc. IEEE Ultrason. Symp. 1993 pp.237-241.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THE ANOMALOUS ELASTIC ANISOTROPY OF $Li_2B_4O_7$ AND ITS INFLUENCE ON SAW PROPERTIES

Natalya F.Naumenko

Crystallography Dept., Moscow Steel & Alloys Institute,
117936, Leninski prosp.,4. Moscow, Russia

Abstract

The anomalous elastic anisotropy of $Li_2B_4O_7$ caused by the "incorrect" relation between two elastic modula, c_{33} and c_{44}, is proved to be responsible for the existence of earlier found undamped longitudinal horizontally polarized (LH) type leaky waves. A simple proof is presented to show that "ray-polarized" quasilongitudinal bulk wave propagating in the plane of reflectional symmetry satisfies the stress-free conditions when the boundary plane is parallel to polarization vector and orthogonal to the plane of reflectional symmetry. The close relation between the particular directions, such as acoustic axes, and the existence of "non-leaky" waves is analyzed using the "exceptional wave line" method. LH type leaky waves exist in crystal cuts with Euler angles (φ,θ,90), where φ is arbitrary when θ=40...46°. If φ=45° the permitted interval of θ angle is the largest: θ=35-68°.

1.Introduction

The acoustic anisotropy of any crystal which means the anisotropy of bulk acoustic wave (BAW) velocities in infinite medium plays a very important role in the investigation of surface acoustic waves (SAW). The geometrical image of acoustic anisotropy is the refraction (or slowness) surface. While analyzing this characteristic surface one can predict the behavior of SAW in selected crystal cuts, for example, the splitting of general SAW solution into two separate waves with different velocities and mechanical displacement (polarization) vectors or the existence of SAW with "supersonic" velocity in relation to the slowest bulk mode.

The analysis of acoustic anisotropy should include the search for particular directions [1]: longitudinal and transverse normals and acoustic axes. While propagating along two first particular directions one of three bulk modes is pure longitudinal or pure transverse respectively. If propagation direction is parallel to one of acoustic axes, at least two bulk modes have the same velocity and thus degenerate. The last type of particular directions in crystals of middle symmetry systems - trigonal, tetragonal and hexagonal - is usually localized in the planes of reflectional symmetry or the planes orthogonal to even-fold symmetry axes. The bulk waves, propagating along 3-,4- and 6-fold symmetry axes also degenerate. However in some crystals the acoustic axes of "general position" can exist. The example is quartz, a crystal with very strong acoustic anisotropy.

An infinite number of bulk modes with different polarization vectors can propagate parallel to the acoustic axis and if the last one is of conical type [2], the Poynting vector strikes a cone while propagation direction rotates around the axis. It is obvious that all acoustic properties change rapidly in the vicinity of this particular direction. The refraction surface is characterized by the negative curvature in this special area. Thus if the acoustic axis occurs to be in the sagittal plane for the specified SAW orientation it leads to the appearance of additional "limiting bulk waves"[3], propagating parallel to the surface. The corresponding surface skimming bulk waves (SSBW) can be observed experimentally.

Usually only slowest and middle bulk modes degenerate but some exclusions are known, so called "longitudinal acoustic axes". For example, in paratellurite TeO_2 (symmetry 422) in the plane (100) the velocities of the fast and the middle bulk waves coincide. As a result there is a strong deviation of power flow vector from the propagation direction for quasilongitudinal wave. In (001) plane the angle between these two vectors exceeds 73°. The other important result is the existence of crystallographic orientations in TeO_2,

0-7803-2500-1/95 $4.00 © 1995 IEEE

where the quasilongitudinal bulk wave satisfies the stress-free mechanical boundary conditions. This phenomenon was discovered first in [4]. The other example of crystal with anomalous elastic anisotropy is $Li_2B_4O_7$.

2. Elastic anisotropy of $Li_2B_4O_7$ and ray-polarized waves

The crystal of lithium tetraborate $Li_2B_4O_7$ (symmetry 4mm) is one of the most promising new piezoelectric materials for SAW devices due to the high piezoelectric coupling and the existence of temperature compensated cuts for SAW [5-6]. The crystal is already used in SAW devices [7]. Though currently there are many publications devoted to theoretical and experimental investigation of $Li_2B_4O_7$ one important subject, to the author's best knowledge, was not discussed. It is the anomalous elastic anisotropy of this material and its influence on SAW and leaky wave properties.

Despite of some difference in the values of material constants of lithium tetraborate obtained by different researchers [8] the elastic stiffness module c_{33} is usually smaller than c_{44}. It is known that if piezoelectric coupling is ignored the longitudinal BAW propagates along Z axis of tetragonal crystal with velocity $V_l = \sqrt{c_{33}/\rho}$ and all degenerated shear waves polarized in XY-plane have velocity values $V_s = \sqrt{c_{44}/\rho}$, where ρ is a mass density. Usually c_{33} is larger than c_{44} and the longitudinal wave is faster than shear ones. The anomalous relation between these constants leads to the degeneration of the bulk acoustic mode with the highest velocity.

Fig.1 shows the cross section of the refraction surface by the plane (100) calculated without (a) and with (b) piezoelectric effect. Because of symmetry one bulk mode, pure shear, is polarized orthogonal to (100) plane. Two other bulk modes - quasishear and quasilongitudinal - have polarization vectors within this plane. There is an acoustic axis at an angle 69° to Z-axis. It is of ordinary type: the slowest and the middle bulk modes degenerate. One more acoustic axis is parallel to [001] direction (4-fold symmetry axis). If piezoelectric effect is ignored (fig.1a) then two internal (fast and middle) branches degenerate. Since both degenerated modes have polarization vectors in the plane (001), they should be pure shear, while one of them becomes pure longitudinal when propagation direction is parallel to [010]. The non-degenerated mode is pure longitudinal for [001] and pure shear for [010] direction. It is obvious that these bulk modes must have polarization angles varying in the interval 0...90°.

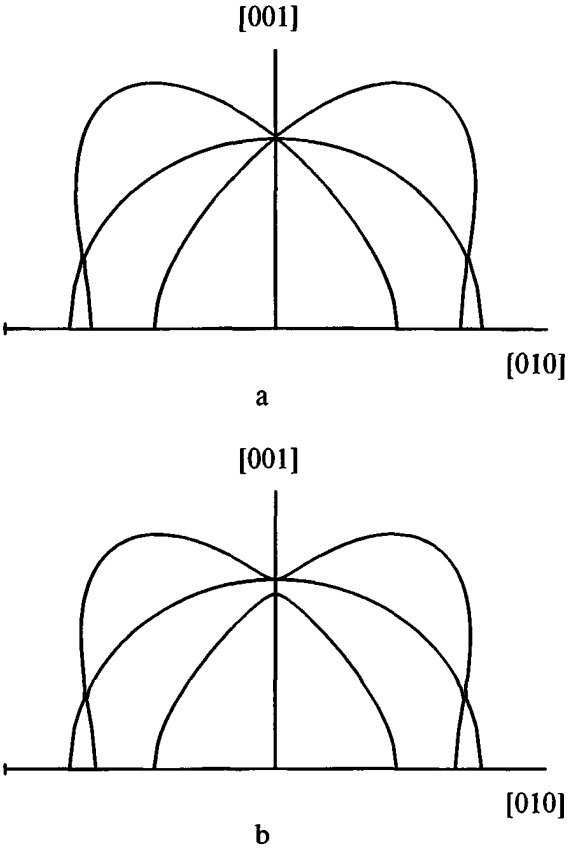


Figure 1. Cross-section of the refraction surface for $Li_2B_4O_7$ without (a) and with (b) piezoelectric effect.

If piezoelectric effect is taken into account the situation changes (fig.1b). The anomalous acoustic axis is transformed to the ordinary one. However the strong anisotropy is observed for the quasilongitudinal wave and the power flow angle exceeds 32°. The material constants of $Li_2B_4O_7$ from [6] were used for calculations.

Such behavior is typical for any crystallographic plane parallel to [001] direction. Fig.2 shows the polarization angle β and the power flow angle ϕ for the bulk mode with the highest velocity in (110) plane of $Li_2B_4O_7$ as functions of propagation angle φ to [110] direction. Piezoelectric effect is ignored. The angle ϕ reaches the value 36.7° when the

propagation direction is close to Z axis ($\varphi \approx 87°$) being zero for $\varphi=0$, 46 and 90°. Zero values of ϕ correspond to longitudinal normals (one of bulk modes has $\beta=0$).

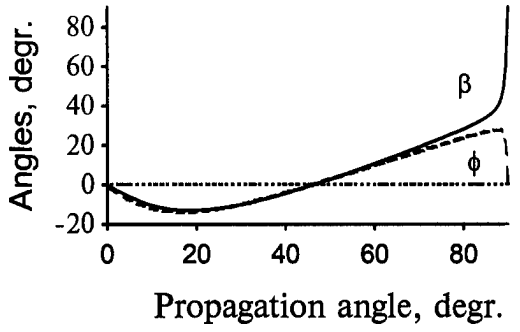


Figure 2. Polarization angle β and power flow angle ϕ of quasilongitudinal wave as functions of propagation angle related to [110] direction in the plane (110).

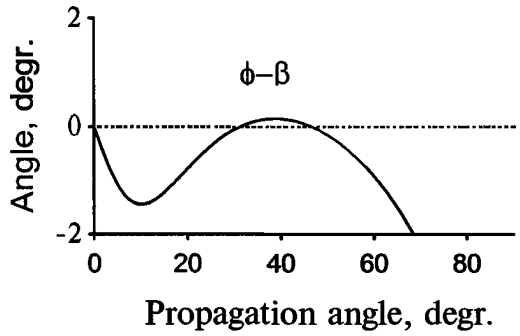


Figure 3. The difference (ϕ–β) as function of propagation direction in (110) plane for quasilongitudinal bulk wave.

In fig.3 the difference between two angles (ϕ–β) is shown for the same wave. It is important that in (110) plane the directions can be found where these two angles are equal and the power flow vector is parallel to polarization.

If piezoelectric effect is ignored, the wave equation for bulk acoustic waves is [1]

$$c_{ijkl}\, n_j\, n_k\, u_l = \rho V^2 u_i \qquad (1)$$

where $\mathbf{n}$ is a unit wave normal, $\mathbf{u}$ is a polarization vector, ρ is a mass density and c_{ijkl} is an elastic stiffness tensor. Equation (1) multiplied by the vector $\mathbf{m}$ which is orthogonal to $\mathbf{u}$ and lies within the plane (110) yields

$$c_{ijkl}\, n_j\, n_k\, u_l\, m_i = 0 \qquad (2)$$

For quasilongitudinal waves polarized in (110) plane with polarization angle β one can write

$$\mathbf{n} = \cos\beta\, \mathbf{u} + \sin\beta\, \mathbf{m} \qquad (3)$$

After substitution of (3) the equation (2) changes to

$$\cos\beta\, c_{ijkl} n_k u_l u_i m_j + \sin\beta\, c_{ijkl} n_k u_l m_i m_j = 0 \qquad (4)$$

The first term in (4) with account of expression for group velocity [1]

$$V^g_i \sim c_{ijkl}\, u_j\, u_k\, n_l \qquad (5)$$

and the orthogonality relation for $\mathbf{V^g}$ and $\mathbf{m}$ gives

$$\mathbf{V^g}\, \mathbf{m} = 0 \qquad (6)$$

Then the equation (4) is satisfied if $\sin\beta=0$ or

$$c_{ijkl} n_k u_l m_i m_j = 0, \qquad (7)$$

which can be written as

$$\mathbf{T}\, \mathbf{m} = 0, \qquad (8)$$

where

$$T_i = c_{ijkl} n_k u_l m_j \qquad (9)$$

are the normal stresses caused by the propagation of the bulk wave with wave vector $\mathbf{n}$ in the half-infinite medium with the boundary surface normal to $\mathbf{m}$. Since the analyzed quasilongitudinal wave is polarized in the plane of reflectional symmetry (110), the vector $\mathbf{T}$ is localized in the same plane. It can be easily shown that this vector is orthogonal to $\mathbf{u}$, since the equation (9), multiplied by $\mathbf{u}$ with account of (5) yields

$$\mathbf{T}\, \mathbf{u} = \mathbf{V^g}\, \mathbf{m} = 0. \qquad (10)$$

Then vector $\mathbf{T}$ must be parallel to $\mathbf{m}$ and from (8) it follows that $\mathbf{T}=0$.

It means that the quasilongitudinal (but not pure longitudinal) wave, propagating in the plane of reflectional symmetry and polarized in the same plane does not disturb the free surface with normal **m** lying within the same crystal plane, if its power flow (group velocity) vector is parallel to polarization direction. This theorem was proved for general case in [9] and such waves were called "ray-polarized".

It is important to mention that, though the wave can be quasishear as well as quasilongitudinal, the power flow angle is usually smaller than 45° and thus the "ray-polarized" waves should be quasilongitudinal rather than quasishear.

Since both the polarization and the power flow vectors are normal to **m** and localized in the boundary plane, the corresponding SAW is pure longitudinal and horizontally polarized (LH) wave.

In addition to Z axis there are two points in fig.3, where the difference function vanishes: φ=31 and 46° with Euler angles (45, 41, 90) and (45, 46, 90). While the second point corresponds to the longitudinal normal (β=0), the first one allows the propagation of LH wave if the crystal is non-piezoelectric.

The piezoelectric effect while taken into account usually changes the structure of LH wave, making it quasi-LH, with small contribution of other partial modes. It can be transformed to leaky wave or disappear as non-physical solution, which was theoretically shown for SH waves in transversely isotropic media [10]. However nearly undamped SSBW with quasilongitudinal polarization usually can be found in these specified crystal cuts.

3. EXCEPTIONAL WAVE LINE ANALYSIS

The collinearity of power flow and polarization vectors is a sufficient but not necessary condition for bulk wave to satisfy the mechanical boundary conditions. Similar bulk waves, called "exceptional waves"(EW)[3] can be found in non-symmetric crystallographic orientations. The general criterion was established in [2]: the bulk wave is exceptional if:

$$\det(\mu_{ij})=0, \qquad (11)$$

where

$$\mu_{ij}= c_{ijkl}\, n_k\, u_l \qquad (12)$$

Such wave satisfies stress-free boundary conditions only on the selected boundary plane with normal **m** from

$$\mu_{ij} m_j=0 \qquad (13)$$

EW is always horizontally polarized [2]. Thus to answer the question whether a bulk wave propagating in an arbitrary crystallographic direction **n** is exceptional or not one must solve the wave equation (1) to find polarization vectors **u** of three bulk modes and substitute them into equations (11)-(13). If (11) is satisfied for any of bulk waves this one is exceptional. While vector **n** scans the three-dimensional space in crystallographic basis, the one-dimensional space of permitted directions for EW can be found. These directions form the continuous "exceptional wave lines"(EWL) on the stereographic projection of unit wave normal sphere $\mathbf{n}^2=1$ [2]. Their configuration depends on the crystal symmetry and the position of acoustic axes. The analysis of this dependence as well as some numerical results for crystals of different symmetry can be found in [11]. Only three of six analyzed materials - paratellurite TeO_2, lithium tetraborate and quartz - allow the propagation of quasilongitudinal EW, while quasishear one exist in all crystals. In TeO_2 the occurrence of EWL for quasilongitudinal mode is caused by the presence of "longitudinal acoustic axis", since according to general theory [2,11] EWL must go from one degenerated branch to another in this particular point.

Fig.4 shows the stereographic projection of unit wave normal sphere with EWL for $Li_2B_4O_7$ without (a) and with (b) piezoelectric effect. Only the part of sphere is plotted with account of crystal symmetry. The acoustic axes are marked by bold points. For the trivial pure SH waves both unit wave vectors **n** and plane normals **m** belong to the planes of reflectional symmetry. These waves are polarized normally to symmetry planes and in the points of conical acoustic axes EWL goes from one degenerated branch to another.

In addition to these "trivial" lines the "non-trivial" EWL have been found for all three bulk modes. Their configuration depends on piezoelectric effect. The permitted unit wave vectors for quasi-longitudinal EW in non-piezoelectric crystal form a cone (fig.4a) with the rotation axis parallel to Z direction. The last direction being the acoustic axis of "tangent" type allows the propagation of infinite number of pure

shear EW [12], one of these waves belongs to the branch with the highest velocity. Thus it may be considered as one more cone, degenerated. While piezoelectric effect becomes appreciable both cones are coupled and for the real values of piezoelectric constants in $Li_2B_4O_7$ there are four small cones with the axes in (110) planes (fig.4b). If piezoelectric coupling grows, these four cones disappear. It means that the existence of LH type waves is the result of the remaining elastic anomaly.

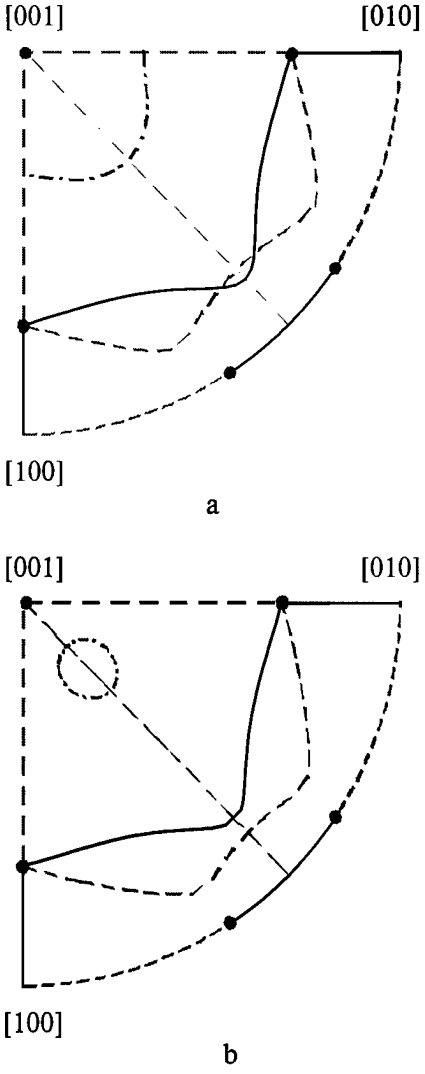


Figure 4. Stereographic projection of exceptional wave lines in lithium tetraborate without (a) and with (b) piezoelectric effect: quasilongitudinal waves (dashed-dotted line), fast shear (dashed line) and slow shear (solid line) waves.

The intersections of quasilongitudinal EWL with the plane (110) in fig.4b give two cuts with Euler angles (45, 41, 90) and (45, 46, 90).

To analyze the structure of modified EW with account of electric boundary conditions it is necessary to solve the boundary problem by means of some numerical techniques.

4. LH type leaky waves in $Li_2B_4O_7$

The rigorous solution of boundary problem with both mechanical and electric boundary conditions reveals that in lithium tetraborate longitudinal horizontally polarized (LH) type leaky waves with decay coefficients smaller than 0.001 dB/wavelength can propagate in orientations with Euler angles $(\varphi,\theta,90)$, where φ is arbitrary when $\theta=40...46°$.

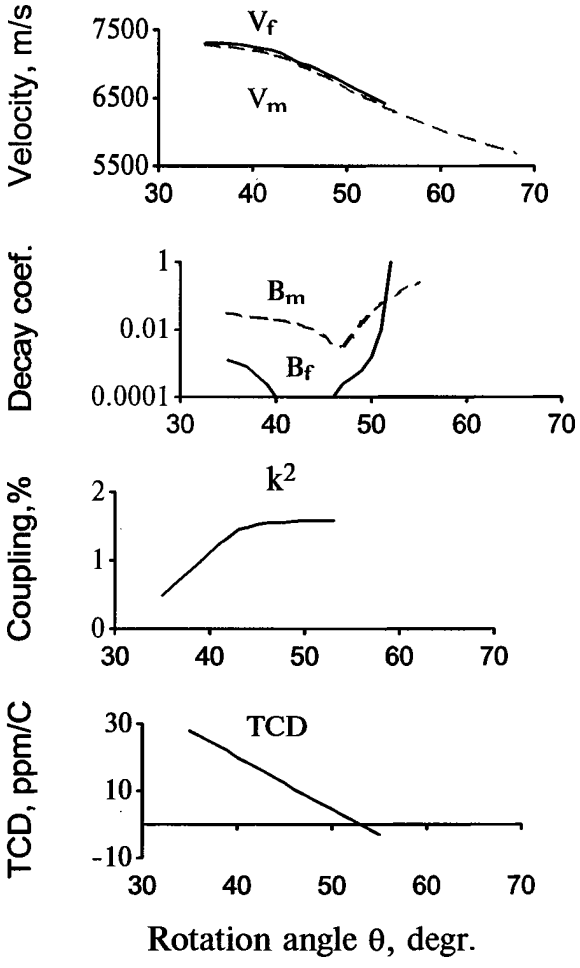


Figure 5. LH type leaky wave characteristics in $(45,\theta,90)$ cuts of $Li_2B_4O_7$

However nearly undamped quasi-longitudinal SSBW can be found outside this area. When $\varphi=45°$ the permitted interval of angle θ for leaky waves is the largest: $\theta=35...54°$ for free surface and $35...68°$ for metallized one.

Leaky waves propagating in orientations with Euler angles $(45,\theta,90)$ were analyzed earlier in [13-15]. The velocities are slightly smaller than cut-off velocities of quasi-longitudinal SSBW for the same sagittal planes. The leaky waves are three-partial and include only one homogeneous mode, shear vertical one. SH wave is uncoupled. The predominate quasilongitudinal partial wave is tilted in relation to the boundary surface, the tilt angle is large and reaches 25°.

Fig.5 shows the calculated values of leaky wave velocities **V**, decay coefficients **B**, electromechanical coupling coefficients k^2 and temperature coefficients of delay (TCD) for crystal cuts with Euler angles $(45,\theta,90)$. The temperature coefficients from [5] were used. Coupling coefficients k^2 were calculated through the velocity difference between free and shorted surfaces. Maximum coupling for LH type leaky waves $k^2=2.1\%$, the velocity **V** varies within the interval 6424-7308 m/s for free surface and there is a cut, where TCD=0.

It is important to mention here that leaky wave becomes "non-leaky" with zero decay coefficient for two cuts: (45, 41, 90) and (45, 46, 90), where exceptional waves exist. In the interval between these two cuts the wave degenerates into SSBW if the surface is unmetallized. For metallized surface the minimum value of decay coefficient $B=0.002$ dB/λ is observed for $\theta=46.5°$. The orientation with minimal decay for application in SAW devices, (45, 46, 90), has the following parameters: $V_f=6980.2$, $V_m=6926.1$ m/s, $k^2=1.6\%$; $B_f=0$, $B_m=0.003$ dB/λ, TCD=10 ppm/°C. For SAW in the same cut $V_f=3095.3$, $V_m=3081.0$ m/s, $k^2=0.92\%$.

The effective dielectric permittivity function [16] versus slowness is shown for this cut in fig.6. One can see that only two waves can be radiated: LH leaky wave and SAW. Since both real and imaginary parts of function vanish in the same cut-off point for LH-wave, it means that this wave has zero attenuation on the free surface.

The longitudinal character of leaky wave is proved by the depth dependence of polarization components in (45, 46, 90) cut, shown in

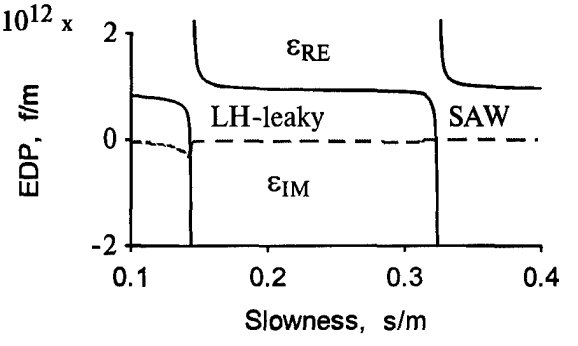


Figure 6. Effective dielectric permittivity function for (45, 46, 90) cut of lithium tetraborate.

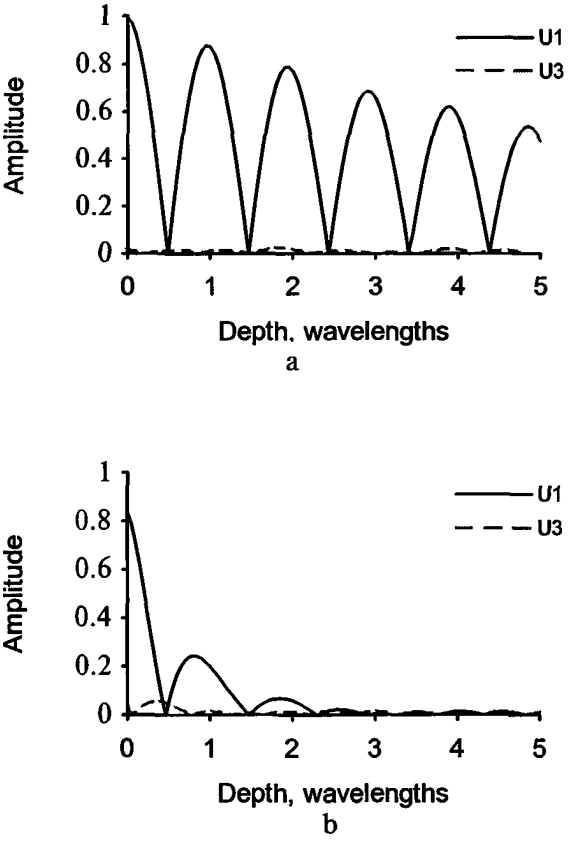


Figure 7. Variation versus depth of amplitude components for LH type leaky wave propagating in (45,46,90) cut: a - free surface, b - metallized surface.

fig.7. There is practically only one longitudinal component, when the surface is free. The vertical

shear component is very small, and horizontal shear one is zero because of symmetry. The metallization (fig.7b) causes an energy concentration within 4-5 wavelengths from the surface.

5.Conclusion

The anomalous elastic anisotropy is shown to be responsible for the existence of LH type leaky waves with small and zero decay coefficients. The exceptional wave line analysis can be successfully used to find orientations in any crystal, allowing the propagation of bulk acoustic waves along the free surface. In contrast to numerical techniques widely used for SAW analysis in the specified crystallographic orientation this method reveals the decisive role of crystal symmetry in the existence of SAW solutions with "transonic" velocities and the close relation between particular directions, first of all acoustic axes, and the permitted orientations for leaky and non-leaky waves.

However it should be mentioned that such anomaly is not necessary. For example, LH type leaky waves have been recently found existing in quartz.

6.References

[1] F.I.Fedorov, The theory of elastic waves in crystals, New York: Plenum Press, 1968.

[2] V.I.Alshits, J.Lothe "Elastic waves in triclinic crystals. III. The problem of existence of exceptional surface waves and some of their general properties," Sov.Phys.- Crystallography, vol.24, pp. 644-648, 1979.

[3] P.Chadwick, G.D.Smith, "Foundations of the theory of surface waves in anisotropic elastic materials", in Advances in Applied Mechanics, vol.17, New York: Academic Press, p.303,1977.

[4] N.F.Naumenko,V.S.Bondarenko,N.V.Perelomova, "Features of Volume and Surface Acoustic Wave Propagation in Paratellurite Crystals.", Sov.Phys.-Solid.State, vol.25, pp.1512-1513, 1983.

[5] N.M.Shorrocks,R.W.Whatmore,F.M.Ainger, I.M.Young "Lithium tetraborate - a new temperature compensated piezoelectric material for surface acoustic wave devices", in Proc.IEEE Ultrasonics Symp., 1981, pp.337-340.

[6] M.Adachi, T.Shiozaki, H.Kobayashi, O.Ohnishi, A.Kawabata "Temperature compensated piezoelectric lithium tetraborate crystal for high frequency surface acoustic wave and bulk wave device applications, " in Proc.IEEE Ultrasonics Symp., 1985, pp.228-232.

[7] H.Abe, M.Ohmara, H.Saitou "SAW devices on lithium tetraborate ($Li_2B_4O_7$)", in Proc.1994 IEEE International Frequency Control Symp., 1994, pp.289-295.

[8] J.A.Kosinski, A.Ballato, Lu Yicheng "Pure-mode measurements of dilithium tetraborate material properties," in Proc. 1993 IEEE International Frequency Control Symp., 1993, pp.359-370.

[9] P.Chadwick "A general analysis of transonic states in an anisotropic elastic body", Proc. Roy. Soc. Lond., vol. A401, pp.203-223, 1985.

[10] V.I.Alshits, V.N.Lyubimov, "Two Types of Exceptional Body Waves in Piezoelectrics and the Sectors of Existence of Bleustein-Gulyaev Waves", Sov.Phys.-Crystallography, vol.30, pp.252-256, 1985.

[11] V.I.Alshits, V.N.Lyubimov, N.F.Naumenko, N.V.Perelomova,A.L.Shuvalov, "Exceptional Elastic Body Waves in Crystals of Various Symmetries", Sov.Phys.- Crystallography, vol.30, pp. 123-126, 1985.

[12] V.I.Alshits, V.N.Lyubimov, A.L.Shuvalov "Exceptional Body Waves in the Neighborhoods of Acoustic Axes of Various types", Sov.Phys.- Crystallography, vol.32, pp.828-835, 1987.

[13] N.F.Naumenko "Quasihorizontal Polarized Quasi-Bulk Acoustic Surface Waves in Piezoelectric Crystals", Sov.Phys.-Crystallo

graphy, vol.37(2), pp.220-223, 1992.

[14] N.F.Naumenko "Leaky surface acoustic waves with quasi-longitudinal polarization in the crystal of lithium tetraborate", Sov.Phys.-Crystallography, vol.32, pp.520-522, 1992.

[15] N.F.Naumenko "Longitudinal horizontally polarized leaky and non-leaky SAW in lithium tetraborate", Phys.Lett., vol.A195, pp.258-262, 1994.

[16] R.F.Milsom, N.C.H.Reilly, M.Redwood "Analysis of generation and detection of surface and bulk acoustic waves by interdigital transducers", IEEE Trans.Sonics Ultrason., vol.SU-24, pp.147-166, 1977.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SYNCHROTRON RADIATION TOPOGRAPHY OF THICK SWEPT QUARTZ

C. K. Suzuki, X. W. Zhang*, M. Ando*, Y. Yoda**, S. Kikuta**,
K. Hamaguchi***, K. Nagai***, and S. Taki***

UNICAMP, University of Campinas, Faculty of Engineering, C.P. 6122,
13081-970 - Campinas, SP, Brazil
*Photon Factory, National Laboratory for High energy Physics, Tsukuba,
Ibaraki 305, Japan
**Department of Applied Physics, Faculty of Engineering, University of Tokyo,
Bunkyo-ku, Tokyo 113, Japan
***Toyo Communications Equipment Co., Ltd., 2-1-1 Koyato, Samukawa,
Kanagawa, Japan

ABSTRACT

Ag- and Au-diffusion swept synthetic quartz samples as thick as 10 mm along the [20$\bar{2}$0] direction were characterized by synchrotron radiation topography at a superconducting vertical wiggler beamline of Photon Factory. The high resolution topographic image allows the visualization of the Ag (Au) diffusion paths along the Z-direction. The stereographic pair topographs can ben used to reproduce the 3-dimensional configuration of the dislocation network. Based on the diffracted and direct (transmitted) images of swept and unswept quartz some new hypothesis concerning the etch-channel suppression mechanism are presented.

INTRODUCTION

X-ray topography with conventional sealed tube and synchrotron radiation have been used for characterization of sweeping effect on the crystalline perfection of quartz crystals.[1-4] The sweeping or electrodiffusion in quartz consists on a selective exchange of impurity ions by applying an electric field along the c-axis of the crystal maintained at ~500°C.[5] The electrodiffusion performed in vaccum produces etch-channel free materials[6], differently from the electrodiffusion done in air, where the airborne impurities can replace some of the originally existing impurities[7], nullifying the resistance effect of etch-channel formation. The technological interest of sweeping has been the radiation hardening of quartz resonator in terms of frequency stability for applications in military and space communications. More recently, the increasing interest of etching techniques for microlithography and thinning techniques for high fundamental frequency resonators and filters[8] has aroused a renewed interest on sweeping processes.

Sweeping performed in air with Cr and Au evaporated metallization electrodes has the same effect of Ag electrodes, in order to reduce or to avoid the etch-channel formation.[9,10] It is supposed that Au and Ag diffusion can inhibit the airborne impurities to enter into the crystal. Even though many contributions have been reported on this subject, the mechanism by which the etch-channels are reduced is not fully understood.

In the present research, Ag and Au swept quartz samples as thick as 10 mm in the [20$\bar{2}$0] direction were characterized by synchrotron radiation X-ray topography with the objective of supplying new informations on the sweeping mechanism of Ag- and Au-diffusion in synthetic quartz.

SPECIMEN PREPARATION AND SWEEPING

Pure-Z synthetic quartz samples with dimensions 20, 10 and 15 mm along X [2$\bar{1}$10], Y[20$\bar{2}$0] and Z [0001] directions were prepared. Metallization of Z-surfaces with Ag and Au was according to the following conditions,

Ag: 1200 Å
Au: Cr 150 Å + Au 1200 Å.

The sweeping was performed in air, whose experimental apparatus and operation conditions had been described before.[10] Table 1 lists the typical conditions of Ag- and Au-diffusion and the remanent Ag and Au concentrations after the sweeping. It is interesting to observe that the Ag-content is of the order of ppm, and Au-content is much less, in the range of ppb's. The synthetic quartz samples becomes etch-channel free using these sweeping conditions in the case of 24 h diffusion.

Table 1
Sweeping Conditons and Ag and Au Concentrations

Specim.	Diffusion	Temp. (°C)	E (V/m)	Time (h)	Concentration
1	none	520	none		Ag, Au (N.D.)
2	Ag	520	250	2	Ag: 10.8 ppm
3	Ag	520	250	24	Ag: 4.8 ppm
4	Au	520	1000	24	Au: 10 ppb

X-RAY TOPOGRAPHY

PRINCIPLE OF IMAGE FORMATION

Suppose the (hkl) crystalline planes of the sample are diffracted by a parallel and monochromatic X-ray beam. In a condition satisfying the Bragg law, there will be a diffracted and a transmitted or direct beam (Fig. 1). If the crystal is perfect, there will be no contrast, but in the case of a defect, for example, a dislocation line here represented by the misorientation cylinders around an edge dislocation, there will be a contrast in the diffracted beam due to the localized strain. Therefore, one part of the transmitted energy will flow to the

0-7803-2500-1/95 $4.00 © 1995 IEEE

diffracted direction, which contributes with an intensity enhancement. In this case, the direct image will present a negative intensity of the defect contrast in comparison with other parts of the crystal.

The radiation intensity (I) going through a crystal is exponentially decreased relative to the incident intensity (I_o), according to

$$I = I_o \exp(-\mu t),$$

where μ is the linear absorption coefficient, which depends on the material and the wavelength of the radiation, and t, is the thickness of the sample. For $\mu t \sim 1$, the dislocation image can be formed by kinematical diffraction,[11] and for $\mu t = 3 \sim 5$, the topographic image will be formed by dynamical and intermediary diffraction effects,[11] which is much more sensitive than kinematical diffraction images. Figure 2 is the $(2\bar{1}\bar{1}0)$ double crystal topograph of a Y-cut special bar with $(11\bar{2}2)$ seed with $\mu t \sim 3$. It was recorded using a conventional X-ray Cu tube, and so, the exposure time is 12 hours even for the sample thickness of ~300 μm. Therefore, the sweeping characterization by synchrotron radiation is fully justified, especially taking into consideration the technological necessity of thick crystals characterization, besides the interesting features of diffraction effects.

SYNCHROTRON RADIATION AND EXPERIMENTAL SETTING

The experiment was conducted at the superconducting vertical wiggler beamline of Photon Factory, National Laboratory for High Energy Physics operating with energy E=2.5 GeV and current i_{max}=340 mA. Figure 3 shows the BL-14 wiggler brilliance as a function of photons energy.[12] For the 20 keV photons, the brilliance is ~20 times more intense in comparison with the bending magnet, and of the order of 1 million times more intense than the conventional sealed X-ray tube. The experimental setting of the X-ray optics is represented in Fig. 4. The Si (111) pre-monochromators tune the radiation to the desired energy, and the asymmetric collimator Si (333) (or Si (800)) makes the beam parallel (plane wave) and with an enlarged cross-section. Quartz $(2\bar{1}\bar{1}0)$

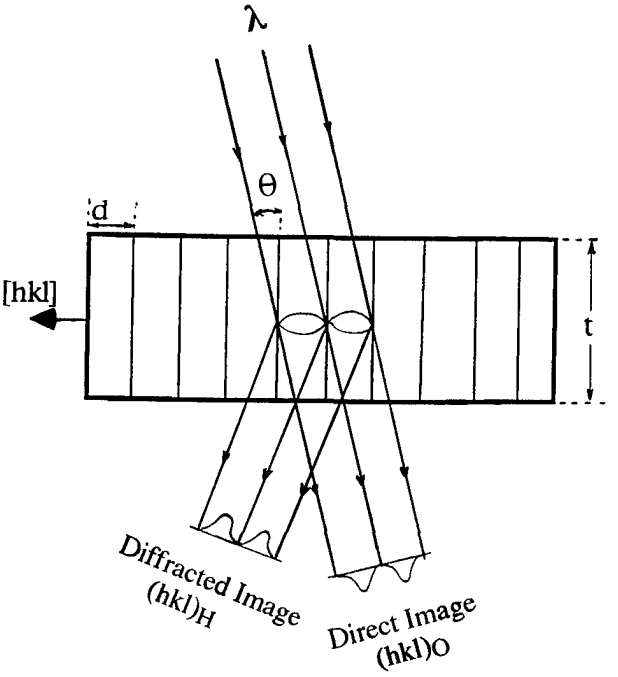


Fig. 1. Image formation in X-ray diffraction topography.

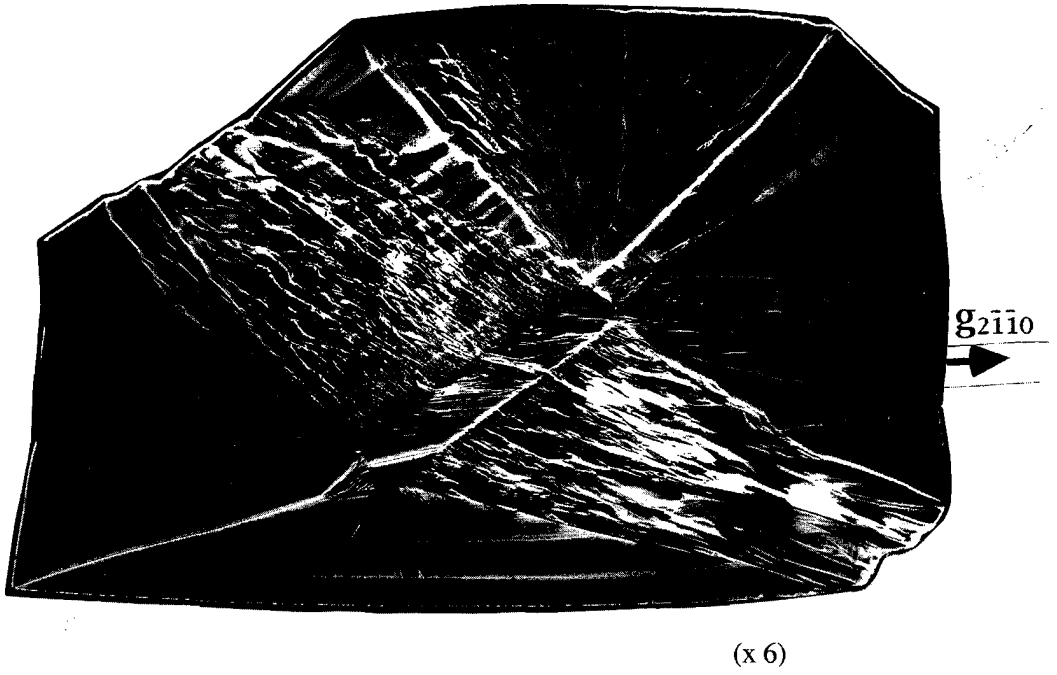


Fig. 2. Conventional CuKα X-ray topograph of Y-cut special bar with (1122) seed.

reflection in the transmitted geometry forms a non-parallel (+,-) setting.[13] In this experiment, the stereographic pair topographs have been recorded by the exposures of $(hkl)_H$ and $(hkl)_O$ images. One advantage of this technique is the possibility of reconstructing the tridimensional network of dislocation lines and/or the growth tunnels[14] across the XY-planes for posterior correlation with some possible appearance of etch-tunnels. For the sample thickness 10 mm, the μt-value is approximately 5. The topographs were recorded on Fuji #50 and Fuji #80 X-ray films with the exposure times 10-30 min. for $(2\bar{1}\bar{1}0)$ reflections.

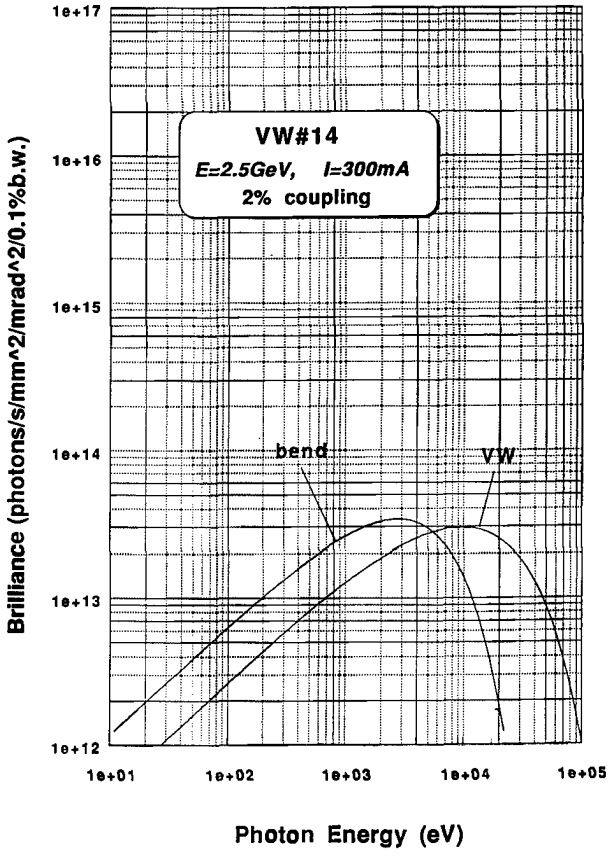


Fig. 3. Wiggler brilliance as a function of photon energy at Photon Factory, KEK.[12]

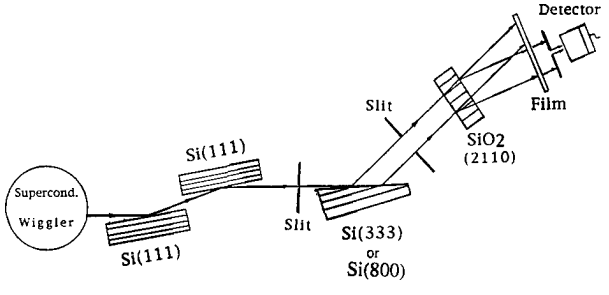
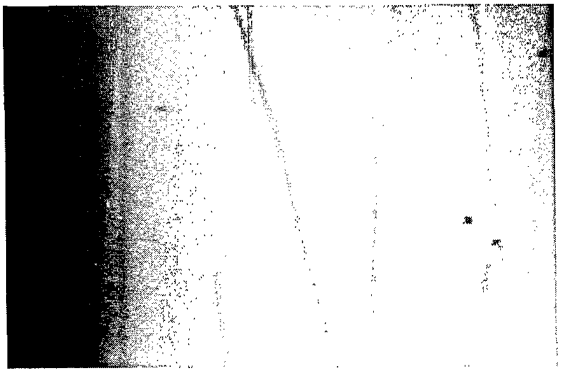


Fig. 4. Experimental setting of synchrotron radiation topography.

RESULTS

Figures 5 (a) and 5 (b) show the $(2\bar{1}\bar{1}0)$ stereographic pair topographs of unswept synthetic quartz for diffracted and transmitted images, respectively. The dislocation lines are less than 10 lines/cm^2, and the white-black contrast of these images are inverted in the diffracted and direct images. Plane wave dislocation topographic image can present two lines consisting on the propagation of X-ray wave along the $\mathbf{k}_H$- and $\mathbf{k}_O$- directions, which are the wave vector of O- and H-waves.[14] In our case, after the propagation of X-ray beam through the distorted regions of dislocations, it flows to the perfect region outside the strained region, and the wave spreads over a large area of the dispersion surface. Therefore, the dislocation images have much better contrast when the dislocation is close to the exit surface of the beam, that can be observed in Fig. 5 (a). The rectangular small contrast are solid inclusions (I). It is interesting to observe the double lines dislocation contrast of direct image (Fig. 5 (b)). From this stereographic pair, the tridimensional configuration of the dislocation network in the sample can be geometrically reproduced in Fig. 6. ABCD is the entrance surface, and EFGH is the exit surface of the X-ray beam. It is possible to localize with good accuracy the intersections of the dislocation lines or growth tunnels with the





(b) direct image

Fig. 5. Synchrotron radiation stereographic pair topographs of unswept synthetic quartz.

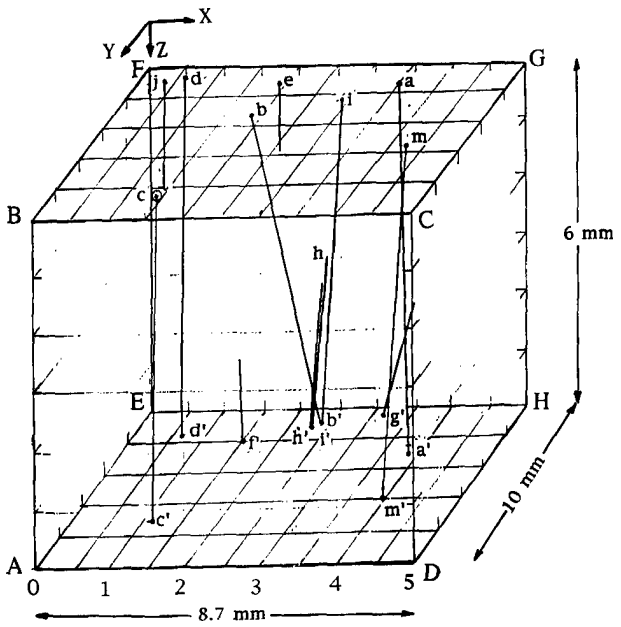


Fig. 6. Reproduction of dislocation network in synthetic quartz from the stereographic pair topographs of Fig. 5.

top and the botton surfaces (XY-surfaces).

The effect of Ag-diffusion are presented in the (2110) topograph of Fig. 7. The vertical lines along the Z-direction are the "diffusion paths" caused by a very localized micro-strain, which suggests the presence of Ag-atoms (Fig. 7 (a)). It is possible to observe that the diffusion process is not homogeneous in the XY-plane. The concentration of Ag is 4.8 ppm in this sample (specimen 3 of Table 1), which is of the same order of Al-concentration, 4.4 ppm, in this same sample. Therefore, the possibility of Ag^{+} as a charge compensator of Al^{3+} can be considered. The transmission topograph shows a weak contrast of layered diffusion of Ag. The double lines of the dislocation images (Fig. 7 (b)) indicate no change concerning some possible decoration effect by the Ag-atoms in comparison with the unswept crystal. The increase of Ag-concentration to 10.8 ppm by applying a shorter diffusion time (2 h), causes a sufficiently high lattice strain to destroy the contrast of the diffusion paths (Fig. 8 (a)). This result shows that the Ag-atoms contribute for the image formation of the diffusion paths of Fig. 7 (a). In the direct image of Fig. 8 (b), the dislocation images are blurred, which indicates that decoration effect has occurred to a certain degree.

For Au-diffusion, there is also a "diffusion paths" contrast along the Z-direction (Fig. 9 (a)), similar to the Ag-diffusion. The concentration of remanent Au in this sample is very small (10 ppb), but the contrast of the diffusion paths is considerable. The direct image (Fig. 9 (b)) shows that Au-diffusion does not follow the layered diffusion process, which is different from Ag-diffusion.

The amplified direct images of dislocations (Fig. 10) show in detail the double lines of unswept quartz, which are still maintained after sweeping with Ag- and Au-diffusion. This result indicates that there is no detectable decoration effect of the dislocation cores in the case of 24 h diffusion.

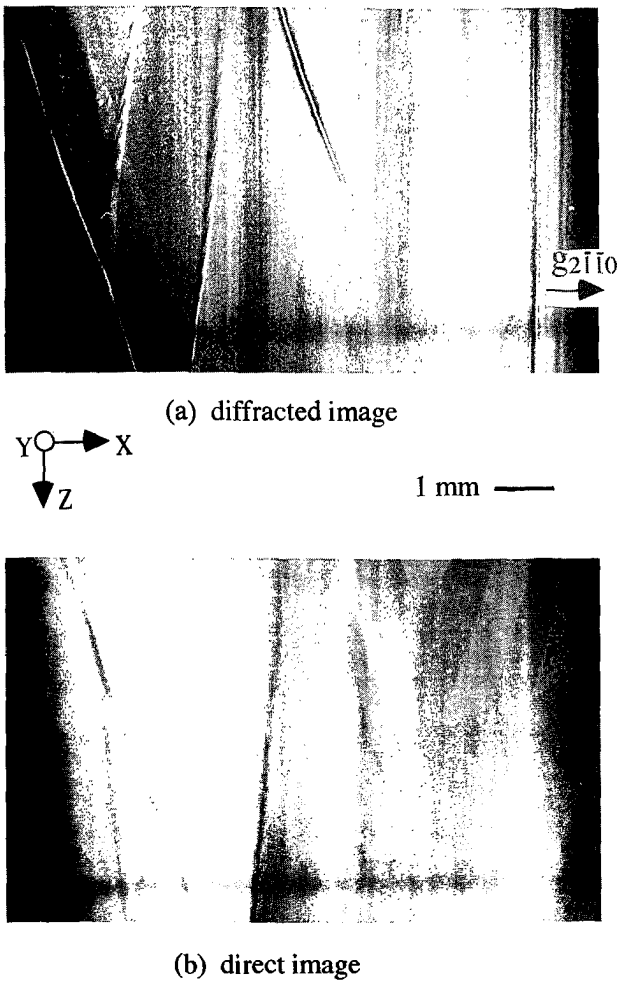
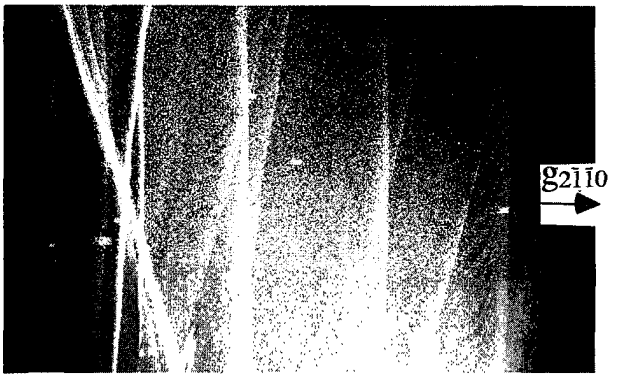


Fig. 7. Synchrotron radiation topographs of Ag-diffusion swept quartz. The "diffusion paths" are parallel to to the Z-direction. Remanent Ag-concentration is 4.8 ppm.

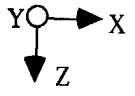
ETCH-CHANNEL SUPPRESSION MECHANISM

Various hypothesis have been proposed by many authors to explain the sweeping effect to the etch-channel suppression. Some of these hypothesis are:

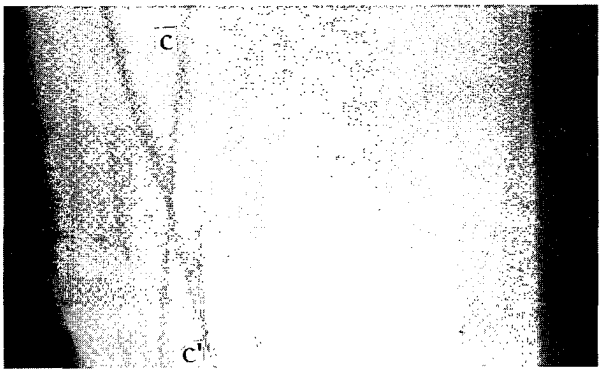
- the presence of Ca^{++} and Mg^{++} impuritities in the neighbourhood of dislocation lines, whose strain energy can be minimized by electrodiffusion;[10]
- related to the model of the oxygen atoms in a silica tetrahedron exposed to a surface unbounded to the substructure,[15] the possibility of swept ions to become bounded to these non substructure-bound oxygen[9] can be considered;
- the possibility of electromigration of impurities into vacancies bound to dislocation or into growth tunnels make them unreactive to etchant;
- the electromigration of Au or Ag in quartz under various sweeping conditions of temperature and electric field, conducted in air and in vacuum, shows a similar result in avoiding the diffusion of airborne impurities.



(a) diffraction image

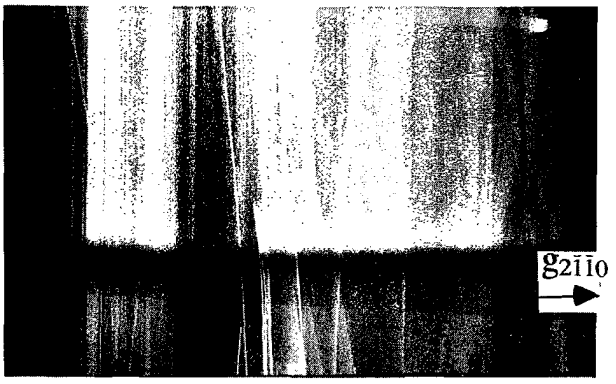


1 mm

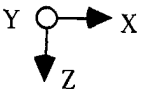


(b) direct image

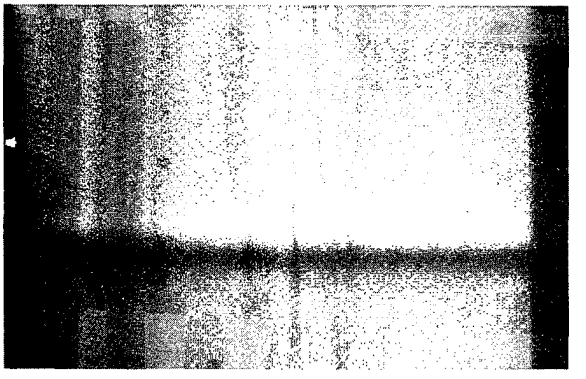
Fig. 8. Higher concentration of remanent Ag-concentration (10.8 ppm) in synthetic quartz. The diffusion paths are not observed due to the lattice strain.



(a) diffracted image



1 mm



(b) direct image

Fig. 9. Synchrotron topographs of Au-diffusion swept quartz. "Diffusion paths" can be observed.

The synchrotron topography shows the presence of remanent Ag (Au) atoms along the diffusion paths, that intercepts the dislocation lines in all their extension. The metal-complex bound with the SiO_2 structure is not clear, but the presence of Ag (Au) at the diffusion paths-dislocation line interceptions can inhibit the etch-channel generation.

CONCLUSIONS

Based on the synchrotron radiation topographic observation of Ag- and Au-diffusion, the following conclusions can be presented:

- the electrodiffusion of Ag and Au is not homogeneous but they are quite localized along the "diffusion paths" parallel to the Z-direction;
- the diffusion process of Au is different from Ag, which is layered;
- the content of remanent Au is of the order of ppb, in comparison with the case of Ag which is of the order of ppm, but the contrast of lattice strain of the diffusion paths are practically equivalent; that means the bound site of Au in the SiO_2 structure is different from Ag;
- for 24 h diffusion in air for Ag and Au, no detectable decoration effect of dislocation lines has been observed;
- the presence of Ag (Au) at the interceptions of the diffusion paths with dislocation lines inhibit the etch-channel generation.

ACKNOWLEDGMENTS

The authors acknowledge the encouragement and continuous support of Professor Kazutake Kohra, and also Professor Y.E. Nagai for his suggestions and critical reading of the manuscript. One of the authors (C.K.S.) would like to acknowledge his participation in the JAERI-RIKEN SPring-8 Project Team during which SR experiment on quartz was conducted, and the support of Sawyer Research Products, Inc. CKS also acknowledges the finantial support of FAPESP and CNPq.

REFERENCES

1. S. Keerti and A.R. Lang, "Crystal perfection and silver electromigration paths in alpha-quartz", J. Appl. Cryst. vol. 5, pp. 72-78 (1972).

2. A. Zarka, M. T. Sebastian, and B. Capelle, "X-ray topographic study of vacuum swept quartz crystals", Proceed. of 42nd Annual Frequency Control Symposium, pp.208-210 (1988).

3. M.T. Sebastian, "X-ray diffraction study of crystal under a static electric field", Prog. Crystal Growth and Charact., vol. 27, pp. 233-278 (1993).

4. J. Asahara, K. Nagai, K. Hamaguchi, S. Taki, C. K. Suzuki, X.W. Zhang, and Y. Yoda, X-ray topography of thick synthetic quartz crystals", presented at the Symposium on Synthetic Crystals, Osaka, November 18, 1992.

5. J. J. Martin, "Electrodiffusion (sweeping) of ions in quartz: A review", IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control, vol. 35, pp. 288-296 (1987).

6. J. R. Vig, J. W. LeBus, and R. L. Filler, "Chemical Polished Quartz", Proceed. of 31st Annual Frequency Control Symposium, pp.131-143 (1977).

7. G.B. Krefft, "Effects of high temperature electrolysis on the coloration characteristics and OH-absorption bands in alpha-quartz" Radiation Effects, vol 26, pp. 249-259 (1975).

8. O. Ishii, T. Morita, T. Saito, and Y. Nakazawa, "High frequency fundamental resonators and filter fabricated by batch process using chemical etching", in these Proceedings.

9. J. C. Gualtieri, "The influence of temperature and electric field on the etch-channel density in swept cultured quartz", Proceed. 39th Annual Frequency Control Symposium, pp. 247-254 (1985).

10. J. Asahara, K. Nagai, K. Hamaguchi, H. Sone, and S. Taki, "An analytical study of electrodiffusion (sweeping) of synthetic quartz", Proceed. of 45th Annual Frequency Control Symposium, pp. 9-21 (1991).

11. B. K. Tanner, "X-ray diffraction topographs", Oxford: Pergamon Press, 1976, ch. 3, pp. 63-99.

12. H. Kitamura, ed., "Insertion Device Handbook 1990 Photon Factory", National Laboratory for High Energy Physics, (1990) pp. 101-111.

13. K. Kohra, M. Yoshimatsu, and I. Shimizu, "Direct observation of imperfection in crystals", Ed. J. B. Newkirk and J. H. Wenick, Interscience, pp. 461-470 (1962).

14. S. Kats, A. Halperin, and M. Schieber, "Growth tunnels in quartz crystals", Proceed. of 36th Annual Frequency Control Symposium, pp. 193-196 (1982).

15. T. Ishikawa, S. Kikuta, and K. Kohra, "Angle resoved plane wave X-ray topography", Jpn. J. of Appl. Phys., vol 24, pp. L559-L562 (1985).

16. F.M. Ernsberger, J. Phys. Chem. Solids, vol 13, p.347 (1960).

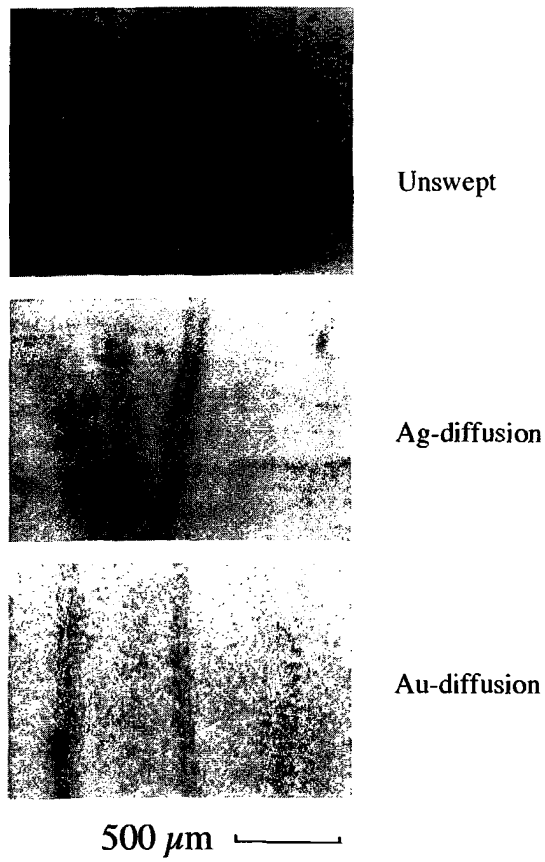


Fig. 10. Double lines contrast of dislocation in the direct images.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SPECIFIC FEATURES OF QUARTZ CRYSTALS LAMELLAR STRUCTURE

N. M. Bauer, A.P. Pogrebnyak, S. N. Abdrafikov, N. A. Mamaev, B. V. Shulgin

The plant "Kristall", Yuzhnouralsk, Russia

It is considered, that reorientation of the growth surface from flat to a stepped one leads to irregular distribution of impurity centres [1, 2]. Theoretical models of heat and mass transfer process kinetics feature the possibility for impurities to be mainly located in the area of incoming angle of the steps (speaking about macrosteps).

So to say, the growth of macrostep should be accompanied by formation on the boundary with substrate of a layer with higher density of impurities comparing it with the array. However, the experimental studies of this phenomenon are quite complicated and up until now, it has not been completely investigated.

It is not clear yet, namely, how effective is the macrostep in redistribution of impurity centres, as well as what is the thickness of the newly formed impurity layer and what factors define it.

This report gives results of experimental investigation of structural Al distribution in quartz crystals grown on seeds making desired angle with the face $\{10\bar{1}0\}$ (nonsingular faces (NF) of the zone [0001]).

It was shown before [3] that regeneration growth pyramids of these crystals possess a stepped growth surface formed by faces $\{\bar{1}120\}$ and $\{10\bar{1}0\}$ areas located in turn and making a less angle with the initial surface. Moreover, certain relationship between the average sizes of the steps and the initial orientation of NF was discovered. Substructure of these crystals and distribution of impurity centres in them are of specific nature, and this fact, to the authors' mind, is to certain extent linked with the problem under discussion and can be of scientific and practical interest.

OBJECTS AND INVESTIGATION METHODS.

Synthetic quartz crystals (SQC) have been investigated which were grown in hydrothermal solution of Na_2CO_3 (1 mol/l, T=623 K, ΔT=15°C, P $\geq$ 700 MPa) on seeds (30x40x3 mm) forming desired angles with the face $(10\bar{1}0)$. The orientation of seeds was determined by the angle of their cutting with the step of 5 degrees within the angle limit $0 < \theta < 30$. Accuracy control of the seeds' cutting angle after grinding of their surfaces showed deviation $\pm 10'$ in axis [0001] and $\pm 15'$ in a perpendicular direction. All the seeds were made from crystal growth <C> pyramid of the earlier generation. The seeds were placed in one and the same zone of autoclave to ensure comparable condition of crystallization. Crystals were grown in

0-7803-2500-1/95 $4.00 © 1995 IEEE

conditions when correlation between the growth rate in an optical axis direction and growth rate of a positive trigonal prism was $V_c/V_x = 1.08$ (agreed as group A crystals). Also some crystals grown in comparable conditions with adding of 3 g/l Li NO_3 in soda solution (group B crystals).

Plane-parallel Z-cut plates were sliced from the middle part of crystals, their surface having undergone grinding and polishing (up to optical requirements). The final thickness of plates was 0.3, 0.5, and 3 mm. The samples with different thickness were addressed to tackle different problems. So, 0.3 mm plates were used to make X-ray topograms (Lang variant) giving impression about the distribution of structural heterogeneities in growth pyramids being investigated (prior to this, the samples were etched for 5 minutes in 30% solution of HF to remove the surface layer damaged by mechanical processing). Plates with thickness of 0.5 mm had undergone further slicing. 2-3 samples of restangular profile with area of 0.5 cm^2 were cut from NF growth pyramids (mass differences of the samples not exceeding 5 mg). They were used directly to register thermal luminescence curves (TL). Samples with the thickness of 3 mm were used for optical spectral analysis and solving other problems.

TL curves were registered using irradiated specimens (^{60}Co, dose $2x10^5$ Gy with linear temperature rise at the rate of 0.1 degree/hr in the range of 300-650°K. Temperature was controlled by means of thermocouple PP-1. Heating was performed in a vacuum chamber of special design, the operating pressure being 0.1 Pa. In the centre of the removable cover of this chamber there was a window of 25 mm diameter made of high quality quartz glass. TL peaks were registered through this window by means of photoelectronic device with the range of spectral sensitivity 160-600 nm (PED-71).

Signal from the PED, having been amplified and processed by computer, was registered by potentiometer KCП-4. The operating conditions of the amplifier were well-calibrated, and that made it possible to ensure reliable registation of TL signals with varied intensity, namely, 10,000-100,000 times. Values of peak intensity were specified for mass units so as to compare TL curves of different specimens, that is, TL curves hither to enclosed give the idea about luminescence outcome from the mass unit of the specimen under investigation.

With the purpose of making additional investigations following similar technique, some Y-cut specimens with the thickness of 0.5 mm were made from <C> pyramid crystal growth of group A. They were used when analysing dose changes of the TL curves.

INVESTIGATION RESULTS.

Radiation influence on quartz leads to crystal structure destruction (amorphing of sample) or radiation stimulated redistribution of impurity centres (a combination of both processes is also possible) [4,5]. Following observation of γ-irradiation of isotope ^{60}Co causing mainly changes in the electronic subsystem of the crystal, not leading to some evident effects of the basic structure of the nuclea dislocations from their standard positions, which is characteristic of reactor irradiation. However, application of the TL method for evaluation of investigated specimens on the basis of impurity centres, requires solving the problem of irradiation dose, since the process of their radiation redistribution is of a probabilistic character and can be multi-stepped.

Specimens from the <C> pyramid crystal growth (group A) were analysed for correlation between TL curve profile and irradiation dose. The average from 3

samples' TL curves corresponding to different doses is given in Fig.1. The results indicate evident changes in TL curves profile with increasing of irradiation dose. These changes show themselves in the number of registered TL peaks with different doses as well as in their intensities correlation. Four rather weak peaks with maximums corresponding to temperatures of 350, 413, 445, 523° K (curve 1, Fig.1) are registered on the TL curves of the specimens being investigated at relatively small doses. Their intensity increases more than 10 times with a dose of 10^4 Gy, though temperature scale location remains the same. When increasing irradiation dose 10 times more, considerable reducing of the first three peaks' intensities on temperature scale is observed with critical increasing of the last peak intensity, its maximum having shifted to the area of higher temperature (curve 3, Fig.1).

Table 1 shows a dramatic correlation between quantitative changes in TL parameteres and irradiation dose. T_m in Table 1 corresponds to the maximum temperature of the corresponding peak; value I, its intensity in relative units at this point; value E corresponding to activation energy in eV, defined by the half width of its peak and temperature of its maximum. TL peaks are enumerated as they follow each other on the temperature scale.

By analyzing Table 1 it becomes clear that dose changes of the second and third TL peak intensities are of similar character. No critical temperature shifts of these peaks are observed when increasing the dose. As for the fourth peak, its intensity is constantly increasing with increasing of the irradiation dose and its maximum shift to the area of higher temperature. It is noteworthy that these shifts become more evident when intensities of the second and third peaks are reducing. It is should be mentioned that dose changes of TL curves are studied on specimens of identical orientation, thickness, sectorial characteristics, the specimens having been taken from the crystals grown in comparable conditions.

Fig.2 shows specific TL curve specimens taken from NF growth pyramids of different orientation, irradiated by $2x10^5$ Gy dose. Two peaks with the maximum being 360 and 365° K are registered on TL curves for crystals of both groups, peaks of group B crystals being of comparable intensity values, while the second peak intensity of group A crystals is much more than that of the first one.

Studying TL curves on the specimens taken from both groups of crystals, a specific orientation effect can be traced, that is, consecutive changes of both peaks' intensities following changes in orientation of NF. Intensities of the peaks decrease following an increase in θ angle. Table 2 gives the idea about relative value of orientation effect; there you can find integral intensities of peaks I_1 and I_2 in relative units; their activation energy of luminiscence E_1 and E_2 in electronvolts; as well as the average size of growth steps of NF with corresponding orientation on data basis (3) expressed in mkm. Table 2 gives TL parameters of the duplicating specimens for each NF orientation. Integral intensity peaks' values of TL as regards crystal groups A and B are given in relative units, which makes it possible to compare them to each other.

Studying Table 2, we find out that there is an orientation effect in crystals of both groups. This effect can be traced not only on duplicate specimens, but can also be traced when repeating the experiment after annealing of the specimens and their repeated irradiation TL characteristics of group A crystals with $\theta = 10°$ orientation should be noted. Their first peak's intensity corresponds to the general correlation tendency between this TL parameter and NF

orientation. As for the second peak's intensity it is less than similar TL peaks' intensity for orientation θ is 15° and 20°. Impurity centre's density might be very high in this case, causing density suppression (luminescence outcome reduces with increasing of centre's density). TL curves of group B crystals may prove it. Intensities of their peaks, including specimens from NF 10° growth pyramids, follow a general pattern of orientation effect. Table 2 lacks the data of NF 5° (group A crystals). NF growth pyramids turn out to be rather narrow in this case, which makes it impossible for the specimens to follow a chosen pattern.

Structural study shows that NF growth pyramids of the zone [0001] of SQC possess lamellar structure. There are a lot of bands of dark contrast in them (Fig.3) revealed by method of x-ray topography. They are not inherited from the seed, but are formed at the initial growth stage in an area close to the seed. Their bands are stretched across NF pyramid. For different NF orientations they are parallel to hexagonal prism face, making a smaller angle with the seed. The density of these bands is rather great. They are rarely found alone, even in specimens with thickness of 0.3 mm. Their combinations, which look like wide bands of higher picture contrast, are usually seen on topograms.

Defects of this type are found in different reflections. They are seen most clearly with symmetrical reflection when diffraction vector is perpendicular to them. With asymmetric reflections, in reflections of type $(10\bar{1}1)$, their picture clearness becomes problematic, indicating stretching of these defects across the specimen and in the optical axis direction. Studying topograms, we find also some defects in NF growth pyramids' orientation, which is close to the direction of the crystal's electric axis (Fig.3.). They are partially inherited from the seeds. The density of these defects is rather small and varies in different crystals from 50 up to 100 per cm^2. They can be considered as linear defects by the picture contrast. Their local changes of orientation can be found in the points of intersection with the above mentioned defects. That's why they cannot be strictly linear. When magnifying topograms 20-30 times, they look like broken lines of a complicated profile. When hydrothermal etching is applied, a lot of etching bands are formed on the specimens' surface in NF pyramids, separating them into zones (lamellas). On its way into the depth of the crystal, the width of etch such lamella is not changed markedly as a rule (Fig.3-c). The sizes of different lamellas relative to each other for each certain orientation may vary markedly. The lamellas are parallel to each other and in Z-cut planes they coincide with face $\{10\bar{1}0\}$ orientation, the face being the closest to NF.

Hydrothermal etching reveals also thin structure of lamellas. Small bands of etching are found which usually start and end on the opposite boundaries of each lamella. Normally they are not inherited from one lamella into another. There does not exist any interconnection with regard to distribution of small grooves of etching in neighbouring lamellas, that is, the substructure of each lamella is autonomous. Small bands of etching make an angle of 30° with the boundaries of the lamellas. Thus, their orientation coincides with that of steps but with reconstructed profile of NF [3].

RESULTS REVIEW AND CONCLUSIONS

Geometric correlation between defects having growth origin on topograms and stretched boundaries between lamellas on pictures of hydrothermal etching, proves their common nature. Evaluation of average sizes of lamella images using pictures of hydrothermal etching for NF orientations 15° and 20° correlates quite well with the step

height of their reconstructed profile [3]. Therefore, lamellas structure of NF zone [0001] growth pyramids of SQC is called forth by the step profile of their growth front and is the result of autonomous formation of steps. Since areas of steps formed by face [10$\bar{1}$0] are passive while growing, they serve as substrate for the growing of each step. That is why lamella boundaries possess the above-mentioned orientation in NF growth lamella while moving along the substrate in a tangential way, the size of which corresponds to the height of the step (Fig.4-a). The fact that boundaries of growth lamellas are detected by different methods, proves imperfection of the crystal in the boundary area, that is, the conjugation area of the forming macrostep with substrate may markedly vary from the rest of growth lamella, speaking about structural characteristics. This effect may be caused by several reasons, including redistribution of impurity centres on the stepped growth front. Since appearance of impurities distribution in height is not known a priori in the area of incoming angle, it can be considered as having approximately a rectangular profile (Fig.4-a), where r_0 corresponds to step size with impurity density as N_0 specific for arrays with the given conditions of crystal formation, and features conjugation area of the step with substrate where impurity density is N_x. To be more exact, it is considered that $n_x > n_0$ $\delta << \Gamma_0$, $\Gamma_0 + \delta = r$ where r = step height.

Simple calculations show that in model as represented in Fig.4a, average impurity density of specimen will be defined by the expression:

$$n = n_0 + (n_x - n_0) \times \delta / r \ldots (1)$$

If we assume that values n_o, n_x and δ do not depend on r (which is quite sensible for macrosteps), then from (1) it follows that the average impurity density in crystal will depend on step size by which it is formed. Thus growth pyramids of NF being researched and formed by steps with different height (Table 2), may have an average density of impurities. Value $(n_x - n_o) \times \delta$ in equation (1) features activity of macrostep.

It follows from estimation, the empirical relation between integral intensity of TL peaks (namely, peaks T_2) and density of steps (value 1/r) can be described by the following equation $I = I_o + I_x/r$ (2)

where $I_o = 0.29 \pm 0.03$

$I_x = 6.66 \pm 0.90$

We assume that the intensity of the second peaks of TL in crystals being researched is proportional to the average density of "smoky" centres, which serve in these crystals as centres of recombinational luminiscence, then parametrical correlation comes from (1) and (2):

$$n_x/n_o - 1 = I_x/I_o \times 1/\delta$$

Independent estimation of value δ makes it possible to define correlation n_x/n_0 from (2) and (3). On the pictures of hydrothermal etching, the boundary area between lamellas is etched like a band of 3 mkm width. Taking into consideration that areas close to the surface are etched more, the value can be 1.

Then the correlation n_x/n_0 will $\sim$ 20, that means that density of the "smoky" centres in boundary areas will be at least 20 times more than their density in the array. Performed researches make it possible to conclude:

1. Orientation effect has been revealed regarding "smoky" centres distribution in z-zone of SQC growth pyramids displaying itself in reducing of its average density with corresponding increase of the seed cut angle.

2. It is discovered that NF growth pyramids of the zone [0001] of SQC possess lamellar structure caused by a stepped front of their growth. It is shown that in the area of macrosteps incoming angle of crystal under investigation, the density of "smoky" centres (lowest values are taken) is 10 times more than that in the array.

LITERATURE

1. R.L. Parker; Crystal Growth Mechanismus: Energetics, Kinetics and Transport. - Solid State Physics, v. 25, Academic Press, New York and London, 1970.
2. A. A. Chernov; Formation of Crystals. - Modern Crystalography, Moscow, Nauka, 197, v. 3, p. 3- .
3. N. A. Mamaev, A. F. Kuznetzov, A. F. Zatzepin, B. V. Shulgin; Nonsingular Faces Reconstruction of Quartz Crystal in Hydrothermal Conditions, Crystalography, 1987, v. 32, p. 196-202
4. J. A. Weil; A review of Electron Spin Spectroscopy and It's Application to the Study of Paramagnetic Defects in Crystaline Quarts. - Phys. Chem. Minerals, 1984, v. 10, p. 149-165.
5. A. R. Silin, A. N. Truchin; Point defects and elementary excitation in crystalline and became glassy Si O . - Riga: Zinatne, 1985, p. 244.

Table 1.
Effect X-irradiation dose on the thermal luminescence parameters of artificial quartz crystals.

Peaks No	Peaks parameters	X-irradiation dose, Gy 10	10		
	T max, K	350	348	346	345
1	I,intensity	0.012	0.110	0.075	0.035
	E, eV	-	0.30	0.30	0.30
	T max, K	413	413	413	-
2	I,intensity	0.008	0.09	0.025	-
	E, eV	-	-	-	-
	T max, K	445	445	448	-
3	I,intensity	0.010	0.110	0.035	-
	E, eV	-	-	-	-
	T max, K	523	525	535	543
4	I,intensity	0.008	0.180	0.220	0.2
	E, eV	-	0.95	0.95	0.96

Fig.1 Changers in the thermal luminescence curves with the radiation dose increase

1-doze 10^3 Gy,
2-doze 10^4 Gy,
3-doze 10^5 Gy.

Table 2.
Quartz crystals thermal luminescence parameters in nonsingular faces growth pyramids of the zone [0001].

Crystal	O	I,intensity in rel unit	E, eV	I, intensity in rel unit	E, eV	r, mkm
	5	-	-	-	-	11.9
	10	0.071; 0.075	0.2	0.43; 0.40	1.1	15.9
	15	0.057; 0.059	0.2	0.98; 1.056	1.1	20.9
A	20	0.031; 0.044	0.2	0.66; 0.731	1.0	32.0
	25	0.015; 0.016	0.2	0.452; 0.477	1.0	100.3
	30	0.006; 0.011	0.2	0.312; 0.294	1.0	-
	10	0.118; 0.128	0.2	0.131; 0.153	1.0	
B	15	0.105; 0.091	0.2	0.116; 0.099	1.0	
	20	0.066; 0.071	0.2	0.058; 0.068	1.0	

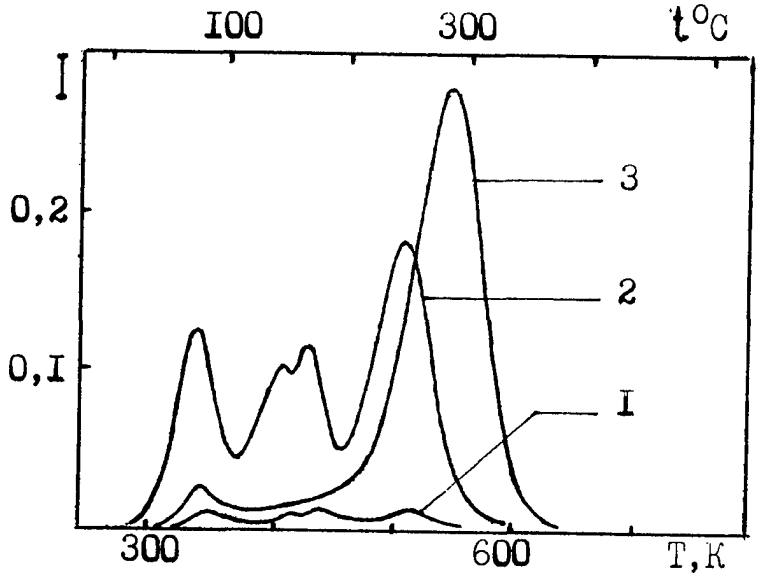


Fig. 1 - Changes in the thermal luminescence curves with γ-radiation dose increase.
1 - dose 10_3 Gy; 2 - dose 10_4 Gy; 3 - dose 10_5 Gy

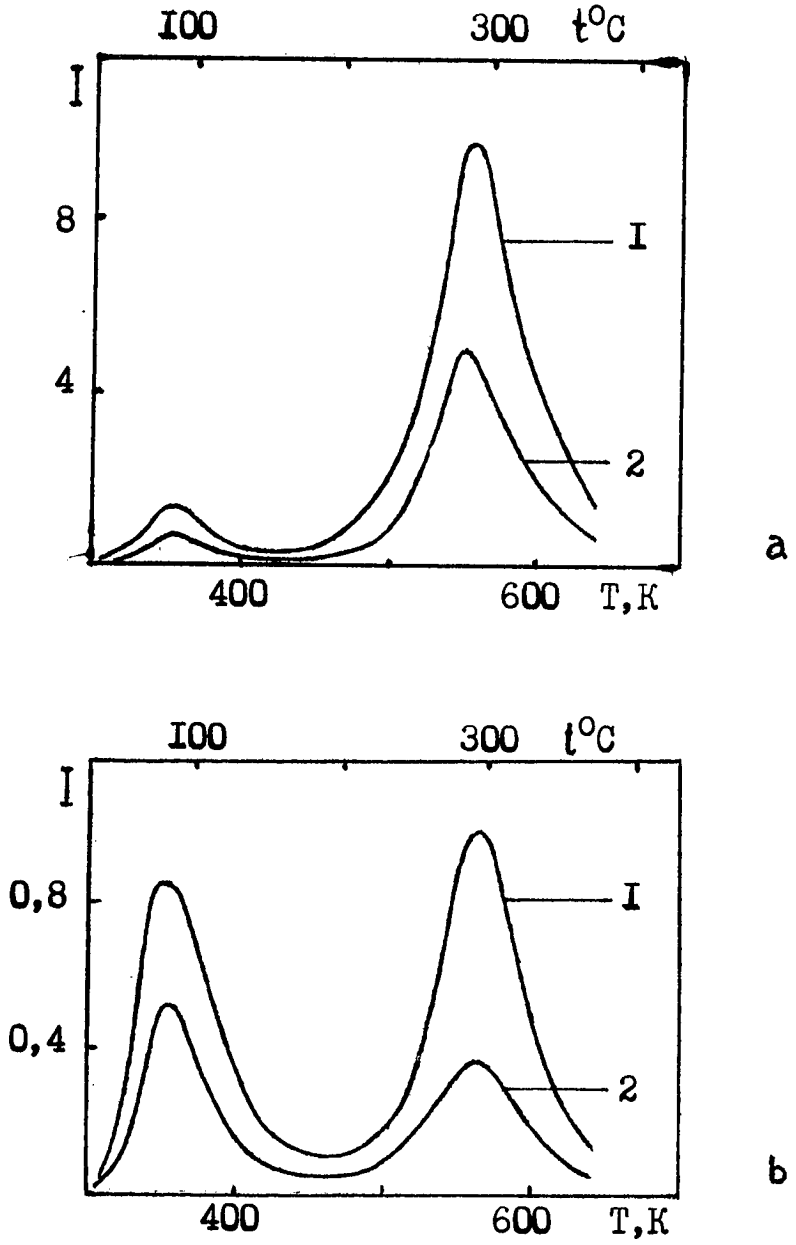


Fig.2. Thermal luminescence curves for different orientations of nonsingular faces growth pyramids;
a) crystal A, cut (0001), thickness 0.5 mm, dose - $2*10^5$ Gy; 1 - $\theta=15^\circ$, 2 - $\theta=25^\circ$;
b) crystal B, cut (0001), thickness 0.5 mm, dose - $2*10^5$ Gy; 1 - $\theta=10^\circ$, 2 - $\theta=20^\circ$.

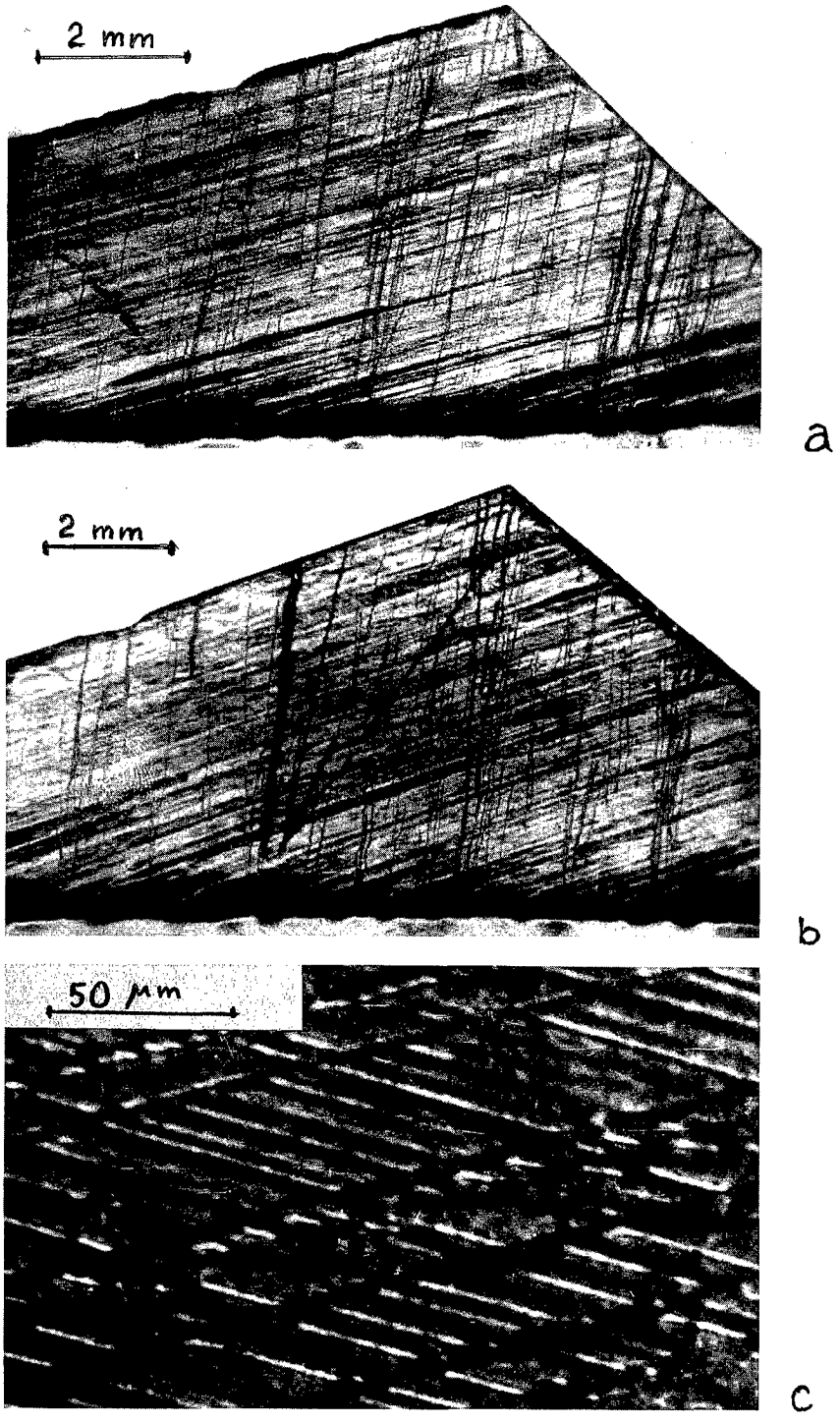


Fig.3. Lamellar structure of the zone [0001] nonsingular faces growth pyramids in syhthetic quartz;
a) X-ray topogram of the crystal B z-cut, $\theta=15^{\circ}$, thickness 0,3 mm, Mo-K_{α} - radiation, reflaction $(\bar{1}\bar{1}20)$;
b) X-ray topogram of the crystal B z-cut, $\theta=20^{\circ}$, thickness 0,3 mm, Mo-K_{α} - radiation, reflaction $(\bar{1}\bar{1}20)$;
c) hydrothermal etching pattern of the crystal B z-cut, $\theta=15^{\circ}$.

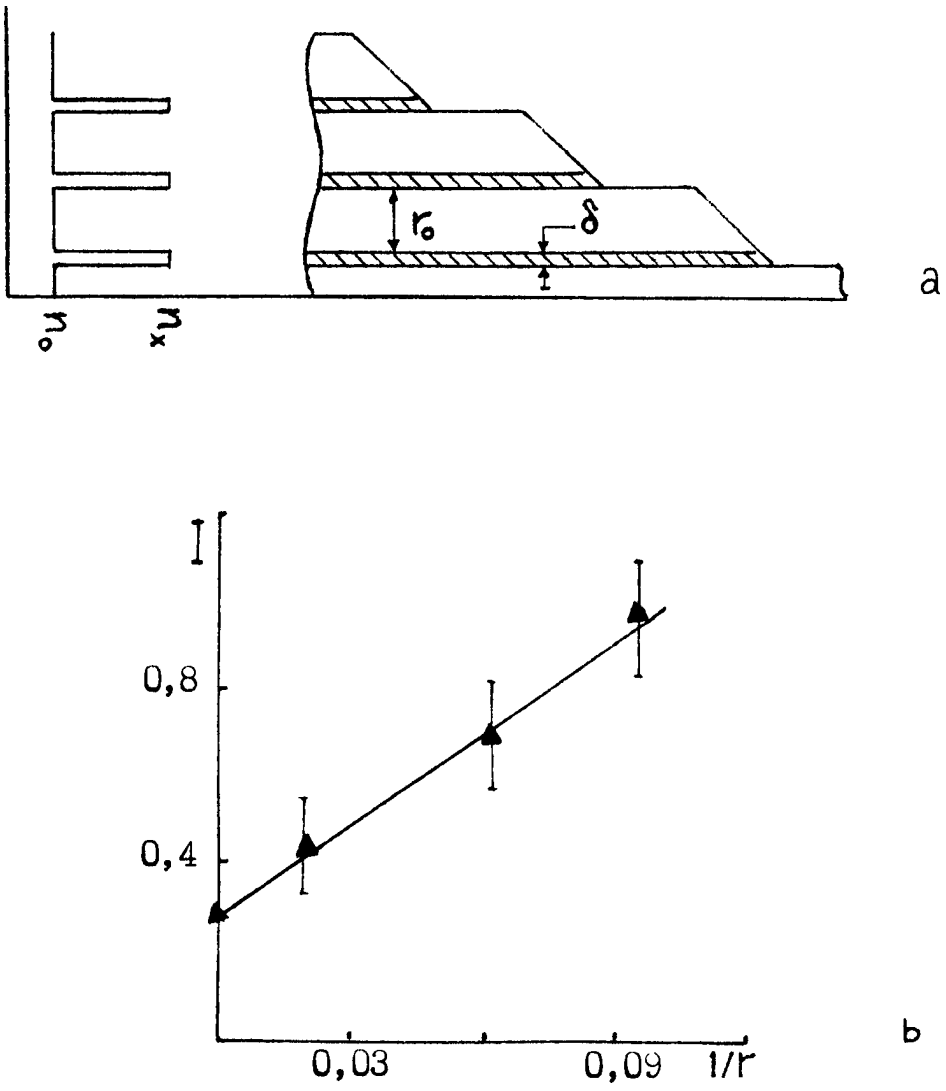


Fig. 4 - Impurity centers distribution on the stepped growth surface:
a) scheme of impuriry centers distribution
b) relation between thermal luminescence peaks integral intensity and steps density.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THERMAL DESORPTION OF QUARTZ CRYSTALS

R.A. Murray,[1] D.E. Pierce,[1,2] R. Lareau,[1] S. Laffey,[3] and J.R. Vig[1]

[1]U.S. Army Research Laboratory, Fort Monmouth, N.J., 07703-560, [2]National Research Council Associate,
[3]Vitronics Inc., Eatontown, N.J., 07724

ABSTRACT

Thermal desorption mass spectroscopy was used to measure the outgassing of water and ammonia from quartz wafers after different processing steps typically used in resonator fabrication. The activation energies of the two species were calculated and compared with a test bakeout. It was determined that the water could be represented well by multiple activation energies. The effect this has on bakeout parameters is discussed. Also discussed are the possible sites where water and ammonia may be bound to the quartz surface. It was determined that part of the ammonia that was detected in these experiments originates in the vacuum system. Although UV ozone cleaning is primarily used to remove carbonaceous contaminants, it was found to substantially reduce the amount of ammonia desorbing from quartz samples.

I. INTRODUCTION

The aging of quartz resonators is one of the important problems that persists in limiting the performance of high precision oscillators. Although there is no theoretical lower limit to aging, at present the best attainable in a real resonator is approximately 10^{-12}/day [1]. The primary causes of aging [2] include mass transfer to and from the crystal surface and, possibly, diffusion from the bulk of the crystal. Previously [3], the results of the first experiments using Thermal Desorption Mass Spectrometry (TDMS) to measure the outgassing of alpha quartz wafers were reported. This is a very sensitive method to look at what desorbs from quartz crystals. Thermal desorption of quartz can not only determine contamination on the surface, but also the activation energies of the surface species. The activation energies can affect bakeout schedules and phase noise [4]. It was reported that H_2O, OH, CO_2, minor amounts of hydrocarbons, and NH_3 (ammonia) were found, but, F, H, and H_2 were not [3]. In this paper the investigation focused on water and ammonia. These two species were found to produce the highest total amounts (outgassing rates integrated over time) of molecules in the temperature range of interest, 60° to 573°C.

The outgassing of quartz wafers was measured after different combinations of chemical cleaning, etching, and UV ozone cleaning to determine how the amounts of water and ammonia on the surface change at different points in a typical resonator fabrication process. The hypothesis was presented earlier [3] that the ammonia desorbing from the samples came from one of the processing steps that used ammonia, either in the cleaning or in the etching. Here, the results of experiments designed to trace the source of the ammonia are reported, and ways to reduce its outgassing rate are suggested. Bakeout times necessary to reduce the water and ammonia are examined, along with whether or not any water is diffusing from the bulk.

II. EXPERIMENTAL METHODS

In this set of experiments, we used a 30 Watt CO_2 laser to heat the samples in a vacuum chamber, as illustrated in Fig. 1. The laser was external to the vacuum system and the beam was admitted into the desorption chamber through a ZnSe window, which is not shown, and which is transparent at 10.6 μm, the wavelength of the laser. The sample holder was stainless steel, and the quartz wafers were supported by thin W-Re wires that were gold coated. The laser was focused on the sample so that the sample was the hottest object in the desorption chamber. In earlier experiments, resistance heaters were used inside the vacuum system to heat the samples radiantly, but the outgassing of the

0-7803-2500-1/95 $4.00 © 1995 IEEE

heaters was large compared to the outgassing from the samples. The experiment required the ability to observe low outgassing rates from the samples, which required very low contributions from other sources. The laser system shown provides a clean way of heating the samples. The laser was computer controlled to heat the samples linearly from 60°C to the upper setpoint, (usually 500° to 570°C) in 90 seconds. Experimental evidence indicated that if this temperature schedule was followed, the sample holder did not get hot enough to contribute appreciably to the outgassing.

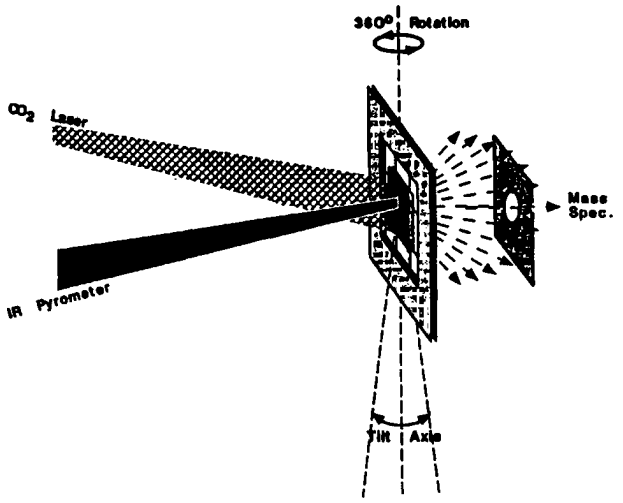


Figure 1: Laser desorption setup.

The temperature was measured with a pyrometer, designed to work with quartz, which has an estimated accuracy of ±10°C. A typical temperature profile is shown in Fig. 2. The laser was turned on and in 90 seconds the sample was heated from 60° to 585°C, then the laser shut off automatically and the sample cooled down. The dashed line from 0 to 15 seconds represents the extrapolated temperature because the minimum sample temperature the pyrometer will detect is approximately 120° C. It was found that samples that were heated to temperatures that exceeded the quartz alpha-beta transition temperature broke, as the sample in Fig. 2 did. This gave a good external check on the temperature accuracy of the pyrometer. The desorption system used two turbo vacuum pumps and could reach pressures in the mid-10^{-9} torr range in under two hours. The quartz wafers were positioned in the line of sight to the mass spectrometer ion source.

The mass spectrometer was a double focusing model with very high resolution, i.e., M/ΔM=30,000. Performing experiments at a higher resolution than necessary reduces the number of ions that are collected and lowers signal strength of the desorption curves so experiments done in this work were performed at a resolution of 1800. This mass spectrometer allowed special techniques to be used under certain circumstances that improved the signal-to-noise ratio of the desorption data. If the masses are close enough, the different ions could be accessed by electrostatic scanning, which can be done much faster, and with less noise, than normal mass spectrometer quadrupole operation. Water and ammonia were the species of interest, so instead of measuring the main water peak, at 18.015 atomic mass units (AMU), and the main ammonia peak, at 17.027 AMU, a different approach was used. When water is ionized in the mass spectrometer it breaks into fragments, such as OH and O with a fixed percentage. For example, the OH represents 21% of the H_2O [5]. The mass of the OH is 17.003 AMU, which was much closer to the main peak mass of ammonia. So OH was used instead of H_2O in all the desorption experiments described in this paper. Several times during the experiment the OH to water ratio was checked to make sure the 21% ratio was still holding, and it was, indicating that all of the OH came from H_2O. That is, none of the OH that was observed was desorbed from the surface in the form of OH. In all the experimental data that are presented here, when an amount of OH outgassing is discussed, the amount of water can be obtained by multiplying the OH value by five.

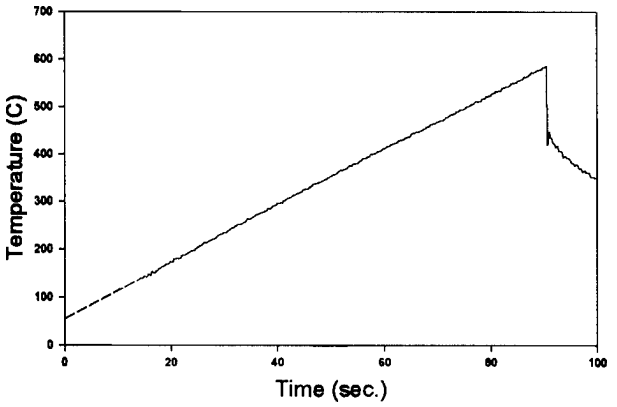


Figure 2: Temperature profile.

All of the samples used in this study were AT-cut wafers from a single bar of sodium carbonate grown quartz ($alpha_{3500}$ = 0.08). In order to determine the effect of cleaning and etching on outgassing, the wafers were prepared as if they were being made into resonators. The desorption was then measured at different parts of the process. All of the samples were prepared by lapping

with 5- then 3- μm alumina grit. Over the years, various cleaning procedures for wafers have been adopted [6-8]. The cleaning procedure used for preparing these samples was as follows. After lapping, the wafers were boiled in trichloroethane, then boiled in isopropyl alcohol. They were then cleaned in Standard Cleaning Solutions 1 (SC-1), which is six parts water, one part ammonium hydroxide, one part hydrogen peroxide, and then in SC-2, which is five parts water, one part hydrochloric acid, and one part hydrogen peroxide. A few of the samples were processed without SC-1 to provide ammonia free preparation. The chemical cleaning of the wafers in SC-1 and SC-2 was done at 80°C. The wafers were rinsed four times in deionized water at 80°C after the SC-1 cleaning and again after the SC-2 cleaning. If the wafer was to be etched, it was UV ozone cleaned for 20 minutes on each side before etching. Since all of the wafers used were AT-cuts, saturated ammonium bifluoride at 75°C was used for the etching. After etching, the samples were rinsed four times in deionized water at 80° C and blown dry with UHP N_2. Some of the samples were UV ozone cleaned without subsequent etching. The wafers were then put immediately into the TDMS.

III. RESULTS

The first set of samples were chemically cleaned in SC-1 and SC-2, using the method described in the previous paragraph, but not UV ozone cleaned and not etched. The results are shown in Fig. 3. The y axis in this figure represents the outgassing rate, in counts per second. As the temperature of the sample increases, shown in the x axis, molecules start to desorb. The temperature at which the outgassing occurs is important. Higher temperature desorption peaks represent higher activation energies. The OH had the highest outgassing rate, rising nearly monotonically from around 150°C to the maximum temperature of the desorption run, here around 550°C. The OH never actually formed a peak in this figure. None of the OH desorption measurements using the laser showed a true desorption peak. This caused some problems in assigning the water activation energy, which will be discussed later. The NH_3 also started to outgas at approximately 150°C, but unlike the OH, it exhibited a well formed peak. Integrating under the desorption curves yields a value that is proportional to the total desorbed species. The mean ratio of the total integrated OH to the NH_3 was approximately five.

In the next experiments, the samples were prepared using SC-1 and SC-2, then UV ozone cleaned for 20 minutes on each side, but not etched, then put immediately into the TDMS. UV ozone cleaning is

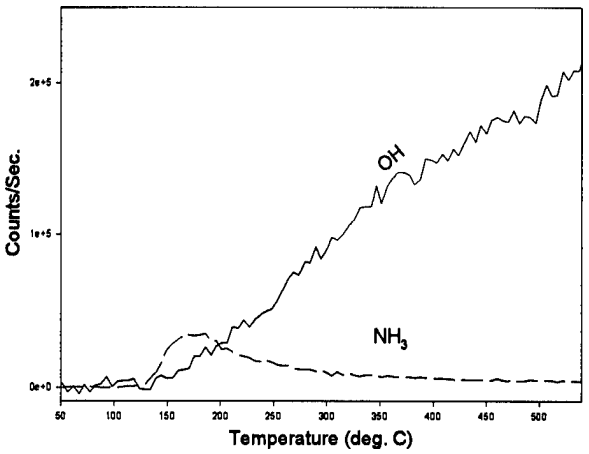


Figure 3: Outgassing of OH and NH_3 from unetched sample.

excellent for cleaning carbonaceous contamination from quartz, but it has not been recommended for the removal of inorganic contamination [7-9]. The samples prepared with UV ozone showed a drop in the ammonia of roughly a factor of six, while the OH showed no significant change. In Fig. 4 the ammonia desorption from a sample without UV ozone cleaning and one with UV ozone cleaning are plotted together so the differences between them can more easily be seen.

For most of the samples measured after UV ozone cleaning, a reduction of the low temperature

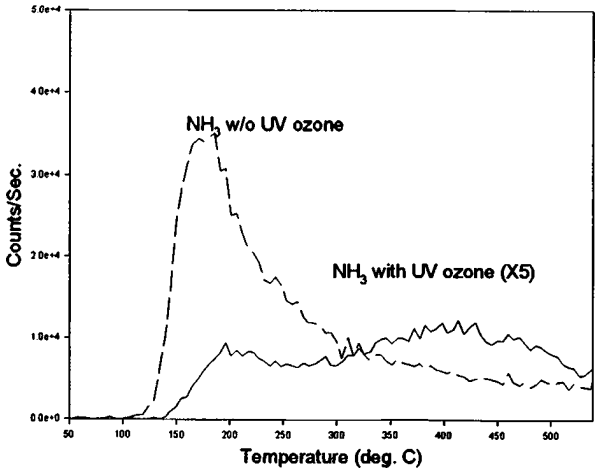


Figure 4: Outgassing of NH_3 from unetched sample cleaned with and without UV ozone.

(150° to 200°C) ammonia peaks was observed. However, a second higher temperature peak was promoted, usually around 390° to 450°C. The high temperature peaks were sometimes seen in other desorption measurements, but they were very prominent in the UV cleaned samples. In Ref. 6 it was reported that for some thick layers of contaminants, increasing the time samples spent in UV ozone cleaning from one hour to one day improved the cleanliness of the surface. So one sample was UV ozone cleaned overnight, but the amount of ammonia did not show any additional reduction.

The next experiment examined samples that were chemically cleaned using SC-1 and SC-2, UV ozone cleaned for 20 minutes on each side, and then etched in a saturated solution of ammonium bifluoride in water, at 75°C for one hour. This chemically polished the samples. The desorption is shown in Fig. 5. These samples desorbed approximately 7 times less OH than the unetched samples. While some decrease was expected due to the smoother surface of the sample after the chemical polish, the reduction of surface area is not sufficient to explain the results. For one thing, the amount of ammonia did not change

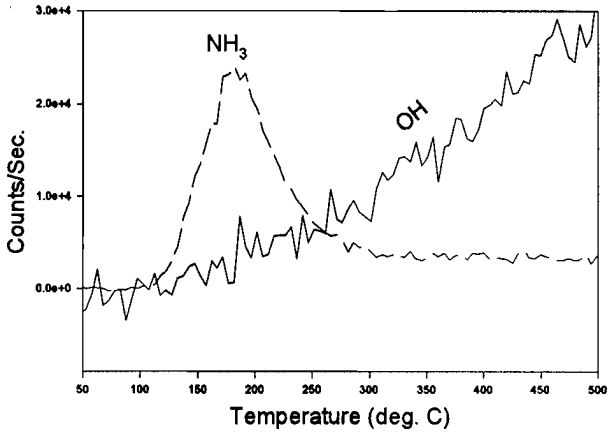


Figure 5: Outgassing of OH and NH_3 from etched sample.

significantly from the unetched results. Kiselev and Krylov [10], stated that fluorinating the surface of SiO_2 made the surface hydrophobic as well as acidic. They also reported that fluorination of the surface by 3% F, by mass, resulted in the reduction of OH groups on the surface by nearly a factor of 3. It is possible that part of the reduction in the desorption of OH in the etched samples is due to the presence of F, left on the surface from the ammonium bifluoride etching solution, while the rest is due to the smaller surface area of the etched surfaces. Also, Little [11] states that F on silica forms acid sites, and the ammonia could bond to these. This could be the reason why the amount of ammonia measured from the etched samples was not reduced compared to the unetched samples. This will be discussed in more detail later.

With the next samples, everything was processed as in the sample shown in Fig. 5, but after etching, and rinsing, the samples were UV ozone cleaned for 20 minutes on each side. This reduced the mean amount of ammonia with respect to the etched samples that had not been UV ozone cleaned after etching by approximately a factor of 10. The OH was unchanged and the low temperature ammonia peaks were depopulated while the higher temperature peaks were enhanced (like the unetched UV ozone cleaned desorptions). The ammonia desorbed from etched, UV-ozone-cleaned and etched, not-UV-ozone cleaned samples are shown in Fig. 6.

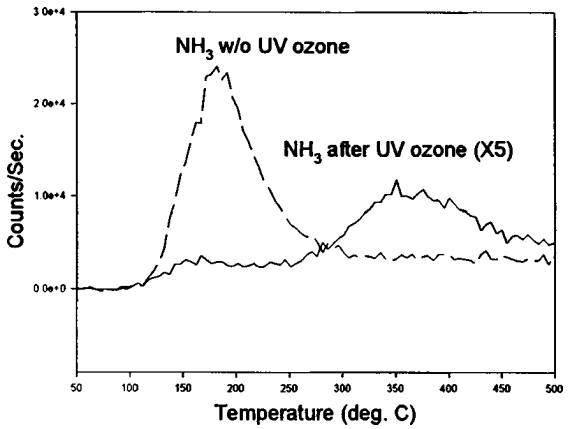


Figure 6: Outgassing of NH_3 from etched samples cleaned with and without UV ozone.

The OH, representing water, was expected because it had been reported in silica outgassing previously[10,11]. Ammonia was detected in both the etched and unetched samples, but both of them had been processed using some form of ammonia, so samples were lapped and cleaned without ammonia to try to isolate the point at which the ammonia got on the surface. Plotted in Fig. 7, is the desorption of ammonia from a sample that was prepared without ammonia cleaners or etches, just SC-2 and water rinses. The desorption of a sample prepared with ammonia (SC-1 and water rinses) is also shown. The ammonia was still very much present on the sample that was prepared without ammonia. It appeared that the ammonia was being adsorbed onto the quartz while in the vacuum system. It had been previously

assumed that all the ammonia in the desorption experiments had been from the ammonia in the processing. Apparently that was not the case.

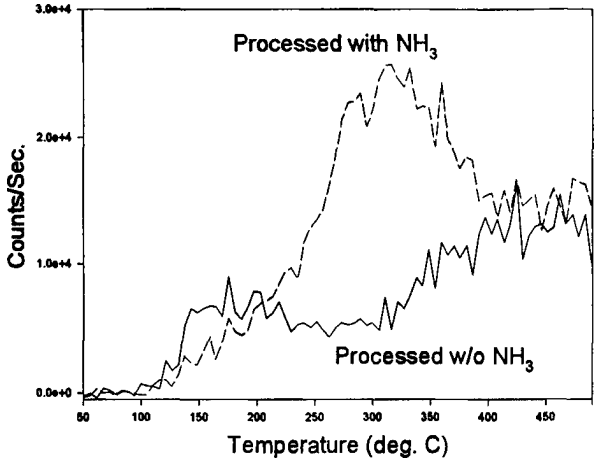


Figure 7: Outgassing of NH_3 from sample processed with and without NH_3.

When the background gas in the TDMS vacuum system was examined, a very small partial pressure of ammonia was found, approximately 10-50 times less than the OH background. To insure that the ammonia that was being detected was not some artifact of hydrothermally grown quartz, a piece of high purity silica was plasma etched and placed in the TDMS. When it was heated, it also gave off ammonia. The exact source of the ammonia remains undetermined, however, it could be something as simple as nitrogen and water reacting with the hot filaments in an ion gauge. Given that the TDMS had a very good vacuum system (the desorption runs were usually done in the mid to low 10^{-9} torr range) the small partial pressure of ammonia that is present here may be hard to eliminate. Also, since the mass of the ammonia lies so close to the OH, ammonia in the amounts that were seen in the TDMS vacuum system would almost certainly be undetectable using commercially available mass spectrometers.

Since the UV ozone affected the outgassing, an etched sample was exposed to a high-intensity Xe UV light source while it was inside the TDMS, at approximately 10^{-6} torr. The lamp produced 7 Watts between 200 and 400 nm and the sample was illuminated through a 1 mm thick sapphire window. The sample was exposed for four minutes on each side. The OH was reduced slightly and the ammonia was reduced by roughly a factor of 10 when the thermal contributions were factored out. This was roughly the same as UV ozone cleaning for etched samples. The desorption of the vacuum UV (VUV) cleaned sample is plotted with the desorption of an etched sample that did not use any type of UV cleaning after etching, in Fig. 8. Note that in this figure, both OH curves have been offset by $5x10^4$ to avoid being obscured by the ammonia curves.

Desorption curves that have completely resolved maxima, like ammonia, can be analyzed using the kinetics methods of Redhead [12] or Chan et al. [13], to obtain the activation energies. The ammonia peaks appear to obey simple, first order desorption kinetics. The low temperature activation energy was calculated to be at approximately 25 kcal/mole. This was a low enough activation energy so that it should desorb quickly at a relatively low temperature, even if the calculated activation energy were slightly in error. The higher temperature ammonia activation energy was calculated to be in the mid 40 kcal/mole range.

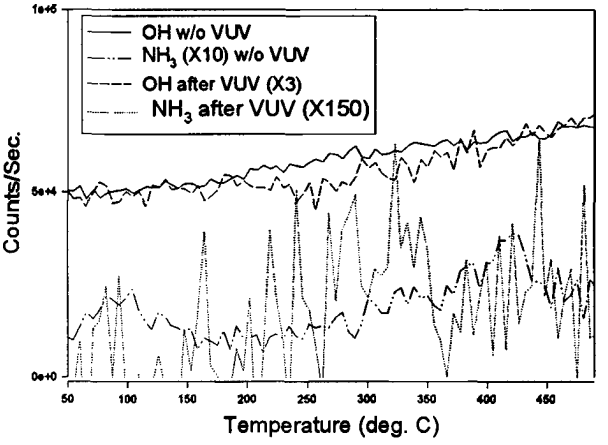


Figure 8: Outgassing of OH and NH3 from samples cleaned with and without VUV.

Estimating the activation energy, or energies, of the OH was undertaken as well. There were two problems that had to be resolved in order to do this. The first was the fact that in all the desorptions that were done with the laser, no desorption maxima were seen. That was because if the temperature exceeded the alpha-beta transition the samples cracked and fell from the sample holder. A maximum in the OH desorption was reported in last year's paper, when the sample was heated in the Ta cell [3]. The temperature of the peak was estimated by modeling the radiant heat transfer to the sample from the cell to be in the 450° to 650°C range. The Ta cell was not used in the current work because the low outgassing rates we were observing would be obscured by the Ta outgassing. The second problem was whether or not the OH desorption was due to one single

peak or to many peaks at different temperatures, or due to diffusion from the bulk. One way to test if the activation energy was correct was to heat the sample at a constant temperature for a given amount of time and observe if the surface species decrease to the amount that was predicted.

A bakeout experiment was prepared that would test the activation energy, or energies, of the OH and determine if the correct activation energy for the high temperature ammonia peak had been calculated. The sample was heated using the same linear temperature profile as all the other desorption experiments until it reached setpoint, 335°C then held constant for 13.5 minutes. When the sample was desorbed after the bakeout, the ammonia was reduced by a factor of 50 and the OH was reduced by a factor of 10. This gave an activation energy of the high temperature ammonia peak of 42 kcal/mole. As an example of how the OH outgassing data can be fit using multiple activation energies, in Fig. 9, a model of the OH outgassing, made using six activation energies, is plotted with the experimental OH outgassing, before the bakeout. The same OH model after a simulated bakeout is plotted with the experimental OH desorption in Fig. 10.

There are some indications in the literature that water may be mobile enough at elevated temperatures to diffuse through the bulk in quartz [14-16]. A test was conducted to determine how much, if any, of the OH that was being desorbed from the samples was coming from the bulk. The surface of one of the samples was etched with HF in heavy water. It was hoped this would label the surface desorption as D_2O, so any H_2O diffusing from the bulk, would be identifiable. However, it was not possible to use this method to separate surface water from bulk water because the H_2O in the vacuum exchanged with the D_2O. That is, when samples prepared this way were desorbed, the H_2O in the background gas reacted with D_2O on the surface to form HOD, and it was impossible to tell where it originated

The dehydroxylation of silica gels and quartz glass has been thoroughly investigated [17-22] (that is, heating the silica in a vacuum and driving the surface water away). At some temperature, starting at approximately 400°C for silica, the surface becomes irreversibly dehydrated, that is, hydrophobic. If this could be done with quartz resonators, it could reduce the adsorption and desorption of water after the enclosure was sealed, thus possibly reducing the aging. An experiment was designed to try to make an alpha quartz sample partially hydrophobic. A sample was desorbed 10 times without leaving the vacuum, using a linear 90 second temperature profile from 60° to 500° C, then put in deionized water for 5 minutes. When it was desorbed again, there was a slight reduction in OH, however the reduction was not enough to be significant based on this single experiment.

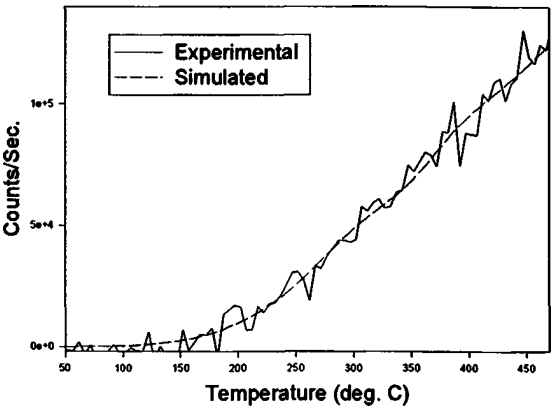


Figure 9: Experimental and simulated outgassing of OH before bakeout.

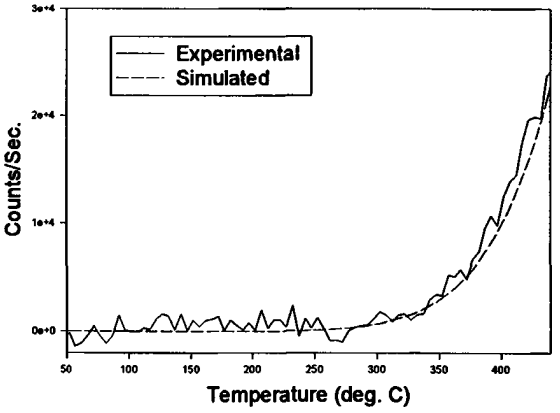


Figure 10: Experimental and simulated outgassing of OH after bakeout.

IV. DISCUSSION

The source of some of the ammonia adsorbed on the quartz surface does seem to be the vacuum system. The first question that will be discussed is: to what types of sites on the samples is the ammonia binding? The first possibility is that it is binding to some contaminant

atoms left on the surface after cleaning. This does not necessarily mean that the cleaning process used here left contaminants on the surface. It may be that the chemicals used in the cleaning process itself may leave behind atoms bound to the surface that the water rinses did not completely remove. For example, Little [11] stated that chlorine on the surface of silica can form an adsorption site for ammonia. This has relevance here because hydrochloric acid is used in one of the cleaning solutions (SC-2). Fluorine atoms may also affect the adsorption of ammonia. Kiselev and Krylow [10] have reported that fluorine on silica reduced the amount of OH on the surface, but that it also causes nearby water molecules to become proton donor centers. Proton donors are acid sites which can bond to basic molecules, such as ammonia. So the possibility exists that fluorine left on the surface from etching in ammonium bifluoride may increase the number of adsorption sites for ammonia in the etched samples enough to compensate for the reduction of surface area with respect to the unetched samples. This would explain why the amount of ammonia that was detected desobing from the etched and unetched samples was roughly constant, while the OH that desorbed from the etched samples was less than the OH from the unetched samples.

The next area to be addressed is what type of bonding sites are there for the ammonia atoms to bond with on a clean quartz surface? Morrow, et al. [17] report observing an ammonia adsorption peak on untreated silica (using IR spectroscopy) which they assigned to hydrogen bonded ammonia. Also, strained siloxane rings and bridges that can occur on the silica surface have been found to produce bonding sites that are very reactive. [17-22] In Fig. 11, the diagram shows how the ammonia can attach to the ring. Strained rings can be induced by applying a force or, in silica, by baking under vacuum at temperature above approximately 800°C. In alpha quartz, even if it could be caused by lapping damage, etching should remove it.

Next, the possible binding sites for OH on quartz is examined. The desorption of water on silica has been studied extensively. The generally accepted binding sites are shown in Fig 11, as described by Hair [23] and Little [11]. At room temperature, several types of bonding are occupied, including water physically adsorbed on OH sites. Heating to above 150°C removes most of the physisorbed water, but the rest desorb at higher temperatures. Pairs of OH, if they are close enough to each other, bond to form a siloxane bridge and release a water molecule, starting at about 400°C. Single OH are stable up to at least 800°C. So clearly, in silica, there are multiple activation energies that desorb at temperatures comparable to what were observed in this work for the alpha quartz.

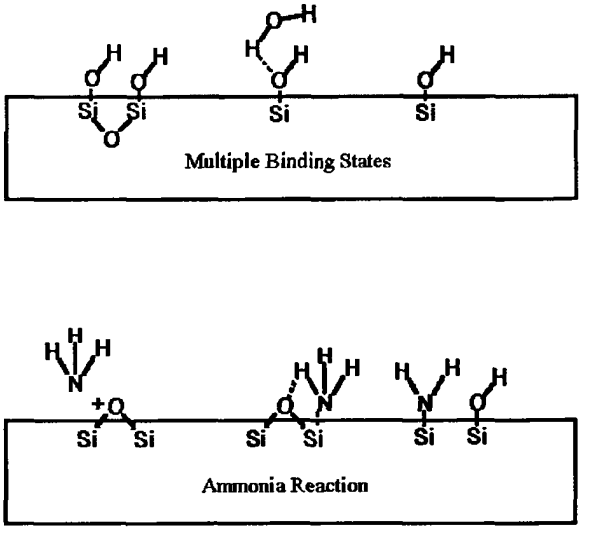


Figure 11: Bond sites.

Table 1 shows the bakeout times required at various bakeout temperatures to reduce an adsorbed monolayer to half a monolayer (half life) for different activation energies and the corresponding desorption peak temperatures, assuming first order kinetics, and a pre-exponential constant of 10^{13}/second. These were calculated using Eq. (1) from Redhead [12],

$$N(t) = -d\sigma / dt = \upsilon\sigma \exp(-E / RT) \qquad (1)$$

where N is the rate of desorption from the surface (molecules/cm^2 sec.), σ is the surface coverage

	Activation Energy(kcal/mole)/ Peak Desorption Temperature (°C)		
Bakeout Temp	28(kcal/ mole) 200°C	35(kcal/ mole) 300°C	41(kcal/ mole) 400°C
100°C	30 min	.7 years	10^3 years
200°C	1 sec.	.3 hr	7 days
300°C	3 msec.	2 sec.	5 min.

Table 1: Desorption half life.

(molecules/cm^2), ν is the pre-exponential rate constant.(1/sec), E is the activation energy (cal/mole), T is temperature, and R is the universal gas constant = 1.987 cal/mole K

The fit of the bakeout required that more than one activation energy be used. The pre-exponential constant used for the OH fit shown in Figs. 9 and 10 was 10^4/sec. which is within the range of experimentally reported values [13]. Six activation energies, ranging from 10 to 22 kcal/mole, were used to model the water desorption. These gave, using the $\nu = 10^4$ /sec. rate constant, peaks at temperatures from 200° to 620°C. Although the energies used reproduce the OH outgassing before and after the bakeout well, they are not the only possible values that could be fit. The example is included only to illustrate that multiple types of OH sites on the surface could represent the OH desorption.

In addition to bakeout times for achieving acceptable aging rates, the adsorption and desorption rates of surface contaminants have been shown to contribute significantly to phase noise in very high frequency (525 MHz) BAW resonators [4]. The phase noise was found to be dependent on both the surface coverage and desorption rates, among other things. Knowing the activation energies of surface species, could be important in the design and fabrication of microresonators, and microsensors.

V. CONCLUSIONS

It appears that the observed ammonia has multiple sources. Processing the wafers completely without ammonia still resulted in ammonia when the samples were heated in a vacuum. Some of the ammonia may be due to reactive species such as Cl, and maybe F, which can bind to the surface during processing and adsorb the ammonia while in the vacuum. Naturally occuring bond sites on the quartz surface, like hydrogen bonds for example, probably adsorb some of the ammonia from the vacuum system as well. However, if the quartz is stressed or damaged, other highly reactive states due to strained Si-O-Si bonds seen in silica may be possible. Ammonia was found in the TDMS vacuum system with partial pressures in the mid 10^{-11} torr range, 10-50 times smaller than the partial pressures of OH. The source of the ammonia in the system could be water and N_2 combining on hot filaments, or other similar reactions that are possible in a vacuum system that contains analytic components.

The desorptions were analyzed using the methods of Redhead[12] and Chan[13] and the ammonia peaks were found to be first order and approximately 25 and 42 kcal/mole, using the pre-exponential rate constant of 10^{13}/sec. The ammonia was reduced more than the water in the test bakeout. Given that the total amount of ammonia on the samples that were tested was always less than the water, any bakeout in a production line already set up to remove water from the resonator surface should also remove the ammonia. Calculating the activation energy of the water proved harder because no water maxima were observed. Without the maxima, the methods of Redhead and others are harder to apply. It was found that the water could not be adequately fit to the experimental data without using multiple desorption peaks.

Also, in the single trial that was conducted here, UV in vacuum (VUV) appeared to work about as well as UV in air (ozone). However, the UV lamp used in this work did not produce light with wavelengths shorter than around 200 nm. Lamps with wavelengths less than 180 nm [22] may be able to desorb the ammonia and water more efficiently.

VI. REFERENCES

[1] J.R. Vig and F.L. Walls, "Fundamental Limits of the Frequency Instabilities of Quartz Oscillators," Proc. of the 1994 IEEE Int. Freq. Cont. Symp., pp. 506-523.

[2] J.R. Vig and T.R. Meeker, "The Aging of Bulk Acoustic Wave Resonators, Filters, and Oscillators," Proc. of the 45th Ann. Symp. on Freq. Cont., pp. 77-101, 1991.

[3] D.E. Pierce, R.A. Murray, R. Lareau, S. Laffey, and J.R. Vig, "Outgassing of Quartz," Proc. of the 1994 IEEE Int. Freq. Cont. Symp., pp. 107-114.

[4] Y.K. Yong and J.R. Vig, "Resonator Surface Contamination - a Cause of Frequency Fluctuations?," Proc. of the 42th Ann. Symp. on Freq. Cont., pp. 397-403, 1988.

[5] E. Stenhagen, S. Abrahamsson, and F.W. McLafferty, Atlas of Mass Spectral Data, v. 1, John Wiley & Sons, NY, p.1, 1969.

[6] J.R. Vig, H. Wasshausen, C. Cook, M. Katz, and E. Hafner, "Surface Preparation and Characterization Techniques for Quartz Resonators," Proc. of the 27th Ann. Symp. on Freq. Cont., pp. 98-112, 1973.

[7] J.R. Vig, C. Cook, Jr., K. Schwidtal, LeBus, and E. Hafner, "Surface Studies for Quartz Resonators," Proc. of the 28th Ann. Symp. on Freq. Cont., pp. 96-108, 1974.

[8] J.R. Vig, J.W. LeBus, and R.L. Filler, "Further Results on UV Cleaning and Ni Electrobonding," Proc. of the 29th Ann. Symp. on Freq. Cont., pp. 220-229, 1975.

[9] J.R. Vig, "Ultraviolet-Ozone Cleaning of Semiconductor Surfaces," Handbook of Semiconductor Wafer Cleaning Technology, ed. W. Kern, Noyes Publications, pp. 233-273, 1993.

[10] V.F. Kiselev and O.V. Krylow, Adsorption Processes on Semiconductor and Dielectric Surfaces, V.1. Springer-Verlag, NY, 1985.

[11] L.H. Little, Infrared Spectra of Adsorbed Species, Academic Press, NY, pp. 246-249, 1966.

[12] P.A. Redhead, "Thermal Desorption of Gases," Vacuum, v.12, pp. 203-211, 1962.

[13] C.M. Chan, R. Aris, and W.H. Weinberg, "An Analysis of Thermal Desorption Mass Spectra I," Surface Science, v. 1, pp. 360-376, 1978.

[14] J.C. Brice, "Crystals for Quartz Resonators," Rev.Mod. Phys., V. 57, pp. 105-146, Jan. 1985.

[15] D. Griggs, Geophys. J.R. Astr. Soc., v. 14, pp. 19-31, 1967.

[16] A.C. McLaren, J.D. FitzGerald, and J. Gerretsen, Phys. Chem. Minerals, v. 16, pp. 465-482, 1989.

[17] B.A. Morrow, I.A. Cody, and Lydia S.M. Lee, "Infrared Studies on Oxide Surfaces 7. Mechanism of the Adsorption of Water and Ammonia on Dehydroxylated Silica, " J. Phys. Chem. v. 80, no. 23, pp. 2761-2767, 1976.

[18] B.C. Bunker, D.M. Haaland, K.J. Ward, T.A. Michalske, J.S. Binkley, C.F. Melius, C.A. Balfe "Infrared Spectra of Edge Shared Silicate Tetrahedra," Surface Science, v. 210, pp. 406-428, 1989.

[19] B.C. Bunker, D.M. Haaland, T.A. Michalske, and W.L. Smith, "Kinetics of Dissociative Chemisorption on Strained Edge Shared Surface Defects on Dehydroxylated Silica," Surface Science, v. 222, pp. 406-428, 1989.

[20] T.A Michalske and B.C. Bunker, "Slow Fracture on Strained Silicate Structures," J. Appl. Phys., v. 56, pp. 2686-2693, 1984.

[21] M.L. Hair, "The Molecular Nature of Adsorption on Silica Surfaces," Proc. of the 27th Ann. Symp. on Freq. Cont., pp. 73-78, 1973.

[22] M. L. White, "Clean Surface Technology," Proc. of the 27th Ann. Symp. on Freq. Cont., pp. 79-88, 1973.

[23] M.L. Hair, 'Hydroxyl Groups on Silica Surface," Non-Cryst. Sol., v 19, pp. 299-309, 1975.

[24] L.R. Allen, J.M. Grant, T. Nguyen, K.Valiev, L. Velikov, and B. Meshman, "Vacuum Ultraviolet Substrate Etching," Solid State Technology, pp. 77-80, May 1995.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PIEZOELECTRICALLY FORCED THICKNESS-SHEAR AND FLEXURAL VIBRATIONS OF CONTOURED QUARTZ RESONATORS

P. C. Y. Lee and Ji Wang
Department of Civil Engineering and Operations Research
Princeton University, Princeton, NJ 08544-5263

Abstract— **In a previous paper, a set of first-order equations of motion is obtained for contoured crystal plates and for frequencies upto and including those of the fundamental thickness-shear modes.**

In the present paper, the governing equations of contoured plates are extended to include the electric potential which is coupled to the mechanical fields by the piezoelectric effect. These equations are, then, employed for the study of piezoelectrically forced thickness-shear and flexural vibrations of beveled AT-cut quartz plate, *i.e.*, the plate with a portion of uniform thickness between the two wings of the double wedge.

Analytical solutions are obtained by Frobenius method. Displacements, stresses, and electric potential are derivable from six independent functions which are in the form of infinite power series. In addition to the calculations of resonance frequencies and mode shapes, the effect of the contouring on the forced mechanical displacements, electric potential, surface charge, and capacitance ratio are examined.

I. Introduction

Two-dimensional equations of motion of successively higher-order approximations were derived from the three-dimensional theory of piezoelectricity by expanding the mechanical displacement and electric potential in trigonometric series for piezoelectric crystal plates of uniform thickness with or without electrodes.[1] A system of first-order equations of motion for contoured crystal plates was similarly obtained, but an approximation of an integral of the product of stress component and a trigonometric function was introduced. These equations were further reduced to the coupled equations of thickness-shear and flexural vibrations by which the free vibrations of contoured quartz plates of AT-cut were studied.[2]

In the present paper, the governing equations of the thickness-shear and flexural vibrations are extended to the contoured piezoelectric crystal plates by including the coupling with the electric potential through the piezoelectric effect. These equations are employed to study the free and forced vibrations of beveled plates of AT-cut of quartz.

The effect of the slope of the beveled edge of the plate on the resonances, the distribution of mechanical displacements, the electric potential, the surface charge, and the capacitance ratio are studied. It is found that increasing the slope of the beveled edge weakens the coupling between the thickness-shear and flexural modes, lowers the frequencies of the flexural modes, but raises the frequencies of the fundamental thickness-shear and those of its anharmonic overtones. And increasing of the slope of the beveled edge decreases the capacitance ratios of both the flexural modes and the overtones of thickness-shear mode.

II. 2-D Equations for Contoured Crystal Plates

The plate is referred to rectangular coordinates x_i with the faces at $x_2 = \pm b(x_1, x_3)$ and with x_1 and x_3 the coordinates of the middle plane of the plate as shown in Fig. 1.

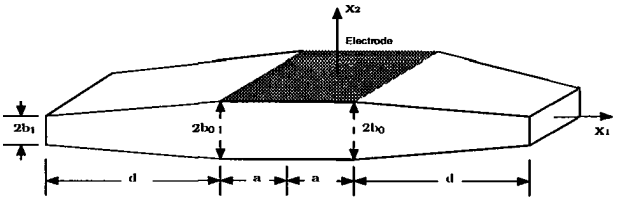


Fig. 1 A partially plated plate with beveled wedges.

We expand the components of the displacement and the electric potential in the trigonometrical series[1]

$$u_i = \sum_{n=0}^{\infty} u_i^{(n)}(x_1, x_3, t)\cos\frac{n\pi}{2}(1-\psi),$$
$$\phi_i = \sum_{n=0}^{\infty} \phi_i^{(n)}(x_1, x_3, t)\cos\frac{n\pi}{2}(1-\psi), \qquad (1)$$

where $\psi = x_2/b$ and $b = b(x_1, x_3)$ is one-half of the thickness of the plate.

By substituting (1) into the variational stress equations of motion, following similar method of deduction employed in Ref. 1, and approximating the integral[2] by

$$\frac{n\pi}{2}\int_{-1}^{1} T_{ij}\sin\frac{n\pi}{2}(1-\psi)d\psi \doteq \frac{1-\delta_{n0}}{2}\int_{-1}^{1} T_{ij}\cos\frac{n\pi}{2}(1-\psi)d\psi, \quad (2)$$

we have the nth-order stress equations of motion

$$T_{ij,i}^{(n)} + \frac{1-\delta_{n0}}{2}\frac{b_{,i}}{b}T_{ij}^{(n)} - \frac{n\pi}{2b}\bar{T}_{2j}^{(n)} + \frac{1}{b}F_j^{(n)} = \rho(1+\delta_{n0})\ddot{u}_j^{(n)},$$

0-7803-2500-1/95 $4.00 © 1995 IEEE

$$D^{(n)}_{i,i}+\frac{1-\delta_{n0}}{2}\frac{b_{,i}}{b}D^{(n)}_i-\frac{n\pi}{2b}\bar{D}^{(n)}_2+\frac{1}{b}D^{(n)}=0, \qquad (3)$$

the nth-order strain-displacement and electric field-potential relations

$$\begin{aligned}
S_{ij} &= \sum_{n=0}^{n}\left[S^{(n)}_{ij}\cos\frac{n\pi}{2}(1-\psi)+\bar{S}^{(n)}_{ij}\sin\frac{n\pi}{2}(1-\psi)\right],\\
E_k &= \sum_{n=0}^{n}\left[E^{(n)}_{k}\cos\frac{n\pi}{2}(1-\psi)+\bar{E}^{(n)}_{k}\sin\frac{n\pi}{2}(1-\psi)\right],\\
S^{(n)}_{ij} &= \frac{1}{2}\left[u^{(n)}_{j,i}+u^{(n)}_{i,j}\right]-\frac{1-\delta_{n0}}{4b}\left[b_{,i}u^{(n)}_j+b_{,j}u^{(n)}_i\right],\\
\bar{S}^{(n)}_{ij} &= \frac{n\pi}{4b}\left[\delta_{2i}u^{(n)}_j+\delta_{2j}u^{(n)}_i\right],\\
E^{(n)}_k &= -\left[\phi^{(n)}_{,k}-\frac{1-\delta_{n0}}{2b}b_{,k}\phi^{(n)}_{,k}\right],\\
\bar{E}^{(n)}_k &= -\frac{n\pi}{2b}\delta_{2k}\phi^{(n)}, \qquad (4)
\end{aligned}$$

and the nth-order constitutive relations

$$\begin{aligned}
T^{(n)}_p &= (1+\delta_{n0})\left[c_{pq}S^{(n)}_q-e_{kp}E^{(n)}_k\right]\\
&\quad+\sum_{m=0}^{\infty}A_{mn}\left[c_{pq}\bar{S}^{(m)}_q-e_{kp}\bar{E}^{(m)}_k\right],\\
\bar{T}^{(n)}_p &= (1-\delta_{n0})\left[c_{pq}\bar{S}^{(n)}_q-e_{kp}\bar{E}^{(n)}_k\right]\\
&\quad+\sum_{m=0}^{\infty}A_{nm}\left[c_{pq}S^{(m)}_q-e_{kp}E^{(m)}_k\right],\\
D^{(n)}_i &= (1+\delta_{n0})\left[e_{ip}S^{(n)}_p+\epsilon_{ij}E^{(n)}_j\right]\\
&\quad+\sum_{m=0}^{\infty}A_{mn}\left[e_{ip}\bar{S}^{(m)}_p+\epsilon_{ij}\bar{E}^{(m)}_j\right],\\
\bar{D}^{(n)}_i &= (1-\delta_{n0})\left[e_{ip}\bar{S}^{(n)}_p+\epsilon_{ij}\bar{E}^{(n)}_j\right]\\
&\quad+\sum_{m=0}^{\infty}A_{nm}\left[e_{ip}S^{(m)}_{ip}+\epsilon_{ij}E^{(m)}_j\right], \qquad (5)
\end{aligned}$$

where

$$\begin{aligned}
T^{(n)}_{ij} &= \int_{-1}^{1}T_{ij}\cos\frac{n\pi}{2}(1-\psi)d\psi,\\
\bar{T}^{(n)}_{ij} &= \int_{-1}^{1}T_{ij}\sin\frac{n\pi}{2}(1-\psi)d\psi, \quad \bar{T}^{(0)}_{ij}=0,\\
D^{(n)}_i &= \int_{-1}^{1}D_i\cos\frac{n\pi}{2}(1-\psi)d\psi,\\
\bar{D}^{(n)}_i &= \int_{-1}^{1}D_i\sin\frac{n\pi}{2}(1-\psi)d\psi, \quad \bar{D}^{(0)}_i=0,\\
F^{(n)}_j &= T_{2j}(1)-(-1)^nT_{2j}(-1),\\
D^{(n)} &= D_2(1)-(-1)^nD_2(-1),
\end{aligned}$$

$$\begin{aligned}
A_{mn} &= \int_{-1}^{1}\sin\frac{m\pi}{2}(1-\psi)\cos\frac{n\pi}{2}(1-\psi)d\psi\\
&= \begin{cases}\frac{4m}{(m^2-n^2)\pi} & m+n=\text{odd}\\ 0 & m+n=\text{even},\end{cases} \qquad (6)
\end{aligned}$$

and $p,q=1,2,3,4,5,6,i,j,k=1,2,3$,and δ_{n0} is the Kronecker delta.

III. Thickness-shear and Flexural Vibrations

A system of first-order equations are obtained from the infinite system, which consists of (3), (5), and (6), by Mindlin's truncation procedure, *i.e.*, letting

$$\begin{aligned}
u^{(n)}_1 &= u^{(n)}_3=0, \quad n>1,\\
u^{(n)}_2 &= 0, \quad n>2,\\
T^{(1)}_2 &= 0, \quad \ddot{u}^{(2)}_2=0, \quad \text{but} \quad \bar{S}^{(2)}_2=\frac{\pi}{b}u^{(2)}_2\neq 0,\\
T^{(n)}_{ij} &= \bar{T}^{(n)}_{ij}=S^{(n)}_{ij}=\bar{S}^{(n)}_{ij}=0 \quad \text{for} \quad n>1. \qquad (7)
\end{aligned}$$

By further removing the first-order thickness-stretch mode $u^{(1)}_2$, the x_2x_3 thickness-shear displacement $u^{(1)}_3$, and neglecting the x_2x_3 transverse shear deflection $u^{(0)}_{2,3}+\frac{\alpha_1}{b}u^{(1)}_3$ in the first-order equations in a manner similar to that in Ref. 2, we obtain a set of six coupled equations on $u^{(0)}_1$, $u^{(0)}_2$, $u^{(0)}_3$, $u^{(1)}_1$ and $\phi^{(0)}$ and $\phi^{(1)}$. Finally, by neglecting the coupling between $u^{(0)}_2$, $u^{(1)}_1$ and $u^{(0)}_1$, $u^{(0)}_3$ due to c_{56}, and also coupling between the $u^{(0)}_3$ and $\phi^{(1)}$ by e_{26}, we have the governing equations of the thickness-shear and flexural vibrations coupled with the first-order electric potential for contoured rotated Y-cut of quartz plates. For motions and contours dependent on x_1 only and faces free of tractions, the equations of motion and the charge equation become

$$\begin{aligned}
&T^{(0)}_{6,1}=2\rho\alpha_2^{-2}(1+R)\ddot{u}^{(0)}_2,\\
&T^{(1)}_{1,1}+\frac{b_{,1}}{2b}T^{(1)}_1-\frac{\alpha_1}{b}T^{(0)}_6=\rho(1+2R)\ddot{u}^{(1)}_1,\\
&D^{(1)}_{1,1}+\frac{b_{,1}}{2b}D^{(1)}_1-\frac{\alpha_1}{b}D^{(0)}_2+\frac{1}{b}D^{(1)}=0, \qquad (8)
\end{aligned}$$

and the constitutive relations

$$\begin{aligned}
\frac{1}{2}T^{(0)}_6 &= c_{66}(u^{(0)}_{2,1}+\frac{\alpha_1}{b}u^{(1)}_1)+\frac{\alpha_1}{b}e_{26}\phi^{(1)},\\
\frac{1}{2}D^{(0)}_2 &= e_{26}(u^{(0)}_{2,1}+\frac{\alpha_1}{b}u^{(1)}_1)-\frac{\alpha_1}{b}\epsilon_{22}\phi^{(1)},\\
T^{(1)}_1 &= \tilde{c}_{11}(u^{(1)}_{1,1}-\frac{b_{,1}}{2b}u^{(1)}_1)+\tilde{e}_{11}(\phi^{(1)}_{,1}-\frac{b_{,1}}{2b}\phi^{(1)}),\\
D^{(1)}_1 &= \tilde{e}_{11}(u^{(1)}_{1,1}-\frac{b_{,1}}{2b}u^{(1)}_1)-\tilde{\epsilon}_{11}(\phi^{(1)}_{,1}-\frac{b_{,1}}{2b}\phi^{(1)}),\\
D^{(1)} &= \frac{2A}{b}(e_{26}u^{(1)}_1-\epsilon_{22}\phi_0e^{i\omega t}), \qquad (9)
\end{aligned}$$

where

$$\tilde{c}_{11} = \hat{c}_{11}-\frac{\hat{c}_{13}\hat{c}_{31}}{\hat{c}_{33}}, \hat{c}_{pq}=\bar{c}_{pq}-\frac{\bar{c}_{p4}\bar{c}_{4q}}{\bar{c}_{44}}, \bar{c}_{pq}=c_{pq}-\frac{c_{p2}c_{2q}}{c_{22}},$$

$$\tilde{e}_{11} = \hat{e}_{11}-\frac{\hat{c}_{13}\hat{e}_{31}}{\hat{c}_{33}}, \hat{e}_{ip}=\bar{e}_{ip}-\frac{\bar{c}_{p4}\bar{e}_{4i}}{\bar{c}_{44}}, \bar{e}_{ip}=e_{ip}-\frac{c_{p2}e_{i2}}{c_{22}},$$

$$\tilde{\epsilon}_{11} = \hat{\epsilon}_{11}+\frac{\hat{e}_{13}\hat{e}_{13}}{\hat{c}_{33}}, \hat{\epsilon}_{ij}=\bar{\epsilon}_{ij}+\frac{\bar{e}_{j4}\bar{e}_{4i}}{\bar{c}_{44}}, \bar{\epsilon}_{ij}=\epsilon_{ij}+\frac{e_{j2}e_{i2}}{c_{22}},$$

$$\alpha_1 = \pi/\sqrt{8}, \quad \alpha_2 = 0.8225, \quad R=\frac{2b'\rho'}{b\rho}. \tag{10}$$

We note that $2b$ and $2b'$ denote the thicknesses, ρ and ρ' the mass densities of the plate and platings, respectively. Also the presence of the impressed alternative voltage on the electrodes or the absence of electrodes is accommodated in $D^{(1)}$ of $(9)_5$ by letting $A=1$ or 0.[1]

For plates with linearly varying thickness in the x_1 direction as shown in Fig. 1, the one half of the thickness of the plate b can be expressed by

$$b(x_1) = b_0(1 \mp \theta X), \tag{11}$$

where

$$\theta = \frac{b_0-b_1}{b_0} = 1-\frac{b_1}{b_0}, \quad X=\frac{x_1}{d}, \tag{12}$$

and the upper and lower signs in (11) apply to the cases where X is larger or smaller than zero. In the present paper, we consider $b_1 \neq 0$. Hence, $\theta < 1$.

For steady vibrations, we let

$$u_2^{(0)}=\eta(x_1)e^{i\omega t}, u_1^{(1)}=u(x_1)e^{i\omega t}, \phi^{(1)}=\sqrt{\frac{c_{66}}{\epsilon_{22}}}\phi(x_1)e^{i\omega t}, \tag{13}$$

where ω is the angular frequency.

By substituting (11) and (13) into (8) and eliminating of η from the coupled equations, we obtain the two coupled equations on u and ϕ

$$\begin{aligned}
&(1-\theta X)^4u^{\mathrm{IV}} - 2\theta(1-\theta X)^3u^{\mathrm{III}} \\
&\quad +(1-\theta X)^2\left[\frac{\theta^2}{4}+\alpha(1-\theta X)^2\right]u^{\mathrm{II}} \\
&\quad +\theta(1-\theta X)\left[\frac{\theta^2}{2}-\beta(1-\theta X)^2\right]u^{\mathrm{I}} \\
&\quad +\left[\frac{\theta^4}{2}+\gamma(1-\theta X)^2+\delta(1-\theta X)^4\right]u \\
&\quad +c_e\left\{(1-\theta X)^4\phi^{\mathrm{IV}} - 2\theta(1-\theta X)^3\phi^{\mathrm{III}}\right. \\
&\quad +(1-\theta X)^2\left[\frac{\theta^2}{4}+\alpha'(1-\theta X)^2\right]\phi^{\mathrm{II}} \\
&\quad \left.+\frac{\theta^3}{2}(1-\theta X)\phi^{\mathrm{I}}+\left[\frac{\theta^4}{2}+\gamma'(1-\theta X)^2\right]\phi\right\}=0, \\
&(1-\theta X)^2u^{\mathrm{II}}+\left[\frac{\theta^2}{4}+\delta'(1-\theta X)^2\right]u \\
&\quad +e_c\left[(1-\theta X)^2\phi^{\mathrm{II}}+\beta'\phi\right]=0,
\end{aligned} \tag{14}$$

where

$$\begin{aligned}
u^{\mathrm{N}} &= \frac{d^{\mathrm{N}}u}{dX^{\mathrm{N}}}, \quad \phi^{\mathrm{N}}=\frac{d^{\mathrm{N}}\phi}{dX^{\mathrm{N}}}, \quad \mathrm{N}=\mathrm{I},\mathrm{II},\mathrm{III},\mathrm{IV} \\
\alpha &= \left(\frac{1+2R}{\check{c}_{11}}+\frac{1+R}{\alpha_2^2}\right)\bar{\Omega}^2, \beta=2(1+2R)\frac{\bar{\Omega}^2}{\check{c}_{11}}, \\
\gamma &= \frac{1+R}{\alpha_2^2}\left[\left(\frac{\theta}{2}\right)^2-\frac{1}{\check{c}_{11}}\left(\frac{\pi d}{2b_0}\right)^2\right]\bar{\Omega}^2, \\
\delta &= \frac{(1+R)(1+2R)}{\alpha_2^2}\frac{\bar{\Omega}^4}{\check{c}_{11}}, \\
\alpha' &= \frac{1+R}{\alpha_2^2}\bar{\Omega}^2, \quad \beta'=\frac{\theta^2}{4}-\left(\frac{\pi}{2}\frac{d}{b_0}\right)^2\frac{q_{12}}{q_{11}}, \\
\gamma' &= \frac{1+R}{\alpha_2^2}\left[\left(\frac{\theta}{2}\right)^2-\frac{\check{e}_{26}}{\check{e}_{11}}\left(\frac{\pi d}{2b_0}\right)^2\right]\bar{\Omega}^2, \\
\delta' &= \frac{1+2R}{p_{11}}\bar{\Omega}^4, c_e=\frac{\check{e}_{11}}{\check{c}_{11}}, e_c=\frac{q_{11}}{p_{11}}, \\
p_{11} &= \check{c}_{11}-\frac{\check{e}_{11}}{\check{e}_{26}}, q_{11}=\check{e}_{11}+\frac{\check{\epsilon}_{11}}{\check{e}_{26}}, q_{12}=\check{e}_{26}+\frac{\check{\epsilon}_{22}}{\check{e}_{26}}, \\
\check{c}_{pq} &= \frac{\tilde{c}_{pq}}{c_{66}}, \quad \check{e}_{ip}=\frac{e_{ip}}{\sqrt{c_{66}\epsilon_{22}}}, \quad \check{\epsilon}_{ij}=\frac{\epsilon_{ij}}{\epsilon_{22}}, \\
\Omega &= \frac{\omega}{\omega_0}, \quad \omega_0=\frac{\pi}{2b_0}\sqrt{\frac{c_{66}}{\rho}}, \quad \bar{\Omega}=\frac{\pi d}{2b_0}\Omega.
\end{aligned} \tag{15}$$

Once u and ϕ are solved from (14), the transverse displacement η, stresses, and electric displacements can be obtained from following relations:

$$\begin{aligned}
\eta(X) &= -\frac{\alpha_2^2}{2(1+R)\alpha_1}\frac{b_0}{d}\frac{1}{\bar{\Omega}^2}\left\{\hat{X}(\check{c}_{11}u^{\mathrm{III}}+\check{e}_{11}\phi^{\mathrm{III}})\right. \\
&\quad -\theta(\check{c}_{11}u^{\mathrm{II}}+\check{e}_{11}\phi^{\mathrm{II}})+\left[\frac{\check{c}_{11}\theta^2}{4\hat{X}}+(1+2R)\bar{\Omega}^2\hat{X}\right]u^{\mathrm{I}} \\
&\quad +\frac{\check{e}_{11}\theta^2}{4\hat{X}}\phi^{\mathrm{I}}+\theta\left[\check{c}_{11}\left(\frac{\theta}{2\hat{X}}\right)^2-(1+2R)\bar{\Omega}^2\right]u \\
&\quad \left.+\check{e}_{11}\theta\left(\frac{\theta}{2\hat{X}}\right)^2\phi\right\}, \\
T_6^{(0)} &= \frac{\hat{X}}{\alpha_1}\frac{b_0}{d^2}c_{66}\left\{\check{c}_{11}u^{\mathrm{II}}+\check{e}_{11}\phi^{\mathrm{II}}+\left[\check{c}_{11}\left(\frac{\theta}{2\hat{X}}\right)^2\right.\right. \\
&\quad \left.\left.+(1+2R)\bar{\Omega}^2\right]u+\check{e}_{11}\left(\frac{\theta}{2\hat{X}}\right)^2\phi\right\}, \\
D_2^{(0)} &= \frac{\hat{X}}{\alpha_1}\frac{b_0}{d^2}\sqrt{\epsilon_{22}c_{66}}\left[\check{e}_{11}u^{\mathrm{II}}-\check{\epsilon}_{11}\phi^{\mathrm{II}}+\left(\frac{\theta}{2\hat{X}}\right)^2(\check{e}_{11}u\right. \\
&\quad \left.-\check{\epsilon}_{11}\phi)+\frac{2A}{\hat{X}^2}\left(\frac{d}{b_0}\right)^2\left(\check{e}_{26}u-\sqrt{\frac{\epsilon_{22}}{c_{66}}}\phi_0e^{i\omega t}\right)\right], \\
T_1^{(1)} &= \frac{c_{66}}{d}\left[\check{c}_{11}u^{\mathrm{I}}+\check{e}_{11}\phi^{\mathrm{I}}+\frac{\theta}{2\hat{X}}(\check{c}_{11}u+\check{e}_{11}\phi)\right],
\end{aligned}$$

$$D_1^{(1)} = \frac{\sqrt{\epsilon_{22}c_{66}}}{d}\left[\check{e}_{11}u^{\mathrm{I}} - \check{\epsilon}_{11}\phi^{\mathrm{I}} + \frac{\theta}{2\hat{X}}(\check{e}_{11}u - \check{\epsilon}_{11}\phi)\right], \quad (16)$$

where

$$\hat{X} = 1 - \theta X. \quad (17)$$

IV. Solutions for the Contoured Portion of the Plate

Solutions of (14) in terms of power series of X are obtained by Frobenius method by letting[2]

$$\begin{aligned} u(X) &= X^s \sum_{n=0}^{\infty} a_n X^n, \quad a_0 \neq 0, \\ \phi(X) &= X^r \sum_{n=0}^{\infty} b_n X^n, \quad b_0 \neq 0. \end{aligned} \quad (18)$$

Substitution of (18) into (14) yields

$$\begin{aligned} &\sum_{n=0}^{\infty}\left[f_i(n,s)a_nX^{n+s+i} + c_e g_i(n,r)b_nX^{n+r+i}\right] = 0, \\ &\sum_{n=0}^{\infty}\left[\bar{f}_j(n,s)a_nX^{n+s+j} + e_c \bar{g}_j(n,r)b_nX^{n+r+j}\right] = 0, \\ &i = -4,-3,\ldots,3,4,\ j = -2,-1,0,1,2, \end{aligned} \quad (19)$$

where

$$\begin{aligned} f_{-4}(n,s) &= (n+s)(n+s-1)(n+s-2)(n+s-3), \\ g_{-4}(n,r) &= (n+r)(n+r-1)(n+r-2)(n+r-3), \\ f_{-3}(n,s) &= -2\theta(n+s)(n+s-1)(n+s-2) \\ &\quad \times[2(n+s)-5], \\ g_{-3}(n,r) &= -2\theta(n+r)(n+r-1)(n+r-2) \\ &\quad \times[2(n+r)-5], \\ f_{-2}(n,s) &= \left\{\theta^2\left[\frac{1}{4} + 6(n+s-2)^2\right] + \alpha\right\} \\ &\quad \times(n+s)(n+s-1), \\ g_{-2}(n,r) &= \left\{\theta^2\left[\frac{1}{4} + 6(n+r-2)^2\right] + \alpha'\right\} \\ &\quad \times(n+r)(n+r-1), \\ f_{-1}(n,s) &= -\theta(n+s)\left\{2(n+s-1)\left[(n+s-2)\right.\right. \\ &\quad \left.\left.\times(2n+2s-3)\theta^2 + \frac{\theta^2}{4} + 2\alpha\right] - \left(\frac{\theta^2}{2} - \beta\right)\right\}, \\ g_{-1}(n,r) &= -\theta(n+s)\left\{2(n+r-1)\left[(n+r-2)\right.\right. \\ &\quad \left.\left.\times(2n+2r-3)\theta^2 + \frac{\theta^2}{4} + 2\alpha'\right] - \frac{\theta^2}{2}\right\}, \\ f_0(n,s) &= \left\{\theta^2(n+s)(n+s-1)\left[\theta^2(n+s-1)(n+s-2)\right.\right. \\ &\quad \left.\left.+6\alpha + \frac{\theta^2}{4}\right] + \theta^2(3\beta - \frac{\theta^2}{2})(n+s) + \frac{\theta^4}{2} + \gamma + \delta\right\}, \\ g_0(n,r) &= \left\{\theta^2(n+r)(n+r-1)\left[\theta^2(n+r-1)(n+r-2)\right.\right. \\ &\quad \left.\left.+6\alpha' + \frac{\theta^2}{4}\right] - \frac{\theta^4}{2}(n+r) + \frac{\theta^4}{2} + \gamma'\right\}, \\ f_1(n,s) &= -\theta\left\{\theta^2(n+s)[4\alpha(n+s-1)+3\beta] + 2(\gamma+2\delta)\right\}, \\ g_1(n,r) &= -2\theta\left[2\alpha'\theta^2(n+r)(n+r-1) + \gamma'\right], \\ f_2(n,s) &= \theta^2\left\{\theta^2(n+s)[\alpha(n+s-1)+\beta] + \gamma + 6\delta\right\}, \\ g_2(n,r) &= \theta^2\left[\alpha'\theta^2(n+r)(n+r-1) + \gamma'\right], \\ f_3(n,s) &= -4\delta\theta^3, g_3(n,r) = 0, \\ f_4(n,s) &= \delta\theta^4, g_4(n,r) = 0, \\ \bar{f}_{-2}(n,s) &= (n+s)(n+s-1), \\ \bar{g}_{-2}(n,r) &= (n+r)(n+r-1), \\ \bar{f}_{-1}(n,s) &= -2\theta(n+s)(n+s-1), \\ \bar{g}_{-1}(n,r) &= -2\theta(n+r)(n+r-1), \\ \bar{f}_0(n,s) &= \theta^2\left[(n+s)(n+s-1) + \frac{1}{4}\right] + \delta', \\ \bar{g}_0(n,r) &= \theta^2(n+r)(n+r-1) + \beta', \\ \bar{f}_1(n,s) &= -2\theta\delta', \bar{g}_1(n,r) = 0, \\ \bar{f}_2(n,s) &= \theta^2\delta', \bar{g}_2(n,r) = 0. \end{aligned} \quad (20)$$

By setting the coefficients of the terms with the lowest power of X in $(20)_1$ and $(20)_2$ equal to zero, *i.e.* $f_{-4}(n,s) = 0$, $g_{-4}(n,r) = 0$, $\bar{f}_{-2}(n,s) = 0$, and $\bar{g}_{-2}(n,r) = 0$, we have the indicial equations

$$\begin{aligned} &s(s-1)(s-2)(s-3) = 0, \quad \text{or} \quad s = 0,1,2,3, \\ &r(r-1)(r-2)(r-3) = 0, \quad \text{or} \quad r = 0,1,2,3, \\ &s(s-1) = 0, \quad \text{or} \quad s = 0,1, \\ &r(r-1) = 0, \quad \text{or} \quad r = 0,1. \end{aligned} \quad (21)$$

It may be seen from (20) that the choice of $s = 0$ and $r = 0$ makes the coefficients of the terms X^{-4}, X^{-3}, X^{-2}, and X^{-1} to vanish for arbitrary values of a_0, a_1, a_2, and a_3 and b_0 and b_1. Hence, the remaining constants in (19) can be expressed in terms of these six arbitrary constants by the coupled nine-term and five-term recurrence relations

$$\begin{aligned} a_{n+8} + c_e b_{n+8} &= \sum_{j=0}^{7}(\Lambda_{n+j}a_{n+j} + c_e\Gamma_{n+j}b_{n+j}), \\ a_{n+8} + e_c b_{n+8} &= \sum_{j=4}^{7}(\bar{\Lambda}_{n+j}a_{n+j} + e_c\bar{\Gamma}_{n+j}b_{n+j}), \end{aligned} \quad (22)$$

where $n > -5$ and

$$\begin{aligned} \Lambda_{n+7} &= 2\theta(2n+9)/(n+8), \\ \Gamma_{n+7} &= 2\theta(2n+9)/(n+8), \\ \Lambda_{n+6} &= -\{\theta^2[6(n+4)^2 + 1/4] + \alpha\}/[(n+8)(n+7)], \end{aligned}$$

$$
\begin{aligned}
\Gamma_{n+6} &= -\{\theta^2[6(n+4)^2+1/4]+\alpha'\}/[(n+8)(n+7)],\\
\Lambda_{n+5} &= \theta\{2(n+4)[(2n+7)(n+3)\theta^2+\theta^2/4+2\alpha]\\
&\quad -(\theta^2/2-\beta)\}/[(n+8)(n+7)(n+6)],\\
\Gamma_{n+5} &= \theta\{2(n+4)[(2n+7)(n+3)\theta^2+\theta^2/4+2\alpha']\\
&\quad -\theta^2/2\}/[(n+8)(n+7)(n+6)],\\
\Lambda_{n+4} &= -\{(n+4)(n+3)\theta^2[(n+3)(n+2)\theta^2+6\alpha+\theta^2/4]\\
&\quad -(\theta^2/2-3\beta)\theta^2(n+4)+\theta^4/2+\gamma+\delta\}/N(8,5),\\
\Gamma_{n+4} &= -\{(n+4)(n+3)\theta^2[(n+3)(n+2)\theta^2+6\alpha'+\theta^2/4]\\
&\quad -\theta^4(n+4)/2+\theta^4/2+\gamma'\}/N(8,5),\\
\Lambda_{n+3} &= \theta[4\alpha\theta^2(n+3)(n+2)+3\beta\theta^2(n+3)\\
&\quad +2(\gamma+2\delta)]/N(8,5),\\
\Gamma_{n+3} &= 2\theta[2\alpha'\theta^2(n+3)(n+2)+\gamma']/N(8,5),\\
\Lambda_{n+2} &= -\theta^2[\alpha\theta^2(n+2)(n+1)+\beta\theta^2(n+2)\\
&\quad +(\gamma+6\delta)]/N(8,5),\\
\Gamma_{n+2} &= -\theta^2[\alpha'\theta^2(n+2)(n+1)+\gamma']/N(8,5),\\
\Lambda_{n+1} &= 4\gamma\theta^3/N(8,5), \Gamma_{n+1}=0,\\
\Lambda_n &= -\delta\theta^4/N(8,5), \Gamma_n=0,\\
\bar{\Lambda}_{n+7} &= 2\theta(n+6)/(n+8), \bar{\Gamma}_{n+7}=2\theta(n+6)/(n+8),\\
\bar{\Lambda}_{n+6} &= -\{\theta^2[(n+6)(n+5)+1/4]+\delta'\}/[(n+8)(n+7)],\\
\bar{\Gamma}_{n+6} &= -[\theta^2(n+6)(n+5)+\beta']/[(n+8)(n+7)],\\
\bar{\Lambda}_{n+5} &= 2\theta\delta'/[(n+8)(n+7)], \bar{\Gamma}_{n+5}=0,\\
\bar{\Lambda}_{n+4} &= -\theta^2\delta'/[(n+8)(n+7)], \bar{\Gamma}_{n+4}=0,\\
N(8,5) &= (n+8)(n+7)(n+6)(n+5). \qquad (23)
\end{aligned}
$$

By replacing the six arbitrary constants by $A_i (i = 0,1,2,3,4,5)$

$$
\begin{aligned}
a_0 &= A_0+A_2, \quad a_2 = A_0-A_2,\\
a_1 &= A_1+A_3, \quad a_3 = A_1-A_3,\\
b_0 &= A_4, \quad b_1 = A_5, \qquad (24)
\end{aligned}
$$

and inserting (23) and (24) into the righthand side of (22), we have

$$
\begin{aligned}
a_n &= A_0U_0(n)+A_1U_1(n)+A_2U_2(n)\\
&\quad +A_3U_3(n)+A_4U_4(n)+A_5U_5(n),\\
b_m &= A_0V_0(m)+A_1V_1(m)+A_2V_2(m)\\
&\quad +A_3V_3(m)+A_4V_4(m)+A_5V_5(m), \qquad (25)
\end{aligned}
$$

where $n > 3$ and $m > 1$.

For instance, the coefficients of b_3 are $V_i(3), i$=0, 1, 2, 3, 4, 5, as given below

$$
\begin{aligned}
V_0(3) &= -\frac{\theta^3}{12e_c}, \; V_1(3) = -\frac{1}{e_c}\left[\frac{1}{6}\left(\frac{\theta^2}{4}+\delta'\right)+1\right],\\
V_2(3) &= -\frac{\theta^3}{12e_c}, \; V_3(3) = -\frac{1}{e_c}\left[\frac{1}{6}\left(\frac{\theta^2}{4}+\delta'\right)-1\right],\\
V_4(3) &= -\frac{\theta\beta'}{3}, \quad V_5(3) = -\frac{\beta'}{6}. \qquad (26)
\end{aligned}
$$

Substitution of (24) and (25) into (18) yields

$$
\begin{aligned}
u(X) &= A_0W_0(X)+A_1W_1(X)+A_2W_2(X)\\
&\quad +A_3W_3(X)+A_4W_4(X)+A_5W_5(X),\\
\phi(X) &= A_0L_0(X)+A_1L_1(X)+A_2L_2(X)\\
&\quad +A_3L_3(X)+A_4L_4(X)+A_5L_5(X), \qquad (27)
\end{aligned}
$$

where

$$
\begin{aligned}
W_i(X) &= \sum_{n=0}^{\infty} U_i(n)X^n,\\
L_i(X) &= \sum_{n=0}^{\infty} V_i(n)X^n, \quad i = 0,1,2,3,4,5. \qquad (28)
\end{aligned}
$$

And further substitution of (28) into (16) gives

$$
\begin{aligned}
\eta(X) &= A_0S_0(X)+A_1S_1(X)+A_2S_2(X)\\
&\quad +A_3S_3(X)+A_4S_4(X)+A_5S_5(X),\\
T_6^{(0)} &= A_0Q_0(X)+A_1Q_1(X)+A_2Q_2(X)\\
&\quad +A_3Q_3(X)+A_4Q_4(X)+A_5Q_5(X),\\
D_2^{(0)} &= A_0H_0(X)+A_1H_1(X)+A_2H_2(X)\\
&\quad +A_3H_3(X)+A_4H_4(X)+A_5H_5(X),\\
T_1^{(1)} &= A_0M_0(X)+A_1M_1(X)+A_2M_2(X)\\
&\quad +A_3M_3(X)+A_4M_4(X)+A_5M_5(X),\\
D_1^{(1)} &= A_0P_0(X)+A_1P_1(X)+A_2P_2(X)\\
&\quad +A_3P_3(X)+A_4P_4(X)+A_5P_5(X), \qquad (29)
\end{aligned}
$$

where

$$
\begin{aligned}
S_i(X) &= -\frac{\alpha_2^2}{2(1+R)\alpha_1}\frac{b_0}{d}\frac{1}{\bar{\Omega}^2}\Big\{\hat{X}(\check{c}_{11}W_i^{\mathrm{III}}+\check{e}_{11}L_i^{\mathrm{III}})\\
&\quad -\theta(\check{c}_{11}W_i^{\mathrm{II}}+\check{e}_{11}L_i^{\mathrm{II}})+\left[\frac{\check{c}_{11}\theta^2}{4\hat{X}}+(1+2R)\bar{\Omega}^2\hat{X}\right]W_i^{\mathrm{I}}\\
&\quad +\frac{\check{e}_{11}\theta^2}{4\hat{X}}L_i^{\mathrm{I}}-\theta\left[\check{c}_{11}\left(\frac{\theta}{2\hat{X}}\right)^2-(1+2R)\bar{\Omega}^2\right]W_i\\
&\quad +\check{e}_{11}\theta\left(\frac{\theta}{2\hat{X}}\right)^2L_i\Big\},\\
H_i(X) &= \frac{\hat{X}}{\alpha_1}\frac{b_0}{d^2}\sqrt{\epsilon_{22}c_{66}}\left[\check{e}_{11}W_i^{\mathrm{II}}-\check{\epsilon}_{11}L_i^{\mathrm{II}}\right.\\
&\quad +\left(\frac{\theta}{2\hat{X}}\right)^2(\check{e}_{11}W_i-\check{\epsilon}_{11}L_i)\\
&\quad \left.+\frac{2A}{\hat{X}^2}\left(\frac{d}{b_0}\right)^2\left(\check{e}_{26}L_i-\sqrt{\frac{\epsilon_{22}}{c_{66}}}\,\phi_0e^{i\omega t}\right)\right],\\
Q_i(X) &= \frac{\hat{X}}{\alpha_1}\frac{b_0}{d^2}c_{66}\left\{\check{c}_{11}W_i^{\mathrm{II}}+\check{e}_{11}L_i^{\mathrm{II}}+\left[\check{c}_{11}\left(\frac{\theta}{2\hat{X}}\right)^2\right.\right.
\end{aligned}
$$

$$+(1+2R)\bar{\Omega}^2]\,W_i+\check{e}_{11}\left(\frac{\theta}{2\hat{X}}\right)^2 L_i\Bigg\},$$

$$M_i(X)=\frac{c_{66}}{d}\left[\check{c}_{11}W_i^{\mathrm{I}}+\check{e}_{11}L_i^{\mathrm{I}}+\frac{\theta}{2\hat{X}}(\check{c}_{11}W_i+\check{e}_{11}L_i)\right],$$

$$P_i(X)=\frac{\sqrt{\epsilon_{22}c_{66}}}{d}\left[\check{e}_{11}W_i^{\mathrm{I}}-\check{\epsilon}_{11}L_i^{\mathrm{I}}+\frac{\theta}{2\hat{X}}(\check{e}_{11}W_i-\check{\epsilon}_{11}L_i)\right],$$

$$i=0,1,2,3,4,5. \tag{30}$$

We note that all the results given in this section upto this point are for $X\geq 0$. For $X\leq 0$, θ must be replaced by $-\theta$ in (11). Hence, in the functions $U_i(n)$ and $V_i(n)$ in (28), θ must be replaced by $-\theta$ for $X\leq 0$. A close examination of $U_i(n)$ and $V_i(n)$ leads us to state, as it may be seen in (26), that with respect to θ

$$\begin{aligned}&U_i(n,\theta)\text{ and }V_i(n,\theta)\text{ are symmetric,}\\ &\text{if }\ i+n=\text{even; anti-symmetric, if }\ i+n=\text{odd.}\end{aligned} \tag{31}$$

Therefore, we see from (29) that W_0, W_2, and W_4 are symmetric and W_1, W_3, and W_5 are anti-symmetric with respect to X. Functions $L_i(X)$ have the same symmetry properties.

All these functions will be needed in order to satisfy the continuity and edge conditions of the beveled plate of Fig. 1.

V. Solutions for the Uniform Portion of the Plate

For the uniform portion of the plate, *i.e.*, $|x_1|\leq a$, $b(x_1)=b_0$, $b_{,1}=0$, and the governing equations (8) and (9) reduce to

$$\begin{aligned}&T_{6,1}^{(0)}=2\rho\alpha_2^{-2}(1+R)\ddot{u}_2^{(0)},\\ &T_{1,1}^{(1)}-\frac{\alpha_1}{b_0}T_6^{(0)}=\rho(1+2R)\ddot{u}_1^{(1)},\\ &D_{1,1}^{(1)}-\frac{\alpha_1}{b_0}D_2^{(0)}+\frac{1}{b_0}D^{(1)}=0,\end{aligned} \tag{32}$$

and

$$\begin{aligned}\frac{1}{2}T_6^{(0)}&=c_{66}(u_{2,1}^{(0)}+\frac{\alpha_1}{b_0}u_1^{(1)})+\frac{\alpha_1}{b_0}e_{26}\phi^{(1)},\\ \frac{1}{2}D_2^{(0)}&=e_{26}(u_{2,1}^{(0)}+\frac{\alpha_1}{b_0}u_1^{(1)})-\frac{\alpha_1}{b_0}\epsilon_{22}\phi^{(1)},\\ T_1^{(1)}&=\tilde{c}_{11}u_{1,1}^{(1)}+\tilde{e}_{11}\phi_{,1}^{(1)},\\ D_1^{(1)}&=\tilde{e}_{11}u_{1,1}^{(1)}-\tilde{\epsilon}_{11}\phi_{,1}^{(1)},\\ D^{(1)}&=\frac{2A}{b_0}(e_{26}u_1^{(1)}-\epsilon_{22}\phi_0e^{i\omega t}).\end{aligned} \tag{33}$$

For symmetric thickness-shear vibrations and with electrodes shorted, *i.e.* $\phi_0=0$, we let

$$\begin{aligned}\bar{u}_2^{(0)}&=C_0\sin\zeta x e^{i\omega t},\\ \bar{u}_1^{(1)}&=C_1\cos\zeta x e^{i\omega t},\\ \bar{\phi}^{(1)}&=\sqrt{\frac{c_{66}}{\epsilon_{22}}}\,C_2\cos\zeta x e^{i\omega t},\end{aligned} \tag{34}$$

where ζ is the wavenumber in the x_1 direction.

Substitution of (34) into (32) results in

$$\begin{pmatrix} Z^2-(1+R)\left(\frac{\Omega}{\alpha_2}\right)^2 & \frac{1}{\sqrt{2}}Z & \frac{\check{e}_{26}}{\sqrt{2}}Z\\ \sqrt{2}Z & 1+\check{\epsilon}_{11}Z^2-(1+2R)\Omega^2 & (\check{e}_{26}+\check{e}_{11}Z^2)\\ \sqrt{2}\check{e}_{26}Z & \check{e}_{11}Z^2+\check{e}_{26}(1-\frac{\beta A}{\pi^2}) & -(\check{\epsilon}_{11}Z^2+1)\end{pmatrix}\begin{pmatrix}C_0\\ C_1\\ C_2\end{pmatrix}=0, \tag{35}$$

where

$$\Omega=\frac{\omega}{\frac{\pi}{2b_0}\sqrt{\frac{c_{66}}{\rho}}},\ Z=\frac{\zeta}{\frac{\pi}{2b_0}}. \tag{36}$$

For nontrivial amplitudes $C_r(r=0,1,2)$, the determinant of the coefficient matrix of (35) must vanish, *i.e.*

$$\begin{vmatrix} Z^2-(1+R)\left(\frac{\Omega}{\alpha_2}\right)^2 & \frac{1}{\sqrt{2}}Z & \frac{\check{e}_{26}}{\sqrt{2}}Z\\ \sqrt{2}Z & 1+\check{\epsilon}_{11}Z^2-(1+2R)\Omega^2 & (\check{e}_{26}+\check{e}_{11}Z^2)\\ \sqrt{2}\check{e}_{26}Z & \check{e}_{11}Z^2+\check{e}_{26}(1-\frac{\beta A}{\pi^2}) & -(\check{\epsilon}_{11}Z^2+1)\end{vmatrix}=0. \tag{37}$$

For a given value of Ω, the dispersion relation in (37) can be solved for three pairs of roots $\pm Z_r(r=0,1,2)$. And by inserting each of these roots into (35), the corresponding amplitudes ratios

$$\beta_{0r}=\frac{C_{0r}(Z_r)}{C_{2r}(Z_r)},\beta_{1r}=\frac{C_{1r}(Z_r)}{C_{2r}(Z_r)},r=0,1,2, \tag{38}$$

can be calculated.

Hence, we have

$$\begin{aligned}\bar{u}_2^{(0)}&=\sum_{r=0}^{2}C_{2r}\beta_{0r}\sin\frac{Z_r\pi}{2}\frac{x_1}{b_0}e^{i\omega t},\\ \bar{u}_1^{(1)}&=\sum_{r=0}^{2}C_{2r}\beta_{1r}\cos\frac{Z_r\pi}{2}\frac{x_1}{b_0}e^{i\omega t},\\ \bar{\phi}^{(1)}&=\sum_{r=0}^{2}\sqrt{\frac{c_{66}}{\epsilon_{22}}}\,C_{2r}\cos\frac{Z_r\pi}{2}\frac{x_1}{b_0}e^{i\omega t},\end{aligned} \tag{39}$$

and

$$\begin{aligned}\bar{T}_6^{(0)}&=\frac{\pi c_{66}}{b_0}\sum_{r=0}^{2}\left[Z_r\beta_{0r}+\frac{1}{\sqrt{2}}(\beta_{1r}+\check{e}_{26})\right]C_{2r}\\ &\quad\times\cos\frac{Z_r\pi}{2}\frac{x_1}{b_0}e^{i\omega t},\\ \bar{D}_2^{(0)}&=\frac{\pi\sqrt{\epsilon_{22}c_{66}}}{b_0}\sum_{r=0}^{2}\left[\check{e}_{26}Z_r\beta_{0r}+\frac{1}{\sqrt{2}}(\check{e}_{26}\beta_{1r}-1)\right]C_{2r}\\ &\quad\times\cos\frac{Z_r\pi}{2}\frac{x_1}{b_0}e^{i\omega t},\\ \bar{T}_1^{(1)}&=-\frac{\pi c_{66}}{2b_0}\sum_{r=0}^{2}Z_r(\check{c}_{11}\beta_{1r}+\check{e}_{11})C_{2r}\sin\frac{Z_r\pi}{2}\frac{x_1}{b_0}e^{i\omega t},\end{aligned}$$

$$\bar{D}_1^{(1)} = -\frac{\pi\sqrt{\epsilon_{22}c_{66}}}{2b_0}\sum_{r=0}^{2} Z_r(\breve{e}_{11}\beta_{1r} - \breve{\epsilon}_{11})C_{2r} \times \sin\frac{Z_r\pi}{2}\frac{x_1}{b_0}e^{i\omega t}. \quad (40)$$

By replacing C_{2r} with $A_{6+r}(r = 0, 1, 2)$ and dropping the time factor $e^{i\omega t}$, (39) and (40) can be rewritten as

$$\begin{aligned}
\bar{u}(\bar{X}) &= A_6\bar{W}_0(\bar{X}) + A_7\bar{W}_1(\bar{X}) + A_8\bar{W}_2(\bar{X}),\\
\bar{\phi}(\bar{X}) &= A_6\bar{L}_0(\bar{X}) + A_7\bar{L}_1(\bar{X}) + A_8\bar{L}_2(\bar{X}),\\
\bar{\eta}(\bar{X}) &= A_6\bar{S}_0(\bar{X}) + A_7\bar{S}_1(\bar{X}) + A_8\bar{S}_2(\bar{X}),\\
\bar{T}_6^{(0)}(\bar{X}) &= A_6\bar{Q}_0(\bar{X}) + A_7\bar{Q}_1(\bar{X}) + A_8\bar{Q}_2(\bar{X}),\\
\bar{D}_2^{(0)}(\bar{X}) &= A_6\bar{H}_0(\bar{X}) + A_7\bar{H}_1(\bar{X}) + A_8\bar{H}_2(\bar{X}),\\
\bar{T}_1^{(1)}(\bar{X}) &= A_6\bar{M}_0(\bar{X}) + A_7\bar{M}_1(\bar{X}) + A_8\bar{M}_2(\bar{X}),\\
\bar{D}_1^{(1)}(\bar{X}) &= A_6\bar{P}_0(\bar{X}) + A_7\bar{P}_1(\bar{X}) + A_8\bar{P}_2(\bar{X}), \quad (41)
\end{aligned}$$

where

$$\begin{aligned}
\bar{W}_r(\bar{X}) &= \beta_{1r}\cos\frac{Z_r\pi}{2}\frac{a}{b_0}\bar{X},\\
\bar{S}_r(\bar{X}) &= \beta_{0r}\sin\frac{Z_r\pi}{2}\frac{a}{b_0}\bar{X},\\
\bar{L}_r(\bar{X}) &= \sqrt{\frac{c_{66}}{\epsilon_{22}}}\cos\frac{Z_r\pi}{2}\frac{a}{b_0}\bar{X},\\
\bar{Q}_r(\bar{X}) &= \frac{\pi c_{66}}{b_0}\left[Z_r\beta_{0r} + \frac{1}{\sqrt{2}}(\beta_{1r} + \breve{e}_{26})\right]\cos\frac{Z_r\pi}{2}\frac{a}{b_0}\bar{X},\\
\bar{H}_r(\bar{X}) &= \frac{\pi\sqrt{\epsilon_{22}c_{66}}}{b_0}\left[\breve{e}_{26}Z_r\beta_{0r} + \frac{1}{\sqrt{2}}(\breve{e}_{26}\beta_{1r} - 1)\right] \times\cos\frac{Z_r\pi}{2}\frac{a}{b_0}\bar{X},\\
\bar{M}_r(\bar{X}) &= -\frac{\pi c_{66}}{2b_0}Z_r(\breve{c}_{11}\beta_{1r} + \breve{e}_{11})\sin\frac{Z_r\pi}{2}\frac{a}{b_0}\bar{X},\\
\bar{P}_r(\bar{X}) &= -\frac{\pi\sqrt{\epsilon_{22}c_{66}}}{2b_0}Z_r(\breve{e}_{11}\beta_{1r} - \breve{\epsilon}_{11})\sin\frac{Z_r\pi}{2}\frac{a}{b_0}\bar{X},\\
\bar{X} &= x_1/a. \quad (42)
\end{aligned}$$

We note that (39) and (40) or (41) are the exact and closed form solutions for plates with uniform thickness and they can be utilized as a check for accuracy of the infinite power series solutions (27) and (29) for the plates with linearly varying thickness by letting $b(x_1) = b_0$.

The six unknown amplitudes A_0, A_1, A_2, A_3, A_4, and A_5 contained in (27) and (29) and the three unknown amplitudes A_6, A_7, and A_8 in (41) are to be determined by satisfying the continuity conditions, at $x_1 = \pm a$

$$\begin{aligned}
\bar{u} &= u,\ \bar{T}_1^{(1)} = T_1^{(1)},\\
\bar{\eta} &= \eta,\ \bar{T}_6^{(0)} = T_6^{(0)},\\
\bar{\phi} &= \phi,\ \bar{D}_1^{(1)} = D_1^{(1)}, \quad (43)
\end{aligned}$$

and the free-edge conditions of traction and charge, at $x_1 = \pm(a + d)$

$$\begin{aligned}
T_6^{(0)} &= 0,\\
T_1^{(1)} &= 0,\\
D_1^{(1)} &= 0. \quad (44)
\end{aligned}$$

VI. The Forced Vibrations of Beveled Plates

The homogeneous solutions of the symmetric thickness-shear and anti-symmetric flexural vibrations in the uniform portion of the plate are given by (39) with A set to 1. For the particular solutions corresponding to an impressed alternating voltage of $\phi = \pm\phi_0 e^{i\omega t}$ at $x_2 = \pm b_0$ and $|x_1| \le a$, we let[1]

$$\begin{aligned}
\bar{u}_2^{(0)} &= 0,\\
\bar{u}_1^{(1)} &= B_1 e^{i\omega t},\\
\bar{\phi}^{(1)} &= \sqrt{\frac{c_{66}}{\epsilon_{22}}}\,B_2 e^{i\omega t}. \quad (45)
\end{aligned}$$

By substituting the above into (32) with $D^{(1)}$ given by $(33)_5$, we have

$$\begin{pmatrix} 1-(1+2R)\Omega^2 & \breve{e}_{26} \\ -\breve{e}_{26}(1-\frac{8A}{\pi^2}) & 1 \end{pmatrix}\begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{8A}{\pi^2}\sqrt{\frac{\epsilon_{22}}{c_{66}}}\,\phi_0 \end{pmatrix}. \quad (46)$$

For frequencies not equal to the thickness-shear cut-off frequencies, *i.e.*

$$D \equiv 1 - (1 + 2R)\Omega^2 + \breve{e}_{26}^2(1 - \frac{8A}{\pi^2}) \neq 0, \quad (47)$$

the amplitudes B_1 and B_2 are

$$\begin{aligned}
B_1 &= -\frac{8A}{\pi^2 D}\sqrt{\frac{\epsilon_{22}}{c_{66}}}\,\breve{e}_{26}\phi_0,\\
B_2 &= \frac{8A}{\pi^2 D}\sqrt{\frac{\epsilon_{22}}{c_{66}}}\left[1 - (1 + 2R)\Omega^2\right]\phi_0. \quad (48)
\end{aligned}$$

By combining the particular solutions (48) and the homogeneous solutions (39), we have the complete solutions under thickness excitation

$$\begin{aligned}
\bar{u}_2^{(0)} &= \sum_{r=0}^{2} C_{2r}\beta_{0r}\sin\frac{Z_r\pi}{2}\frac{x_1}{b_0}e^{i\omega t},\\
\bar{u}_1^{(1)} &= \left(\sum_{r=0}^{2} C_{2r}\beta_{1r}\cos\frac{Z_r\pi}{2}\frac{x_1}{b_0} + B_1\right)e^{i\omega t},\\
\bar{\phi}^{(1)} &= \sqrt{\frac{c_{66}}{\epsilon_{22}}}\left(\sum_{r=0}^{2} C_{2r}\cos\frac{Z_r\pi}{2}\frac{x_1}{b_0} + B_2\right)e^{i\omega t}. \quad (49)
\end{aligned}$$

Substitution of above into (40) results in

$$\bar{T}_6^{(0)} = \frac{\pi}{b_0}\left\{\sum_{r=0}^{2} c_{66}\left[Z_r\beta_{0r} + \frac{1}{\sqrt{2}}(\beta_{1r} + \check{e}_{26})\right]C_{2r} \times \cos\frac{Z_r\pi}{2}\frac{x_1}{b_0} - \frac{4\sqrt{2}A\phi_0 e_{26}}{\pi^2 D}(1+2R)\Omega^2\right\}e^{i\omega t},$$

$$\bar{D}_2^{(0)} = \left\{\frac{\pi\sqrt{\epsilon_{22}c_{66}}}{b_0}\sum_{r=0}^{2}\left[\check{e}_{26}Z_r\beta_{0r} + \frac{1}{\sqrt{2}}(\check{e}_{26}\beta_{1r} - 1)\right]C_{2r} \times \cos\frac{Z_r\pi}{2}\frac{x_1}{b_0} - \frac{4\sqrt{2}A\epsilon_{22}\phi_0}{\pi b_0 D}\left[1 + \check{e}_{26}^2 - (1+2R)\Omega^2\right]\right\}e^{i\omega t},$$

$$\bar{T}_1^{(1)} = -\frac{\pi c_{66}}{2b_0}\sum_{r=0}^{2} Z_r(\check{c}_{11}\beta_{1r} + \check{e}_{11})C_{2r}\sin\frac{Z_r\pi}{2}\frac{x_1}{b_0}e^{i\omega t},$$

$$\bar{D}_1^{(1)} = -\frac{\pi\sqrt{\epsilon_{22}c_{66}}}{2b_0}\sum_{r=0}^{2} Z_r(\check{e}_{11}\beta_{1r} - \check{\epsilon}_{11})C_{2r}\sin\frac{Z_r\pi}{2}\frac{x_1}{b_0}e^{i\omega t}. \tag{50}$$

Both (49) and (50) will be expressed in the same form as that of (41) with $A = 1, R \neq 0$ to represent the solutions of the forced vibrations in the uniform and electrode-plated portions of the plate, *i.e.* for $|x_1| \leq a$ or $|\bar{X}| \leq 1$, where $\bar{X} = x_1/a$. For (27) and (29) to represent the solutions for the contoured and unelectroded portion of the plate, we let $A = 0, R = 0$ for $a \leq |x_1| \leq a + d$ or $0 \leq |X| \leq 1$, where $X = (x_1 - a)/d$.

Finally, by substituting (27), (29), (49), and (50) into the continuity conditions (43) and edge conditions (44), we obtain

$$\begin{pmatrix}
W_0(0) & W_1(0) & W_2(0) & W_3(0) & W_4(0) & W_5(0) & -W_0(1) & -W_1(1) & -W_2(1)\\
L_0(0) & L_1(0) & L_2(0) & L_3(0) & L_4(0) & L_5(0) & -L_0(1) & -L_1(1) & -L_2(1)\\
S_0(0) & S_1(0) & S_2(0) & S_3(0) & S_4(0) & S_5(0) & -S_0(1) & -S_1(1) & -S_2(1)\\
Q_0(0) & Q_1(0) & Q_2(0) & Q_3(0) & Q_4(0) & Q_5(0) & -Q_0(1) & -Q_1(1) & -Q_2(1)\\
M_0(0) & M_1(0) & M_2(0) & M_3(0) & M_4(0) & M_5(0) & -M_0(1) & -M_1(1) & -M_2(1)\\
P_0(0) & P_1(0) & P_2(0) & P_3(0) & P_4(0) & P_5(0) & -P_0(1) & -P_1(1) & -P_2(1)\\
Q_0(1) & Q_1(1) & Q_2(1) & Q_3(1) & Q_4(1) & Q_5(1) & 0 & 0 & 0\\
M_0(1) & M_1(1) & M_2(1) & M_3(1) & M_4(1) & M_5(1) & 0 & 0 & 0\\
P_0(1) & P_1(1) & P_2(1) & P_3(1) & P_4(1) & P_5(1) & 0 & 0 & 0
\end{pmatrix}
\begin{pmatrix} A_0\\A_1\\A_2\\A_3\\A_4\\A_5\\A_6\\A_7\\A_8 \end{pmatrix} =
\begin{pmatrix}
-\frac{8\phi_0}{\pi^2 D}\frac{e_{26}}{c_{66}}\\
\frac{8\phi_0}{\pi^2 D}\left[1 - (1+2R)\Omega^2\right]\\
0\\
-\frac{4\sqrt{2}\phi_0}{\pi b_0 D}e_{26}(1+2R)\Omega^2\\
0\\0\\0\\0\\0
\end{pmatrix}, \tag{51}$$

from which the amplitudes $A_i (i = 0, 1, 2, 3, 4, 5, 6, 7, 8)$ can be computed for given value of ϕ_0 and ω.

For the forced thickness-shear and flexural vibrations of a fully plated crystal plate of uniform thickness, (51) is reduced to

$$\sum_{r=0}^{2}\left(Z_r\beta_{0r} + \frac{\beta_{1r} + \check{e}_{26}}{\sqrt{2}}\right)A_r\cos\frac{Z_r\pi}{2}\frac{a}{b_0} = \frac{4\sqrt{2}\phi_0 e_{26}}{\pi^2 D c_{66}}(1+2R)\Omega^2,$$

$$\sum_{r=0}^{2} Z_r(\check{c}_{11}\beta_{1r} + \check{e}_{11})A_r\sin\frac{Z_r\pi}{2}\frac{a}{b_0} = 0,$$

$$\sum_{r=0}^{2} Z_r(\check{e}_{11}\beta_{1r} - \check{\epsilon}_{11})A_r\sin\frac{Z_r\pi}{2}\frac{a}{b_0} = 0. \tag{52}$$

We note that $\phi_0 = 0$ in the righthand side of (51) if the electrodes are shorted. Therefore, the frequency equation for the free vibrations of the beveled plates with shorted electrodes is obtained by setting the determinant of the coefficient matrix of (51) equal to zero and it may be regarded as a functional relationship between the resonance frequencies Ω and the four parameters of the resonator as

$$F(\Omega, \frac{a}{b_0}, \frac{d}{b_0}, \theta = 1 - \frac{b_1}{b_0}, R) = 0. \tag{53}$$

Once amplitudes are solved from (51), the surface charge Q can be computed by integrating $D_2^{(0)}$ in $(50)_2$ over the electroded area

$$Q = -\int_{-a}^{a}\int_{-c}^{c}\bar{D}_2^{(0)}dx_1dx_3 = -4\sqrt{\epsilon_{22}c_{66}}\sum_{r=0}^{2}\left[\check{e}_{26}\beta_{0r} + \frac{1}{\sqrt{2}Z_r}(\check{e}_{26}\beta_{1r} - 1)\right]A_r \times \sin\frac{\pi Z_r}{2}\frac{a}{b_0} + \frac{8\sqrt{2}\epsilon_{22}\phi_0}{\pi D}\frac{a}{b_0}\left[1 + \check{e}_{26}^2 - (1+2R)\Omega^2\right]. \tag{54}$$

In (54), we have set $2c = 1$. By defining the motional and static capacitances as

$$C_m = \frac{Q}{2\phi_0}, C_s = \frac{2a\epsilon_{22}\phi_0}{b_0}, \tag{55}$$

and substituting (55) into (54), we obtain the capacitance ratio

$$\frac{C_m}{C_s} = -\frac{b_0}{a}\sqrt{\frac{c_{66}}{\epsilon_{22}}}\sum_{r=0}^{2}\left[\check{e}_{26}\beta_{0r} + \frac{1}{\sqrt{2}Z_r}(\check{e}_{26}\beta_{1r} - 1)\right]A_r \times \sin\frac{\pi Z_r}{2}\frac{a}{b_0} + \frac{2\sqrt{2}}{\pi}\frac{1 + \check{e}_{26}^2 - (1+2R)\Omega^2}{1 + \check{e}_{26}^2\left(1 - \frac{8A}{\pi^2}\right) - (1+2R)\Omega^2}. \tag{56}$$

VII. Computational Results

For the computations of AT-cut of quartz plates, Bechmann's value of material constants of quartz are used.[3] For accurate and efficient evaluation of the functions $W_i(X)$ and $L_i(X)$ of the infinite power series in (29),

a multiple precision program developed by Bailey has been employed.[4]

The resonance frequencies Ω are computed from (53), for beveled AT-cut quartz plates with shorted electrodes, as function of a/b_0 with $d/b_0 = 13.00, R = 0.007$ and $\theta = 0.60$, and the results is shown in Fig. 2. Computations for $\theta = 0.00, 0.20$, and 0.40 are also made, but the results are not shown here. The effect of θ, the slope of the bevel, on the resonances can be summarized as follows:

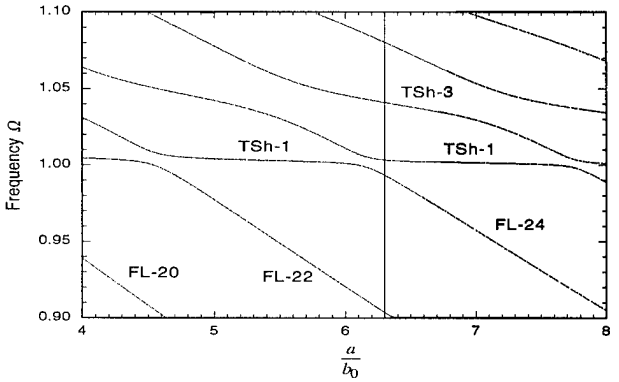


Fig. 2 Frequency spectrum or Ω *vs.* a/b_0 of the thickness-shear and flexural vibrations of beveled plates of AT-cut quartz, with $d/b_0 = 13.00$, $R = 0.007$, and $\theta = 0.60$.

(a) The larger the values of θ, the weaker the coupling between the thickness-shear and flexural modes.

(b) The larger the values of θ, the lower the frequencies of the flexural modes and the higher the frequencies of the fundamental thickness-shear mode and those of its anharmonic overtones.

Effects of $a/b_0, d/b_0$, and θ on the frequency spectra and on the mode shapes computed from present solutions are very similar to those of the previous results computed from the governing equations without piezoelectric coupling. Frequencies are, in general, slightly higher for the present case due to the piezoelectric stiffening effect on the elastic stiffness of the plate. For instance, the thickness-shear cut-off frequencies of an infinite and uniform plate completely covered with electrodes can be obtained from (47) by letting $D = 0 (A = 1)$,[1]

$$\Omega = 1 - R + \frac{1}{2}(1 - \frac{\pi^2}{8})\breve{e}_{26}^2, \qquad (57)$$

where $\breve{e}_{26} = e_{26}/\sqrt{\epsilon_{22}c_{66}}$ defined in (15) is identical to k_{26}, the piezoelectric coupling factor. We see from (57) that the thickness-shear cut-off frequencies are increased by the piezoelectric coupling and decreased by the mass effect of the electrodes.

For forced vibrations, (51) are solved for the belved plate with $d/b_0 = 12.50, R = 0.007, \theta = 0.60$, and $a/b_0 = 6.30$, which is denoted by the vertical line in Fig. 2. At the forcing frequency near the fundamental thickness-shear resonance, distribution of the displacements $u_1^{(1)}$ and $u_2^{(0)}$, the electric potential $\phi^{(1)}$, and the electric displacement $D_2^{(0)}$ are computed and given in Figs. 3, 4, and 5, respectively. Similar computations are made at forcing frequencies near the resonance of other modes, *i.e.* near the intersection points of other frequency branches and the vertical line $a/b_0 = 6.30$ in Fig. 2. It is found that at TSh-1 frequency, amplitudes of $u_1^{(1)}$ and $D_2^{(0)}$ are of two order of magnitude and the amplitude of $\phi^{(1)}$ is of one order of magnitude greater than the corresponding ones at the forcing frequencies near other resonances.

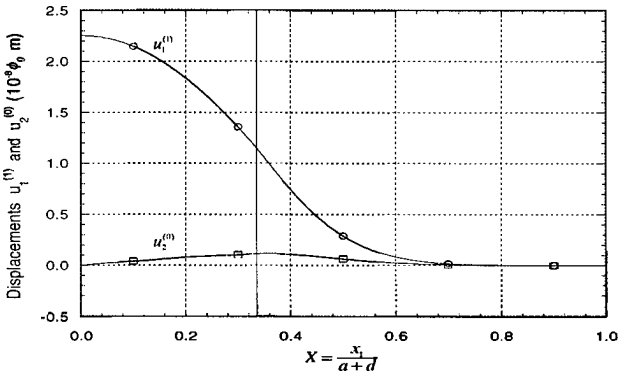


Fig. 3 Distribution of mechanical displacements $u_1^{(1)}$ and $u_2^{(0)}$ at the forcing frequency $\Omega = 1.0125$(near the fundamental thickness-shear resonance), with $a/b_0 = 6.30$, $d/b_0 = 12.50$, $R = 0.007$, and $\theta = 0.60$.

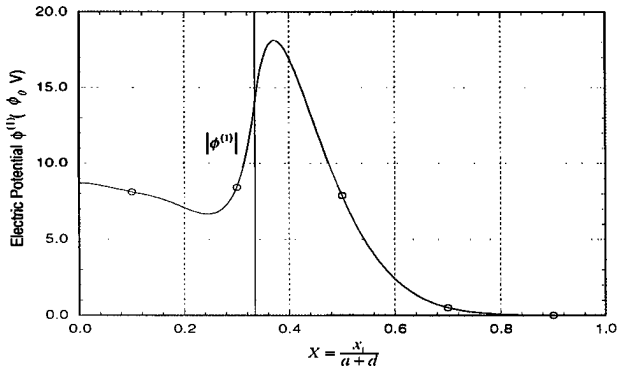


Fig. 4 $\phi^{(1)}$ *vs.* X at Ω at $\Omega = 1.0125$, with $a/b_0 = 6.30$, $d/b_0 = 12.50$, $R = 0.007$, and $\theta = 0.60$.

To check the accuracy of the present solutions, the capacitance ratio C_m/C_s is computed from (56) as a function of forcing frequency Ω for a uniform and rectangu-

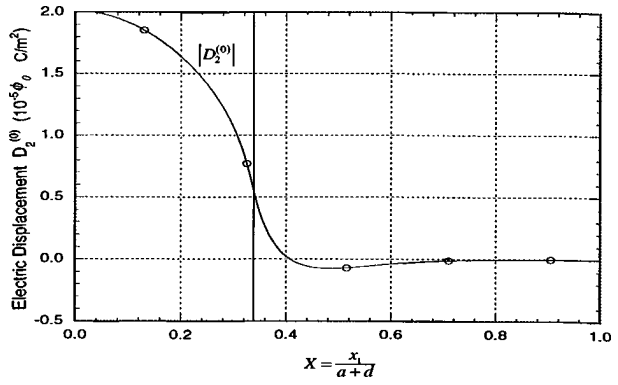


Fig. 5 $D_2^{(0)}$ *vs.* X at Ω at $\Omega = 1.0125$, with $a/b_0 = 6.30$, $d/b_0 = 12.50$, $R = 0.007$, and $\theta = 0.60$.

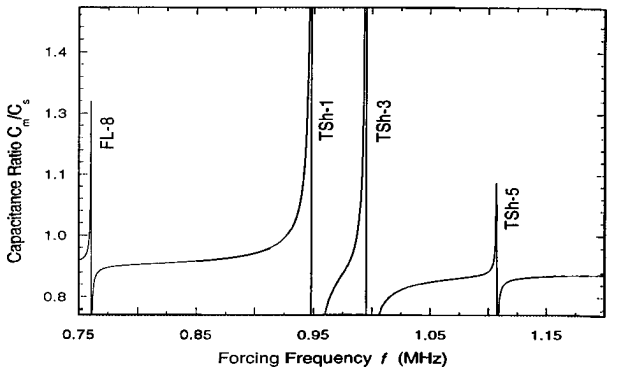


Fig. 6 Computed capacitance ratio C_m/C_s of forced thickness-shear and flexural vibrations of a uniform AT-cut quartz plate, with $a/b_0 = 8.0391(2c = 1)$, $R = 0.007$, and $d/b_0 = \theta = 0.00$.

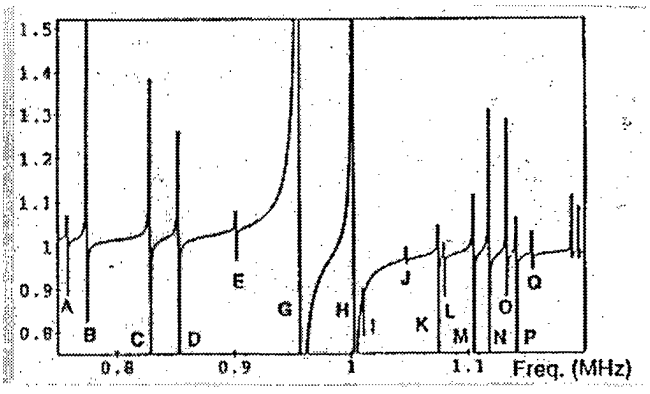


Fig. 7 Measured capacitance ratio C_m/C_s as a function of forcing frequency of a uniform rectangular plate of AT-cut quartz, with $a/b_0 = 8.0391$ and $c/b_0 = 4.03$, by Sekimoto, Watanabe, and Nakazawa(Ref. 5).

lar plate with $a/b_0 = 8.0391, 2c = 1, R = 0.007, d/b_0 = \theta = 0.00$. The predicted result is plotted in Fig. 6 and compared with measured values of Sekimoto, Watanabe, and Nakazawa[5] for which $a/b_0 = 8.0391, c/b_0 = 4.03$ as shown in Fig. 7. Comparison of Fig. 6 with Fig. 7 shows that the distribution and the shape of the C_m/C_s *vs.* Ω curves for the thickness-shear and flexural modes match well with the corresponding ones of the measured values. Excellent prediction of capacitance ratio were presented in Ref. 5 from the finte element solutions of Mindlin's first-order equations(using five modes) for the uniform and rectangular AT-cut quartz plates.

To investigate the effect of contouring on the capacitance ratio, C_m/C_s *vs.* Ω are computed for beveled plates with $a/b_0 = 6.30, d/b_0 = 12.50, R = 0.007$, and $\theta = 0.00, 0.20, 0.40$, and 0.60, and the results are given in Figs. 8, 9, 10, and 11, respectively. Careful examina-

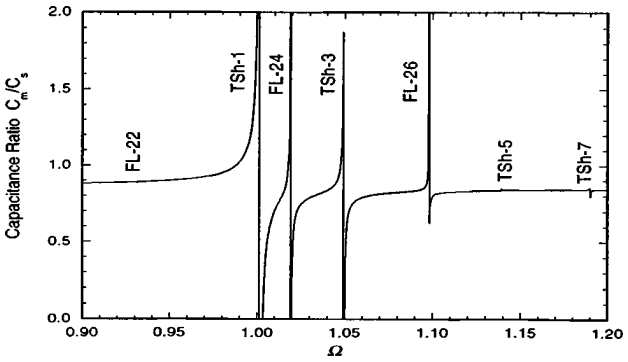


Fig. 8 Computed C_m/C_s *vs.* Ω of a beveled AT-cut plate of quartz with $a/b_0 = 6.30$, $d/b_0 = 12.50$, $R = 0.007$, and $\theta = 0.00$.

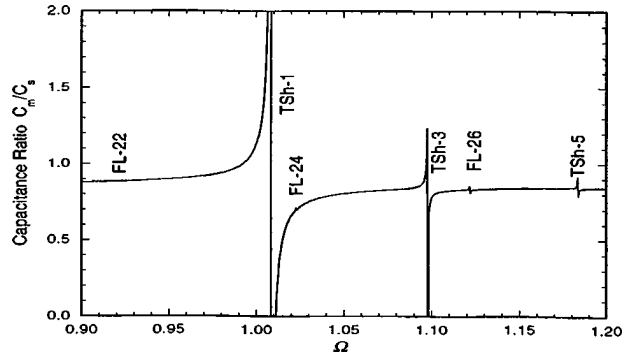


Fig. 9 Computed C_m/C_s *vs.* Ω of a beveled AT-cut plate of quartz with $a/b_0 = 6.30$, $d/b_0 = 12.50$, $R = 0.007$, and $\theta = 0.20$.

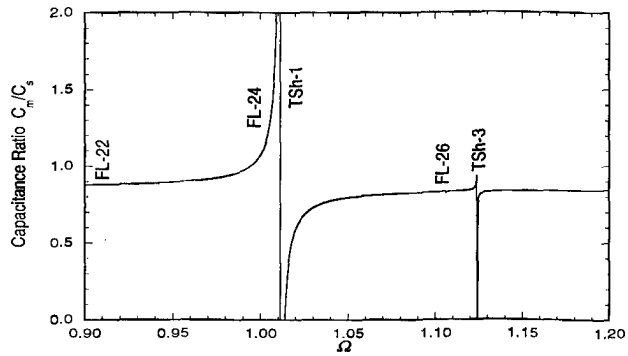


Fig. 10 Computed C_m/C_s *vs.* Ω of a beveled AT-cut plate of quartz with $a/b_0 = 6.30$, $d/b_0 = 12.50$, $R = 0.007$, and $\theta = 0.40$.

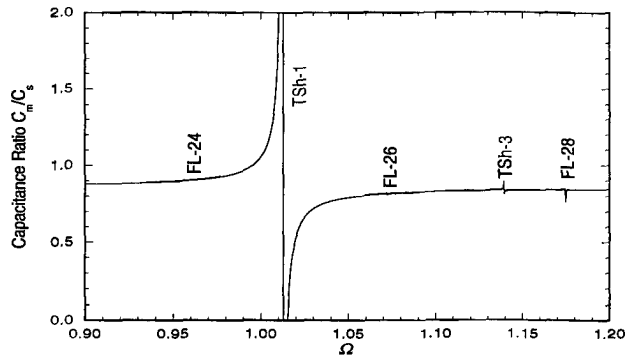


Fig. 11 Computed C_m/C_s *vs.* Ω of a beveled AT-cut plate of quartz with $a/b_0 = 6.30$, $d/b_0 = 12.50$, $R = 0.007$, and $\theta = 0.60$.

tion of Figs. 8 $\sim$ 11 reveals that the increasing the value of θ has the effect of (a) decreasing the magnitude of C_m/C_s of the flexural modes as well as that of the anharmonic overtones of the fundamental thickness-shear mode, and (b) decreasing the frequencies of the spikes of the flexural modes and increasing the frequencies of the spikes of overtones of the thickness-shear mode.

Acknowledgements

The support to Ji Wang from Motorola Inc. through the University Partnerships in Research Program is acknowledged.

References

[1] S. Syngellakis, P. C. Y. Lee, and J. P. Hou. "A two-dimensional theory for high-frequency vibrations of piezoelectric crystal plates with or without electrodes". *J. Appl. Phy.*, **61**(4):1249–1262, 1987.

[2] P. C. Y. Lee and Ji Wang. "Thickness-shear and flexural vibrations of contoured crystal strip resonators". *Proceedings of IEEE 1993 Ultrasonics Symposium*, pages 559–564, 1993.

[3] R. Bechmann. "Elastic and piezoelectric constants of alpha-quartz". *Phys. Rev.*, **110**:1060, 1958.

[4] D. H. Bailey. "Multiprecision translation and excution of FORTRAN programs". *ACM Transactions on Mathematical Software*, **19**(3):288–319, 1993.

[5] H. Sekimoto, Y. Watanabe, and M. Nakazawa. "Forced vibrations of thickness-flexure, face-shear and face-flexure in rectangular AT-cut quartz plates". *Proceedings of the 1992 IEEE Frequency Control Symposium*, pages 532–536, 1992.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A NEW MODEL FOR THE THICKNESS SHEAR RESONATORS WITH SPHERICAL CONTOURS: APPLICATION TO BEVELLED DEVICES.

J.Détaint*, B.Capelle**, A.Zarka**, Y.Zheng**

* FRANCE TELECOM CNET/PAB, BP 107 92225 Bagneux France,
** L.M.C.P. Université Paris VI & CNRS 75252 Paris France,

Abstract: The contoured resonator are mostly employed to obtain the very low losses devices employed to generate very stable frequencies. The partially contoured (bevelled) resonators are also used for filtering. Their design remains still some how empirical, so that a better theoretical understanding of such devices is very desirable. In a preliminary study, experimental determinations of the mode shapes were performed using an important set of bevelled resonators, with designs ranging from nearly biconvexe ones to nearly plane ones having only a very small bevel at their edge. Several conclusions were drawn from these experiments.

The present model of the contoured and the partially contoured resonators, is based upon the Tiersten theory of the transversely varying essentially thickness modes. We propose a new method of resolution of the Tiersten partial derivative equations based upon the algebraical solutions obtained for the same problem in the case of a very close geometry (non spherical contour respecting the lateral anisotropy of the plate). These simple solutions are of a very large generality since they contain nearly all the previously known ones. They bring a better understanding of the different kinds of thickness shear contoured resonators. In this model, for the plane electroded regions (bevelled resonators), we use solutions of the Tiersten equations which are basically series expansions in the exact solutions for a case where the electrodes and the plate have elliptical geometries respecting the lateral anisotropy. For the contoured regions, we use similarly a representation of the shear displacement consisting in an expansion in the eigen solutions obtained for the above mentioned non-spherical contours. The continuity conditions between the different regions and the boundary conditions at the edge are expressed at a discrete number of points to fit nearly or more exactly, the actual geometry of the resonators presenting a spherical contour and a circular geometry. The set of linear equations so obtained constitute an homogeneous linear system whose determinant must vanish to have a solution. This condition is a frequency equation which is solved to find the eigen frequencies and the eigen modes.

The numerical implementation of this model makes use, for the contoured regions, of the confluent hypergeometric (Whittaker) functions [$M_{k,m}(r^2/2)$ or $W_{k,m}(r^2/2)$] which are bounded at zero or at infinity, and of the Jn(r) Bessel functions for the flat region. The convergence of the solution with the discretization appears to be extremely fast, and a very high accuracy seems to be obtained. The comparison with the experimental results indicates a very good agreement. Several important features observed during the experimental study were confirmed and explained by the calculations.

INTRODUCTION

The contoured thickness mode resonators are key devices for the generation of very stable frequencies or for filtering in modern electronics. Oscillators using these resonators constitute the most interesting solution to obtain the very high medium and short term stability required for synchronisation of telecommunication systems particularly those employing high baud rates transmissions. Due to a rapid evolution towards bauds rate exceeding by more than one order of magnitude those presently used, this application will be more and more demanding about the Q.f product obtained from the resonators.

The plano-convexe resonators have, in the frequency range 0.5-50 MHz, the largest Q.f factors observed for quartz and hence lead to outstanding spectral characteristics for the oscillators. Several theories of these resonators were established and were confirmed by the experiments. However, more recent experiments made using either quartz with much improved quality or new crystals (LGS, Lithium tantalate) presenting extremely reduced intrinsic acoustic losses, have indicated that their ultimate performances (Q factors) may still be limited by their design or realisation. So that a renewed interest exists for very accurate models that could bring a new understanding of this question and the possibility to improve further the design and the quality of such resonators.

The partially contoured (bevelled) resonators are of very frequent use for frequency generation or for filtering, in the frequency range from .5 to near 10.Mhz (for the fundamental mode). They have the interest to lead to much larger values of the capacitance ratio C_1/C_0 than the plano-convexe resonators and so to be well adapted to VCXO applications. They can be made to have very attenuated anharmonic responses for filter applications. Their design remains still very empirical and is in most cases based on a cut and try process. Recently, a new interest for this kind of devices has appeared with the advent of automatized elaboration process and the research of a much greater miniaturization. In their practical implementation, it has been observed that their characteristics could be very sensitive to small changes in their design parameters. So that, a better understanding of such devices is very desirable. Furthermore, recent observations have shown that phenomena suspected to exist in very high Q plano-convexe resonators and that are presently very difficult to compute or to observe, can be detected more easily in bevelled resonators.

A new interest for other kind of contoured resonators (such as the convexe resonators using two curvatures) that could bring new possibilities for applications requiring at the same time very high frequency stabilities and a larger capacitance ratio for a given overtone is appearing in relation with a recent advance in their theory.

Beside their theoretical interest, it is more than probable that **the use of non spherical contoured resonators will constitute the optimum solution** for several questions, particularly those concerning the obtention of the ultimate Q factors (frequency stabilities) achievable with piezo-electric devices. They can now be built using the modern techniques employed for non-spherical optics.

More generally, apart from the intrinsic interest resulting of the thermal and time stabilities of the properties of the best piezo-electric materials, and of their very reduced dissipation, most of the favourable properties of the contoured resonators are a consequence of the occurrence in these resonators of a particular type of energy trapping which results mostly from the lateral variations of the thickness.

0-7803-2500-1/95 $4.00 © 1995 IEEE

Mathematical models of these resonators having more and more accuracy were presented in the recent years. They are essentially based upon two related families of plate theories.

1/The plate theories due to Mindlin [1],[2], Lee [3], Tiersten [4], Peach [5], and some other authors, consider several types of expansion of the mechanical and electrical variables as functions of the thickness coordinate. These theories give informations concerning the coupling of the thickness shear displacement with other plate modes. Their implementation in the case of resonators having two comparable lateral dimensions requires a very important amount of numerical calculations due to the number of coupled partial derivative equations and to the possibility to have very fastly varying components. The interest of such solutions to represent very accurately the reality even in intricate cases was demonstrated recently by Y.K.Yong [6] [7].

2/The theories of essentially thickness vibrations of the piezo-electric plates established by H.F.Tiersten and co-workers for the resonators vibrating near one overtone of a thickness mode which have a thickness dependence essentially similar to that of the corresponding one dimensional mode of an infinite piezo-electric plate [8] [9] [10]. This theory lead to very accurate results when there is no too strong coupling with the plate modes.

Models of plano-convex resonators of successively increased generality were established by H.F.Tiersten and co-workers [9] [10]. Resonators with cylindrical bevels were considered by Vangheluwe [11] and Sinha [12]. Plane and corrugated resonators with arbitrary electrode geometry were then considered [13][14]. Peach [15] has proposed a method to establish other approximate equations accounting for the lateral variations of the thickness of the plate. The effect of finite lateral dimensions in plano-convex resonators were studied by Dulmet and Tiersten [16]. Solutions for plano-convexe, bevelled, and non spherical resonators with finite lateral dimensions using either the finite element method or a semi-algebraical technique were then presented [17].

Very recently, the bevelled resonators [18], and the resonators with shallow contours and two different curvatures [19][20] were the objet of new theoretical developments. Plate theories were used to study the influence of the contour on the coupling of flexure with the thickness shear [21] and for the case of contoured strips [21].

In this paper, after reporting preliminary experimental observations of the modes shapes of several resonators made to better understand several points, we propose a new method for the modelization of the contoured resonators. As in most of the recent developments concerning this subject, this analysis is based upon the theory of the essentially thickness modes. The approach chosen is fundamentally a matching technique using a functional basis constituted by the orthogonal eigen-solutions of a very nearby problem relative to totally or partially contoured resonators having specific un-spherical contours. In fact, we have previously shown [17] that in the case of un-spherical contours (such as ellipsoïds, elliptical paraboloïds, etc...) characterized by two principal curvatures being in the same ratio as those of the 2 coefficients M'_n and P'_n of the Tiersten equations, it exists very simple but extremely general solutions. These solutions could be also directly related to the modes (then purely elastic) that exist in isotropic material. These solutions present also the interest to display quite simply, in the un-spherical cases, several general features of the behaviour of partially or totally contoured resonators. A parameter appearing in the analysis allows to approximate the actual geometry of the usual resonators with spherical contours. Since, we express the continuity conditions at the electrode edge, and the boundary condition at the edge of the plate at a discrete number of points, we use this discretization also to fit approximately or more exactly (in a piece-wise approximation) the actual circular geometry of the resonator with a spherical contour.

In a last part, we compare the experimental results with those computed for different resonators geometry and using different types of boundary conditions. A discussion of the interest of the method is then made.

I PRELIMINARY EXPERIMENTS

Many experiment concerning the determination of the mode shape existing in plano-convexe resonators were reported. The most recent observations were made using stroboscopic X-ray topography [23][24]. In most case, they have confirmed the theoretical results. In the contrary, very few experimental observations of the modes shapes in bevelled resonators were reported. Several years ago we have made some experiments on such resonators made at CNET or by quartz companies [17]; but no systematic investigation of the influence of the design parameters were made. In order to get a better understanding of these resonators more systematic investigations were developed recently using a wide range of bevelled resonators having perfectly known geometrical parameters. The samples were mostly chosen among resonators studied at CNET to have very good electrical responses (mostly for filter applications). A great variety devices with a broad range of frequencies (815kHz to 5Mhz) and of geometrical parameters was retained. Here we will present only some results relative to the fundamental mode, chosen to illustrate the most typical facts observed. These experiments were made by X-ray topography using the synchrotron radiation delivered by the storage ring DCI at the L.U.R.E. (Orsay France).

BEVELLED RESONATOR , Large 2h /2l2 ratio,
Nearly biconvex, 2ho=2.275mm
Rc=24.7mm , 2L1= 4.18mm, 2L2=15.25mm

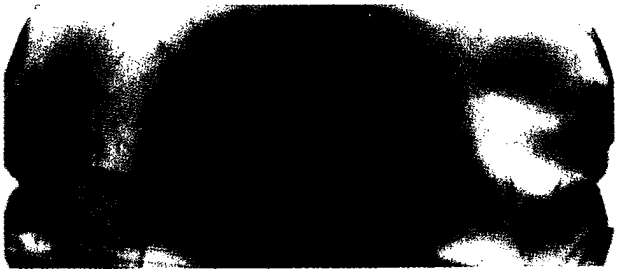


Figure 1a: U1 Component



Figure 1b: U3 Component

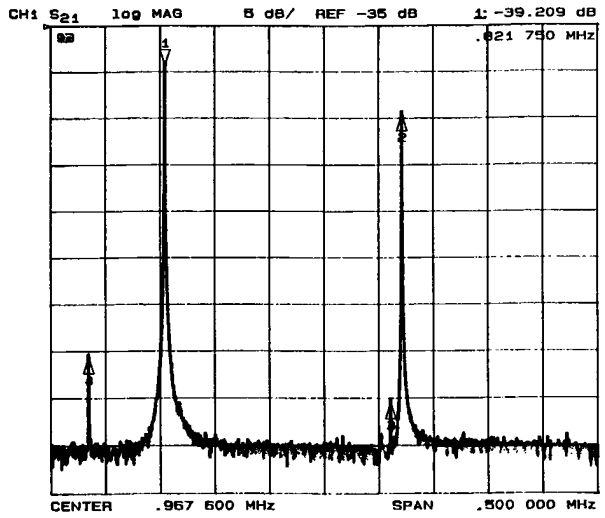


Figure 1c: Response curve of the resonator.

Rather conventional design, 2ho=0.7105mm,
Rc=60mm, 2L1=9.5mm, 2L2=14mm

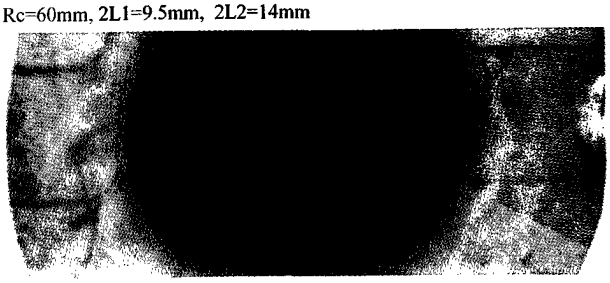


Figure 2a: U1 component , (normal excitation)

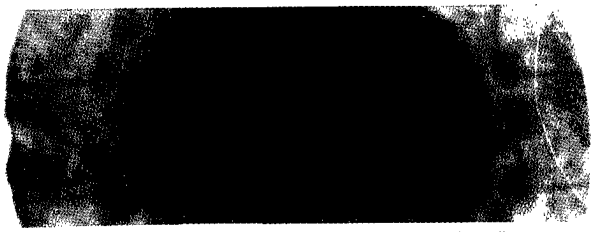


Figure 2b:U1 component (very high excitation level)

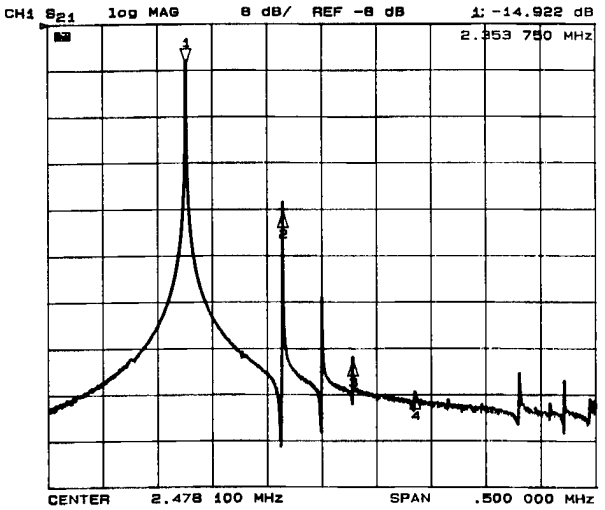


Figure 2c: Response curve of the resonator

The topograph of figure 1a and 1b are relative to a thick plate (821.75Khz, $2L_2/2h=6.70$) which is close to a biconvexe resonator. The main shear component (u_1) has a non negligible amplitude in the bevel particularly near the mounting clips. This resonators display a strong u_3 component (figure 1b) which corresponds much probably to a coupling with a high order face shear mode. This resonator display also a u_2 component (non represented) related to a flexure propagating in an unusual direction (nearly 30° from the z' axis).

On the topograph we can observe that the u_1 component displays a circular shape deformed near the fixation, much probably by an energy leakage. Other thick resonators having the same or similar designs parameters have displayed also strong coupling to the plate modes but the coupled components seems to be variable between a priori identical resonators (sensitivity to the fixation ?). On the response curve of figure 1c, weak plate modes can be observed before the main resonance and the first anharmonic (markers 4 and 3).

The topographs of figures 2a,2b are relative to a bevelled resonator having a rather usual design. It presents negligible non u_1 components. Most of the energy is concentrated in the plane region (totally electroded) where the mode shape is very nearly circular. The amplitude of vibration decreases in the bevel but not to zero at the edge of the plate. Irregularities can be seen in the energy distribution in the bevel. They can be better observed on topograph 2b taken with slightly different observation conditions and a much higher excitation level. These features were observed in several resonators, most often in the z' direction. An interpretation will be given as the result of a calculation in § IV. On the response curve of figure 2c we can observe an anharmonic response very typical of a bevelled crystal. The fast decrease of the amplitude of their response corresponds to the fact that the successive anharmonics have increasing spatial extension and are thus more and more attenuated by energy transfer to the fixation. Other resonators using the same design but using a more rigid fixation have displayed still reduced anharmonic response. The observations made for these resonators are believed to be representative of the behaviour of many bevelled resonators.

The topograph of figure 3a is relative to a large resonator having a plane central region with a diameter being about half the plate diameter. It presents only a u1 shear component we an apparent regular decrease in the bevel. The regular circular fringes near the edges are due to the diffraction process by a plate with varying thickness (see below). The corresponding response curve is displayed in figure 3b, these resonators have a very high Q factor.

The topographs of figure 4 are relative to a thin plate with a short bevel which is quite close to a plane resonator. In figure 4a, the fundamental mode has a nearly circular shape and it

Large Bevel 2l1=8.0mm, 2l2=15.5, Rc=60mm,. 2h=1.082mm,

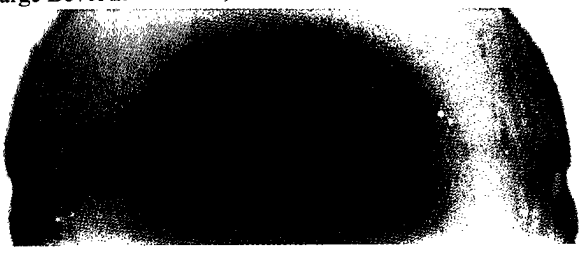


Figure 3: Bevelled resonator with a large bevel

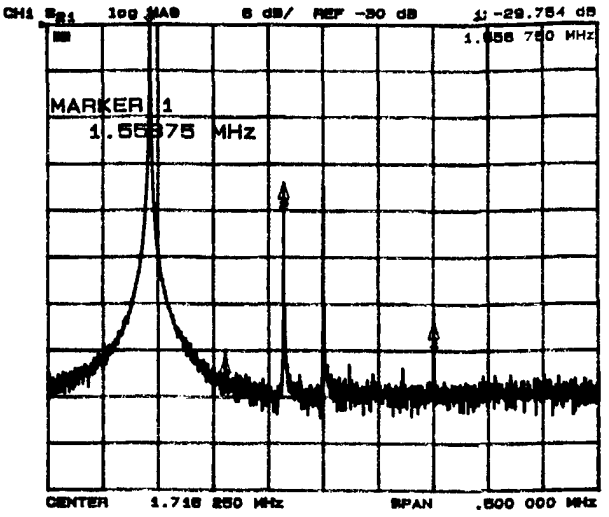


Figure3b: Response of the resonator.

Thin plate (small 2h/2L2) short bevel,
2ho=0.4125mm; 2L1=9.8mm, 2L2=13mm



Figure 4a:Fundamental mode U1

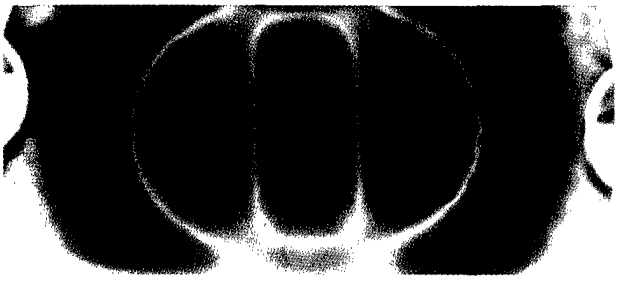


Figure4b:Anharmonic mode U1

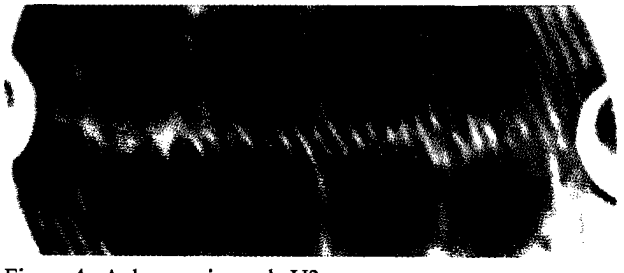


Figure 4c:Anharmonic mode U3

decrease fastly in the bevel, there no detectable acoustic energy near the edge. The fundamental mode has no observable u_2 or u_3 component. The fringes appearing near the edge have their origin in defects (plane growth band) or are x-ray diffraction contrasts due to the varying thickness near the edges (circular fringes). The anharmonic mode, (figures 4b and 4c) has a u_1 component which has a greater extension in the bevel and a u_3 component that indicates a strong coupling with other type(s) of mode. Irregularities in the amplitude variations of the u_1 component can be observed in the bevel particularly near the fixation. The observations made for this anharmonic are very typical of those made for the anharmonics of other bevelled resonators.

On the whole, this study has shown that for most the plates, empirically designed to have a good electrical response, and having a sufficiently great ratio of the diameter to the thickness, no strong coupling with the plate modes occurs. For plates for which the same ratio is under about 10, many cases of coupling where observed. During these experiments we were quite surprised to observe that many resonators having large Q factors (up to 1.10^5 to several 10^5) could have a non negligible displacement near the edge of the plate. When a shear displacement is observable near this edge, the experience indicates that the relevant approximate boundary condition is rather the vanishing of the normal derivative of the displacement than the vanishing of this latter (except quite probably near a fixation).

II RESONATORS WITH UN-SPHERICAL CONTOURS

II.1 The basis of the model.

As a basis of this model, we use the theory of essentially thickness shear modes established by H.F.Tiersten. In this theory, the lateral variations of the main thickness displacement are the solution of a partial derivative scalar equation which is valid for modes (anharmonics) in the vicinity of one overtone (number n) of a given thickness mode (noted (1)). This equation was first established for monoclinic plates then for plates of an arbitrary orientation [10]. In this case the equation has, in a coordinate system with x_2 normal to the thickness and x_1 and x_3 chosen in the plane of the plate so to eliminate the mixed derivatives, the following form :

$$M'_n \frac{\partial^2 \tilde{u}_1^n}{\partial x_1^2} + P'_n \frac{\partial^2 \tilde{u}_1^n}{\partial x_3^2} - \frac{n^2\pi^2 C^*}{4h^2}\tilde{u}_1^n + \rho\omega^2\tilde{u}_1^n = \rho\omega^2(-1)^{(n-1)}/2\,\frac{e_{26}4V_0e^{j\omega t}}{C^{(1)}n^2\pi^2}\{1\}$$

In this equation, $\tilde{u}_1^n(x_1,x_3,t)$ is the lateral dependence of $\hat{u}_1^n(x_1,x_2,x_3,t) = \tilde{u}_1^n \sin(n\pi x_2/2h)$. $\tilde{u}_1$ arises from a transformation of the usual component of the displacement u_1 and of the potential Φ, made to replace the inhomogeneous boundary condition $\Phi = \pm V_0 e^{j\omega t}/2 at x_2 = \pm h$ by the homogeneous conditions $\hat{\Phi} = 0$. The inhomogeneous term of the equation results from this transformation. C* is either $\overline{C}^{(1)}$ for the un-electroded part of the resonator or $\hat{C}^{(1)}$ for the electroded part. $\overline{C}^{(1)}$ Is the stiffened elastic constant relative to the corresponding one dimensional mode (Eigen-value of the Christoffel Matrix).

$\hat{C}^{(1)}$ is a constant that includes the electrical and mechanical effects of the metallization. $4k^{(1)2}/n^2\pi^2$ is the relative frequency lowering due to the electrical effect of the metallization ; $\hat{R}$ = 2 ρ'h'/ρh is the relative frequency lowering due to the inertia of the electrodes. $C^{(1)} = \overline{C}^{(1)}(1\text{-}k^{(1)2})$ is a pseudo ordinary elastic constant. $k^{(1)}$ is the coupling coefficient of the corresponding one dimensional mode. M'_n, P'_n are intricate functions of the material constants and of the plate orientation [10].

Corresponding boundary and continuity conditions: The traction free conditions on the major surfaces, normal to the thickness are automatically verified as a consequence of the method used to establish the equations. The degree of approximation obtained for the contoured region of a resonator is probably somehow lower than for plane regions but for similar curvatures it is the same for spherical or non spherical contours.

On the surface limiting regions with different curvature or limiting the electroded and the un-electroded regions we have to specify the continuity of u_1 and of its normal derivative (approximation of the continuity of $T_{ij}n_j$).

On the external edges of the resonators, the approximate boundary condition to be applied is not uniquely determined and we have to choose in each case what type of mechanical condition, compatible with the scalar equation is the best approximation of the traction free conditions. Usually, for well designed plane or plano-convexe resonators, this choice is of small importance since both u_1 and its normal derivative are extremely small near the edge of the plate, so that the energy in this region is also extremely small. In the experimental study we have observed that this energy condition is not so well fulfilled for many of the bevelled resonators. The experimental results seems to show that the most appropriate condition, far from the fixation, is the vanishing of the normal derivative and not of the displacement; where as the condition u1=0 may be more appropriate near a point of rigid fixation.

II.2 Eigen solutions for regions having a non spherical contour.

The eigen mode at V=0 (short-circuit) are the solutions of the homogeneous form of the Tiersten equations:

$$M'_n \frac{\partial^2 u_1^n}{\partial x_1^2} + P'_n \frac{\partial^2 u_1^n}{\partial x_3^2} - \frac{n^2\pi^2 c^*}{4h^2}\bar{u}_1^n + \rho\omega^2 \bar{u}_1^n = 0 \qquad \{2\}$$

As previously, c* is either $\bar{c}^{(1)}$ for an un-electroded region or $\hat{c}^{(1)}$ for a short circuited electroded region. $\bar{u}_1$ Is identical to u_1 in this case. A simple derivation allows to introduce the corresponding characteristic (cut-off) frequencies in the equations:

$$M'_n u_{1,11}^n + P'_n u_{1,33}^n + 4\pi^2\rho\left(f^2 - f^{*2}\left(1 + \frac{x_1^2}{2R_1 h'_o} + \frac{x_3^2}{2R_3 h'_o}\right)\right)u_1^n = 0 \quad \{3\}$$

Where:

$f^* = f_{ce} = \frac{n}{4h_o}\sqrt{\frac{\hat{c}^{(1)}}{\rho}}$ for the electroded region, or:

$f^* = f_{cl} = \frac{n}{4h_o}\sqrt{\frac{\bar{c}^{(1)}}{\rho}}$ for the un-electroded regions.

As for the spherical contours for which the thickness variations are accounted for by separable approximations like:

$$\left(\frac{1}{2h}\right)^2 = \left(\frac{1}{2h_o}\right)^2\left(1 + \frac{x_1^2 + x_3^2}{2R_c h_o}\right),$$

We use a separable approximation of the same nature for the non-spherical contours having the same symmetry as the equation:

$$\left(\frac{1}{2h}\right)^2 = \left(\frac{1}{2h'_o}\right)^2\left(1 + \frac{x_1^2}{2R_1 h'_o} + \frac{x_3^2}{2R_3 h'_o}\right)$$

The contours approximated by this relation can be quadrics such as elliptical paraboloids or ellipsoids or surfaces with a higher degree such as those defined by the latter relation. We have previously observed [17] that when their two principal curvatures are in the same ratio as the two first coefficients of the equations, (ie $R_1/R_3=M'_n/P'_n$) it is possible to transform equations {2} into an isotropic Helmoltz homogeneous equation.

In facts, the coordinate transformations :

$$X_1 = \left(\frac{C^*}{M'_n}\right)^{1/2}.\gamma x_1 = r\cos t, \qquad X_3 = r\sin t = \left(\frac{C^*}{P'_n}\right)^{1/2}.\gamma x_3$$

where: $\gamma_c^4 = \frac{n^2\pi^2}{2h^3 R'_c}$,

and: $R'_c = R_1\left(\frac{C^*}{M'_n}\right) = R_3\left(\frac{C^*}{P'_n}\right)$

leads to the isotropic equation:

$$u_{1,rr} + \frac{1}{r}u_{1,r} + \frac{1}{r^2}u_{1,tt} + \left(A - \frac{r^2}{2}\right)u_1 = 0 \qquad \{4\}$$

With:

$$A = \left(\frac{n\pi}{2}\right)\left(\frac{R'_c}{2h_o}\right)^{1/2}\left(\frac{f^2 - f^{*2}}{f^{*2}}\right)$$

The coefficient A has a different expression in electroded or un-electroded regions and also between regions with different thickness or curvatures. *It is >0 in the electroded regions, and <0 in the un-electroded regions* .

Equation {4} can be separated as $u_1(r,t) = R(r).T(t)$ into two ordinary differential equations:

$$T'' + \mu^2 T = 0 \qquad \{5\}$$

$$R'' + \frac{1}{r}R' + R\left\{A - \frac{r^2}{4} - \frac{\mu^2}{r^2}\right\} = 0 \qquad \{6\}$$

The separation constant must be an integer $\mu=m$, since the solution of the first equation has to be periodic with a period $2m\pi$

The transformation: $R(r)=(1/r)\,W(r^2/2)$ lead to the equation:

$$W'' + (-1/4 + \frac{A/2}{r^2/2} + \frac{1/4 - m^2/4}{(r^2/2)^2})W = 0 \qquad \{7\}$$

which is confluent hypergeometric equation under the form given by Whittaker [25]. Among the solutions of this equation [26] the Whittaker $M_{A/2,m/2}(r^2/2)$, and $W_{A/2,m/2}(r^2/2)$ functions of index k=A/2, and $\mu = \pm m/2$ are the most interesting ones for our purposes , since the M functions are bounded at zero and the W functions tends towards zero when the argument becomes infinite. They are very appropriate respectively for the internal region (centre) and for the most external region (particularly in the case of "infinite" plate condition).

The properties of the R(r) solutions ($(1/r).M_{A/2,m/2}(r^2/2)$, and $(1/r).W_{A/2,m/2}(r^2/2)$) are very interesting for m=0 since then both functions can be finite at zero [$(1/r).W_{A/2,m/2}(r^2/2)$ is finite for all finite A>0, and for most A<0].

II.3 Case of the non-spherical plano-convexe resonators.

For a plano-convexe non spherical resonator with $R_1/R_3=M'_n/P'_n$, the nth overtone itself and its anharmonics that can be piezoelectrically excited by circular electrodes (in r,t coordinates, elliptical in the x1 and x3 ones) have a revolution symmetry (in the transformed coordinates) and zero or several circular nodal lines. The expressions of the displacement of

these eigen-modes at V=0 normalized so that u_1=1.0 at x_1=0,x_3=0., Are:

$$u_{1int}=\sqrt{2}.\left(\frac{1}{r}\right).M_{A_e/2,0}(r^2/2)$$

$$u_{1ext}=B_0.\left(\frac{1}{r}\right).M_{A_i/2,0}(r^2/2)+C_0.\left(\frac{1}{r}\right).W_{A_i/2,0}(r^2/2)$$

where: u_{1int} and u_{1ext} are respectively the displacement in the electroded and un-electroded regions. The values of A_e, A_i and hence the resonance frequencies $f_r^{(n)}$ of these mode, and the corresponding coefficients B_0 and C_0 are determined by the continuity conditions at the electrode boundary and the boundary conditions at the edge of the plate.

An "infinite plate" solution (condition u=0 at "infinity") can be constructed similarly by omitting the M function) in the expression of u_{1ext}.

The symmetrical or anti-symmetrical eigen-modes having radial nodal lines (they are not excited by symmetrical electrodes respecting the lateral anisotropy of the plates such as, for example, the circular ones in r,t coordinates)) are obtained in using the M and W functions with a non zero second index m/2 (respectively m even or m odd) and a trigonometric function with the appropriate symmetry (cos mt, or sin mt).

II.4. Most general complete solutions.

The most general symmetrical solutions (with respect to the x1 and to the x3 axis) bounded at r=0, is constituted of a linear combination of the successive order symmetrical solutions of equation {7}. It can be decomposed in two sub-expressions: a general solution bounded at zero:

$$u_1=\sum_{m=0}^{\infty} B_m\left(\frac{1}{r}\right)M_{A/2,m/2}\left(\frac{r^2}{2}\right)cosmt$$

where m is an even integer.

And a general symmetrical solution bounded at r ∞:

$$u_1=\sum_{m=0}^{\infty} B_m\left(\frac{1}{r}\right)W_{A/2,m/2}\left(\frac{r^2}{2}\right)cosmt$$

where also m is an even integer.

The first one is useful for the central region of a (non-spherical) contoured resonator, while it is the sum of the two sub-expression (the complete symmetrical general solution) that should be used for finding the forced modes for the external regions of partially .or totally contoured resonators (non-spherical contours) having electrodes of arbitrary geometry. These solutions are also those that should be used for the mode matching technique in the case of resonators with spherical contours.

Similarly general solutions, presenting an antisymmetry with respect to one coordinate axis, either the x1 or the x3 axis can be built using m odd and sin mt or cos mt.

II.5 Properties of these solutions.

1/For the same geometries, other eigen solutions can be found [17] for the n^{th} overtone of non-spherical contoured regions, as: $u_1=P_{a1}(X_1).P_{a3}(X_3)$

Where the P are parabolic cylinder functions of index a_1 and a_3. This indicates that (1/r) $M_{A/2,m/2}(r^2/2)$, and (1/r) $W_{A/2,m/2}$ $(r^2/2)$ with m=0 behave as the product of two parabolic cylinder functions. Moreover, there is a very simple relation between the index A/2 and the index a_1,a_3 (-A= a_1+a_3) of the parabolic functions. The Gauss-Hermite solutions ("infinite" totally electroded plates) which are particular cases of the previous one, corresponding to -a1-1/2=p, -a3-1/2=q (p and q integers) can also be represented by solutions of equation {3} (the absolute value of the first index A/2 is then a half integer since A=p+q+1). It appears thus that the parabolic cylinder functions which are also the solutions in the theory of spherical plano-convexe resonators (Solutions using the Hermite, the Gauss-Hermite or other parabolic cylinder functions.) can be expressed using the Whittaker functions, but then the relation between A and the index of the parabolic cylinder functions is a little different). This, constitute a first, although some how approximate justification of the technique we will employed for the spherical contours. The full justification should also consider the process by which the actual resonator geometry is approximated.

2/ As R'c (and hence A/2) tends toward infinity, it can be demonstrated starting from the homogeneous form of the Tiesten equation or from the properties of the Whittaker functions [26], that the radial part of the solutions tends towards:

$$H_m(r\sqrt{\mp A'})$$

where Hm is a Bessel Function or modified Bessel function (Jm, Ym, Km, Im) depending on the sign of A' and the behaviour of the solution at zero and/or infinity).

with : $$A'=\frac{\pi}{2}\left(\frac{f^2-f^{*2}}{f^{*2}}\right)$$

II.6 Example of behaviour of the solutions using the Whittaker functions.

Many features of the behaviour of the non spherical plano-convexe resonators (and also, more approximately of other types of plano-convexe and of partially or totally contoured resonators) can be determined by the observation of the behaviour of the solutions, considering the previous remark that the Gauss-Hermite solutions constitute an already very good approximation in such cases (this imply that A is close of half an odd integer).

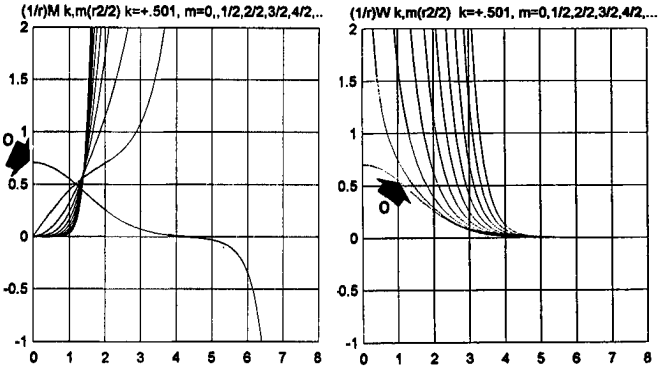
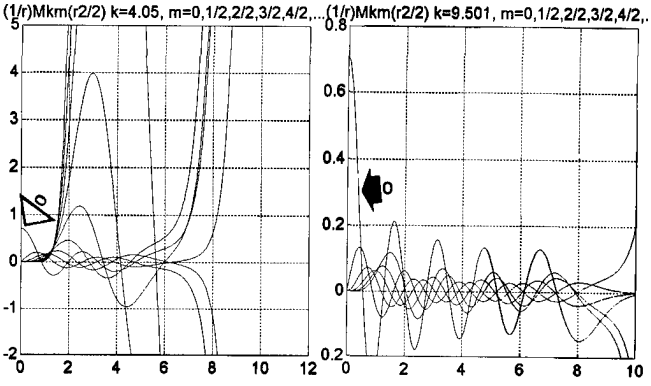


Figure 5 : Solutions of {7} for A/2=+.501

On figure 5a and 5b, we have represented an example of the variations of the functions (1/r) $M_{A/2,m/2}(r^2/2)$, and (1/r) $W_{A/2,m/2}$ $(r^2/2)$, as a function of r, for a positive value of A/2=0.501 and for values of m/2=0,1/2, 2/2, 3/2,... The variation of (1/r)$M_{A/2,0}$ $(r^2/2)$, (and of (1/r) $W_{A/2,0}$ $(r^2/2)$) are then quite close of those existing in the electroded region of a plano-convex resonator near the nth overtone mode. We can notice that (1/r) $W_{A/2,0}$ $(r^2/2)$ is finite at zero (at least for A/2>0), this could be very interesting for other theories of the plano-convexe resonators (the product of two parabolic cylinder functions with non integer index could be finite at zero and infinity).

On figure 6, the variations of (1/r) $M_{A/2,m/2}(r^2/2)$, are represented for a value of A/2=4.05, and half integer values of the second index ranging from 0 to 9/2. This case that could correspond to an anharmonic of a bevelled resonator indicate that the solutions could have an oscillatory behaviour. This is still more observable for a larger values of A/2=9.501 (m/2=0,1/2,...5/2) on figure 7. On the whole, it can be observed that the (1/r) $M_{A/2,m/2}(r^2/2)$ function has, like the series solutions y_1 [26] of the parabolic cylinder equation, a very rapidly varying behaviour with small change of either the positive indexes, or the argument.



Figures 6,7: Solutions of {7} for A/2=+4.05;9.501

On Figure 8a & 8b we have represented the variations of (1/r) $M_{A/2,m/2}(r^2/2)$, and (1/r) $W_{A/2,m/2}$ $(r^2/2)$, for a negative value of A (A= -0.501 corresponding approximately to the behaviour of the displacement in the non electroded region of a plano-convexe resonator near an overtone). The (1/r) $W_{A/2,0}$ $(r^2/2)$ function decreases rapidly but since the total displacement is a linear combination of (1/r) $M_{A/2,m/2}(r^2/2)$, and (1/r) $W_{A/2,m/2}$ $(r^2/2)$ functions, the behaviour of the (1/r) $M_{A/2,m/2}$ $(r^2/2)$ functions can explain that for some non optimized designs the displacement may not decrease uniformely to 0 in the unelectroded contoured external region of a resonator. This explains also the observation made for bevelled resonators of cases where the displacement has one minimum in the bevel region.

For negative value of A/2, the behaviours of the (1/r) $M_{A/2,m/2}(r^2/2)$, and (1/r) $W_{A/2,m/2}$ $(r^2/2)$ function are less sensitive to variations of the two indexes A/2 and m/2, this can be observed in comparing figures 8a & 8b with figures 9a and 9b corresponding to A/2=-1.501.

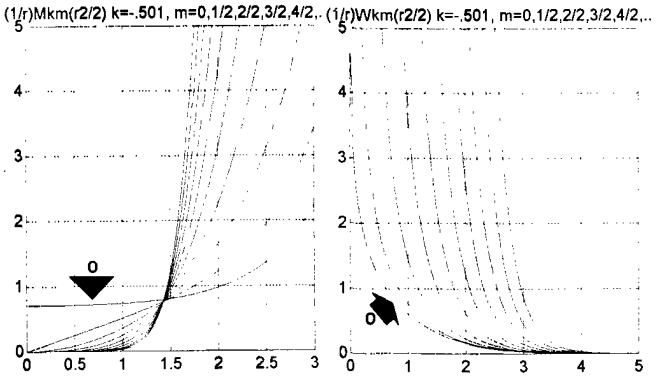


Figure 8a,b: Solutions of {7} for A/2=-.501

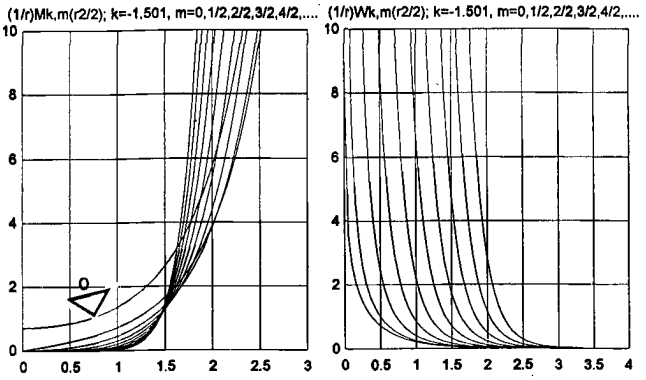


Figure 9a,b: Solutions of {7} for A/2=-1.501

III SOLUTIONS FOR THE PLANE REGIONS

We consider now the eigen-modes at V=0 existing in the plane region of partially contoured resonators. They are the solution of the homogeneous form of the Tiesten equation.

$$M'_n u_1^n, 11 + P'_n u_1^n, 33 + 4\pi^2 \rho (f^2 - f^{*2}) u_1^n = 0 \qquad \{8\}$$

The coordinate transformations :

$$X_1 = \frac{C^{*1/2}}{M'_n}.\gamma.x_1 = r\cos(t) \qquad X_3 = \frac{C^{*1/2}}{P'_n}.\gamma.x_3 = r\sin(t)$$

with: $\gamma^4 = \frac{n^2\pi^2}{4h^4}$

leads to the isotropic partial derivative equation:

$$u_{1,rr} + \frac{1}{r}u_{1,r} + \frac{1}{r^2}u_{1,u} + A u_1 = 0 \qquad \{9\}$$

Where: $A' = \frac{n\pi}{2}\left(\frac{f^2 - f^{*2}}{f^{*2}}\right)$

Equation {9} can be separated as $u_1(r,t) = R(r).T(t)$ in :

$$r^2 R'' + rR' + R(r^2 A - \upsilon^2) = 0 \qquad \{10\}$$

$$T'' + \nu^2 T = 0 \qquad \{11\}$$

with ν=m integer

The general symmetrical solution, with respect to the x_1 and x_3 axis, is, for a plane central electroded region, when the frequency is situated between f_{ce} and f_{cl}:

$$u_1 = \sum_{m=0}^{\infty} B_m J_m(r\sqrt{A'}).\cos(mt) \quad \text{(m even)}$$

and when $f>f_{cl}$:

$$u_1 = \sum_{m=0}^{\infty} C_m I_m(r\sqrt{-A'}).\cos(mt) \quad \text{(m even)}$$

The anti-symmetrical solutions can be similarly constructed using m odd and the appropriate trigonometric functions of mt

III.1 Principle of the solution for a bevelled resonators

For a bevelled resonator having a plane totally metallized region and an un-metallized non spherical bevel such that $R_1/R_3 = M'_n/P'_n$, the symmetrical eigen-modes at V=0, for the n_{th} overtone, normalized so that $u_1=1.0$ at (x1=x3=0.), Are res-

pectively the electroded and un-electroded regions given by the expressions (modes below f_{cl}):

$$u_{1int}=J_0\left(r\sqrt{A'_e}\right)$$

$$u_{1ext}=B_0.\left(\frac{1}{r}\right).M_{A_1/2,0}(r^2/2)+C_0.\left(\frac{1}{r}\right).W_{A_1/2,0}(r^2/2)$$

where: A'_e, A_1 and hence the resonance frequency $f_r^{(n)}$, and the coefficients B_0 and C_0 are determined by the (two) continuity conditions at the ("elliptical" in x1,x3 coordinates, circular in the r,t coordinates) electrode boundary and by the boundary conditions at the (elliptical) edge of the plate. In this case it may be very difficult to obtain simultaneously an extremely small amplitude and a good approximation of the cancellation of the normal derivative of the displacement at the outer edge of the resonator (see the variations of the M functions for the un-electroded region) unless some particular values of the parameter A are obtained. This constitute in fact the design challenge for bevelled resonators.

An "infinite" plate solution can be obtained in retaining only the W function in u_{1ext}. The study of this solution can be very useful to find good designs.

IV A MODEL FOR THE RESONATORS WITH SPHERICAL CONTOURS.

IV.1 Principle:

The solution obtained for the non spherical contours verify an equation which is a good approximation of the corresponding one for spherical contours provided that the both geometry are close (parameters $2h_o$, R_c and $2h_o'$, R_1, R_3). They verify with similar degrees of approximation the same boundary condition on the major surface of the plate (same thickness dependence of the displacement). These solutions can be made to verify, at least at a discrete number of points, the exact continuity conditions or boundary condition at the actual limits of the regions existing in the resonator with a spherical contour.

IV.2 Approximation of the geometry. The non-spherical contours with curvature respecting the lateral anisotropy are naturally an approximation to the spherical ones provide there are appropriate relations between their respective curvatures and thickness. This type of approximation can extremely good if the anisotropy is small as for example for the third overtone of the AT cut of quartz. For example for a central contoured region, a global approximation could be $Rc^2=R_1*R_3$ (or equivalently $Rc^2=R'c.((M'_n.P'_n)^{1/2})/(c^*)$) with the same thickness at the center. In this approximation. Another approximation could be to use the arithmetic mean instead of the geometrical one $Rc=(R_1+R_3)/2$.

However, other approximations can be found in considering the theory of the two types of resonators. As an example, from the models using the Gauss-Hermite approximation for both kinds of resonators it is possible to find [17] a relation between their respective curvatures and thickness parameters permitting to have the same resonance frequency for a given anharmonic of a given overtone, for both types of resonators. This global approximation is particularly adapted to the plano-convex resonators but it give also good results for the bevelled ones (see paragraph V).

A more sophisticated strategy is to obtain a piece-wise approximation of the geometry, in relation with the discretization of the boundary condition. In this case the ratio $(R'_c/2h'_o)^{1/2}$ appearing in the A parameter is varied in each region, by sectors, to obtained the exact value for the spherical contour at a point situated at a limit of the considered region, where the continuity or boundary condition are expressed. In certain cases (external contoured region, bevel), it can be more interesting to chose a value of $(R'_c/2h'_o)^{1/2}$ that gives a good approximation of the curvature between the two circular limits of a region. In this case, a piece-wise un-spherical solution with a $(R'_c/2h'_o)^{1/2}$ parameter (due to thickness and curvature adjustments), varying by sector is used (figure 10). We believe that, the piece-wise solutions could then verify very nearly the exact homogeneous equation for the spherical contour.

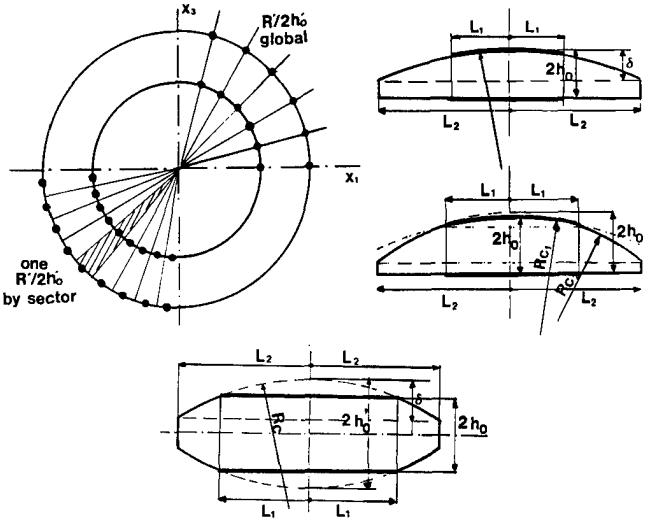


Figure 10: Geometrie of the contoured resonators and discretization used in the model.

IV.3 Expression of the continuity and boundary conditions.

At the limit of two regions (electroded and un-electroded regions, zones with different curvatures,...) we have to express the continuity of the lateral part of the displacement, and of its normal derivative. At the resonator edge we have to express a boundary condition. The experiments have permitted to obtain some informations about the choice of the most adapted conditions (see §I). As proposed originally by Nakamura [27], we express these conditions at a discrete number of point along the limits of the regions and at the edge of the resonator. This leads to truncate the series to a number of terms depending on the number of points at which the conditions are expressed. If we chose p points at each limit, the total number of terms to retain is 3*p since we have 3conditions).Different strategies can be chosen for the distribution of the number of terms of the series between the regions, we have taken more terms for the central region.

VI.4 Forced modes and equivalent scheme.

The determination of the forced modes and of the equivalent scheme follows very exactly the development already given for the planes resonators in reference [28].

IV.4 Numerical Implementation.

This model was implemented as a computer program considering all the geometries of partially or totally contoured

resonators having two regions with the same or different curvatures (including the plane case). The central region is supposed to be electroded. The program is intended to permit a good numerical accuracy. Its principle is described in figure 11.

V APPLICATION TO BEVELLED RESONATORS.

The program was used to compute the resonances frequencies and the mode shapes of bevelled resonators. Different hypothesis were tested for the boundary conditions at the edge of the plate: Diriclet (zero displacement), "infinite" plate (only W functions in the expression of u_{1ext}), Neumann (zero normal derivative).

In a first step we have studied the convergence of the solution as the number of points at which the continuity and boundary conditions are expressed. The results in the case of "infinite" plate conditions are given in table I. For these calculation an equal number of point was chosen at each boundary, an the number given in table I is the number at each limit. In this case an equal number of terms was chosen for the series relatives to the two regions. The calculation were made in the case of resonators very similar to those represented in figure 2, but with a slightly different curvature. An extremely fast semi-alternate convergence of the solution was found either for the fundamental mode and the first anharmonic. The semi-alternate convergence gives easily an idea of the numerical accuracy achieved (this is not the absolute accuracy of the model).

Table I: Convergence properties of the solution for a bevelled resonator.

2ho=0.711, Rc=66.mm 2L1=9.mm 2L2=14.mm

N Points	Frequency fundamental	Frequency 1st anharm.
3	2350300.3700	2406086.5737
4	2350299.4692	2406086.2874
5	2350299.7048	2406086.2657
6	2450299.6418	2406086.2880
7	2350299.6591	2406086.2778
8	2350299.6548	2406086.2818
9	2350299.6554	2406086.2802
10	2350299.6552	-------

for a plane resonator (same param.) fr=2333.347kHz

Table II: Comparison of experimental and computed results.

EXPERIMENTAL AND COMPUTED RESULTS

2ho mm	Rc mm	2L1 mm	2L2 mm	Fr (exper.) Hz	Fr (calcul.) Hz
0.7110	60.0	9.5	14.0	2353750.	2348664.
0.4120	60.0	9.8	13.0	4028800.	4025298.
0.7968	100.0	16.0	22.0	2086900.	2087959.

* 6 points [Rc'(equal frequencies for plano_convexe)]

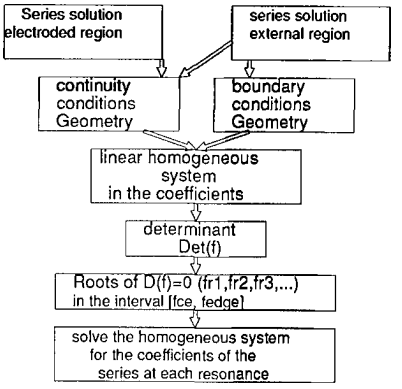


Figure 11: Principle of the numerical implementation of the model.

To obtain a more precise idea of the model accuracy, the resonance frequencies and modes of several of the resonators studied in § I were computed. The results, for the resonance frequencies, obtained with the same "boundary" condition as in table I are represented in table II together with the parameters of the resonators. It can be observed that the agreement with the experimental frequencies is compatible with the accuracy of the determination of the thickness and mass loading (a few thousandth).

Table III: Influence of the boundary conditions on the computed frequencies.

INFLUENCE OF THE BOUNDARY CONDITIONS

DEVICE	dUI/Dn=0	Infinite	Zero UI
normal	2343057.25	2348664.17	2352393.98
thin plate	4023594.21	4025297.99	4026329.83
very large plate	2086443.57	2087959.18	2088866.50

In table III we compare for the same resonators, the frequencies computed for the three types of boundary conditions. It can be observed that the Neumann conditions give the lowest frequencies whereas the Diriclet ones give the highest. We can also observe that the "infinite" plate resonance frequency is always situated between these two latter so that it constitute a very good approximation. The relative differences between them is of the same order of the accuracy on the experimental determination of the parameter so that we cannot conclude about the best choice. If we compare to previous results concerning the influence of boundary conditions on the computed frequencies for high performance plano-convexe resonators using 3rd or 5th overtones modes [16][17], we can observe that in this case of bevelled resonators using the fundamental mode this influence is much greater by about three orders of magnitude. This can allow to detect on bevelled devices, by calculations as well as by the experiments (this was already the case for the results given in §I) many phenomena that could limit the qualities of the presently most performing resonators.

The mode shapes computed using the geometrical approximation taking R'_c such that the ellipsoidal resonators with parameter $(2h_o, R'_c)$ have the same resonance frequency as a spherical one with parameters $(2h_o, R_c)$ for the resonator shown

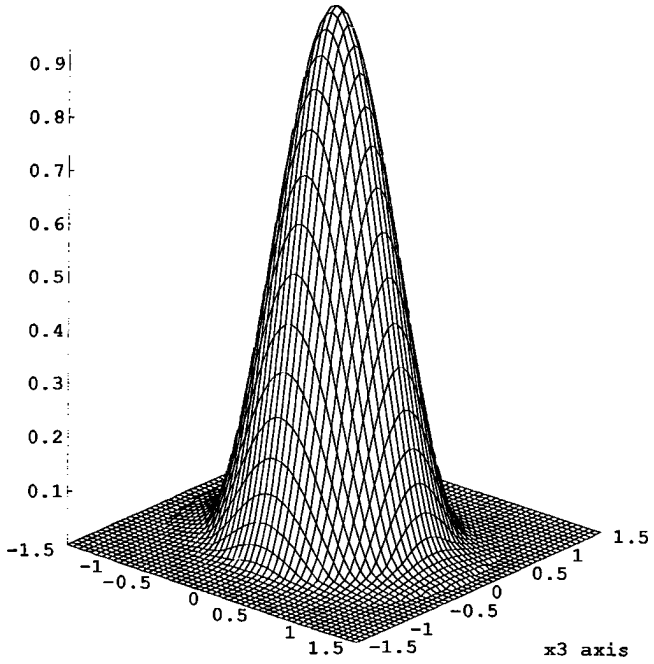


Figure 12: Computed mode shape (u12) for the resonator of figure 4a.

These difference may be due to varied responses in each case of boundary condition, to the discretization and to the truncation of the series. We believe that the discretization underline a real phenomena resulting of the abrupt change in curvature at the boundary of the two regions.

In Figure 14 we consider the case of the resonator of figure 2 with Diriclet conditions at the edge of the plate: The irregularities observed in figures 2a and 2b are well reproduced. Although they are most probably induced by the discretization (4x5 points at each region limit) they reveal, more than in the preceding figure "instabilities" of the mode shape, to perturbations (may be the fixation in the experimental case). As indicated in §I, similar or more pronounced irregularities were very often observed in the experimental study; their origin has been identified only in certain cases (crystal defects such as growth band or dislocation bundles or geometrical imperfections such as defects of circularity of the electrodes, off-centering of the bevel etc....). It seems that it exists designs which are more or less sensitive to such perturbations. These perturbations were never observed in calculations using the "infinite" plate condition. This situate their origin in the boundary conditions at the edge of the plate and also in the terms containing the M Whittaker functions and in their "sensitivity" to small change of the parameters.

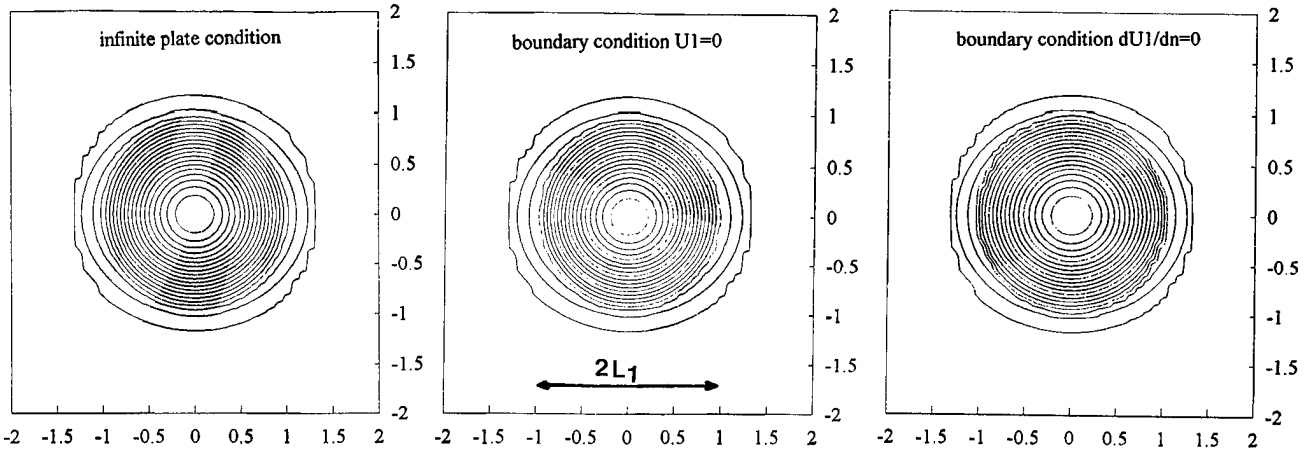


Figure 13: Contour plot of the displacement as a function of the boundary condition (12a "infinite" plate, 12b Diriclet, 12c Neumann).

in figure 4. The square of the displacement for the "infinite" plate condition is drawn in figure 12. In this figure, as in the following, the lateral coordinates are normalized by the value of the radius of the central region (L_1). The agreement with the observed mode seem to be good. As in figure 4a, we can observe that the mode begins to decrease significantly in the central region. The anisotropy of the mode in the contoured region seems to be some how more pronounced for the computed results than in figure 4a. This may result of the geometrical approximation chosen.

In figure 13, contour plots of the displacement computed with the three types of boundary conditions are compared for the same resonator. We can observe that the displacement is quasi-identical for the three cases in the central region, where as slight differences can be observed in the external regions.

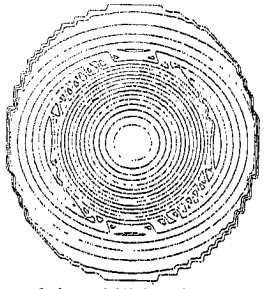


Figure 14: Computed instabilities in the mode shape of a bevelled resonator.

(to be compared with those observed experimentally in figure 2).

CONCLUSION

The experimental study has shown several features of the bevelled resonators. Most of them present energy in the bevel which is often not uniformely decreasing to the edge of the plate.

The theory of partially or totally contoured resonators with contours respecting the lateral anisotropy of crystal plates leads to particularly simple results that are very indicative of the behaviours of the corresponding devices with spherical contours. Further new results could be expected in this direction.

The solutions found in this theory are of a very great generality, they contain as particular case nearly all the solutions already found for all types of energy trapped thickness shear resonators.

The fact that in the formalism used (or for the functional basis considered), the solutions for the non-spherical resonators are the 'minimal' solutions among all existing for other geometries (particularly the spherical one) contains very important consequences that are of the same nature as those found in the case of the plane resonators with elliptical electrodes and plate having an aspect ratio respecting the lateral anisotropy of the plate. Non-spherical contours may constitute the optimum solution to obtain an important improvement of the characteristics of the most performing resonators.

The solutions found for the un-spherical resonators lead to very accurate models for the spherical case provided the displacement is essentially a u_1 shear, and a strategy for approximating the spherical geometry is used.

The application of the model to bevelled resonators has confirmed the possibilities of the method. A very fast semi alternate convergence as a function of the discretization was found which allows to measure the numerical accuracy ($<1.10^{-10}$ for the frequencies). The computed frequencies and modes are in good agreement with the experiment. Several features observed in the experiments could be explained by computed results.

A systematic evaluation of this model is being made for different cases of bevelled and other contoured resonators with the objective to find optimum design rules for resonators which are most often still designed in a very empirical manner.

We believe that an in depth study of the bevelled resonators can bring a new knowledge of all other types of contoured resonators with spherical contours, and particularly, of those used to generate the most stable frequencies. This results of the fact that the phenomena which could limit the performances of the latter, are much more intense in bevelled resonators.

Acknowledgements : The authors wish to acknowledge the skillful assistance of G.Caumont, G.LeTuault and A. Daniel, a for the preparation of the samples.

REFERENCES

[1] R.D.Mindlin Quartely Appl.Mathem. v.XIX n°1 pp.51-61 (1961).

[2] R.D. Mindlin Int.J.Solids.Structures v.20 n°2 pp.141-157 (1984).

[3] P.C.Y. Lee, S.Syngellakis, J.P.Hou. J. Appl. Phys. v.61 n°4 pp.1249-1262 (1987).

[4] H.F.Tiesten. Linear Piezoelectric plates vibrations; Plenum Press (1969).

[5] R.C. PEACH IEEE Trans. Ultrason, Frroelec. Freq.Contr. v.35 n°5 pp.593-611 (1988).

[6] Y.K.Yong, Z.Zhang Proc. of the 1992 IEEE Int. Frequency Control Symposium
pp. 567-581 (1992).

[7] Y.K.Yong, Z.Yang IEEE Trans. Ultrason. Ferroelec. Freq. Contr. v.41 n°5 pp.685-693 (1994).

[8] H.F. Tiersten. J. Acoust. Soc. Amer. v.59 n°4 pp.879-883 (1976).

[9] H.F. Tiersten, R.C. Smythe. J. Acoust. Soc. Am. Vol. 65 n° 6 pp.1455-1460 (1979).

[10] D.S. Stevens , H.F. Tiersten. J. Acoust. Soc. Am. Vol. 79 n°6 pp.1826-1811 (1986).

[11] D.C.L. Vangheluwe. Proc. 34th Annual Frequency Control Symposium pp.412 (1980).

[12] B.K. Sinha, D.S.Stevens. J.Acoust. Soc. Am. v.66 n°1 pp.192-196 (1979).

[13] J. Détaint, J. Schwartzel, C. Joly, E. Philippot. Proc. 41st Annual Frequency Control Symp. p 314 (1987).

[14] H. Seikimoto IEEE Trans. Sonics Ultrason. v.SU-31 n°6 pp.664-669.(1984).

[15] R.C. Peach. Proc.of the 42nd Annual Frequency Control Symposium pp.38-44 (1988).

[16] B. Dulmet, H.F. Tiersten. Proc.1986 IEEE Ultrasonics Symposium. pp. 355-360, (1986).

[17] J.Détaint, H.Carru, J.Schwartzel, B.Capelle, A.Zarka; Proc. of the 42nd Annual Frequency Control Symposium pp.19-28 (1988)

[18] H.F.Tiesten, Y.S. Zhou. Proc. of the 1993 IEEE Int. Frequency Control Symposium pp.431-437 (1993).

[19] H.F.Tiersten, B.J.Lwo, B.Dulmet. pp.172-183 (1994)

[20] J.S.Yang, H.F.Tiersten. to appear in Proc. 1995 Int. IEEE Frequency Control Symposium (these proceedings)

[21] W.Zhang, J.Yang. Proc. of the 1993 IEEE Int. Frequency Control Symposium. pp.448-457 (1993).

[22] P.C.Y. Lee, Ji. Wang to appear in Proc. of the 1995 IEEE Int. Frequency Control Symposium. (These proceedinds).

[23] A. Zarka, B. Capelle, B. Zheng, J. Détaint, J. Schwartzel. Proc. of the 42nd Annual Frequency Control Symposium; pp. (1988)

[24] B.Capelle, Y. Zheng, A. Zarka, J. Détaint, J. Schwartzel; Proc. of the 1992 IEEE Int. Frequency Control Symposium. pp. 553-560. (1992).

[25] E.T. Whittaker ; G.N. Watson
A course of Modern Analysis Cambridge Press 1927

[26] M. Abramowitz ; I. Stegun
Handbook of Mathematical functions ; Dover (1972)

[27] K. Nakamura ; H. Shimizu; Proc. 1976 IEEE Ultrasonics Symp. p. 606 (1976)

[28] J.Détaint, H.Carru, J.Schartzel, B.Capelle, A.Zarka, Y.Zheng, E.Philippot; Proc.of the 45th Annual Frequency Control Symposium pp.166-180 (1991).

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

AN ANALYSIS OF CONTOURED QUARTZ RESONATORS WITH BEVELED CYLINDRICAL EDGES

J.S. Yang and H.F. Tiersten

Department of Mechanical Engineering, Aeronautical Engineering and Mechanics
Rensselaer Polytechnic Institute, Troy, New York 12180-3590

Abstract

The equation for transversely varying thickness modes in doubly-rotated quartz resonators is applied in the analysis of contoured resonators with beveled cylindrical edges. The coefficients appearing in the planar differential operator are written as a sum of a mean or isotropic part plus a deviation. Asymptotic eigensolutions for the nearby isotropic case are obtained for the contoured cylindrical beveled resonator. The resonant frequencies for the actual anisotropic case are obtained from an equation for the perturbation in eigenfrequency from the isotropic solution. A lumped parameter representation of the admittance, which is valid in the vicinity of a resonance, is obtained. Calculated results are presented for some contoured beveled SC-cut quartz resonators.

1. Introduction

The equation for transversely varying thickness modes[1] in doubly-rotated quartz resonators is applied in the analysis of the steady-state vibrations of contoured resonators with beveled cylindrical edges. The contoured inner region is electroded while the more sharply contoured outer region is not. First the planar differential equation for the nth order thickness modes is referred to the coordinate system in which no mixed derivatives occur. Then the coefficients appearing in the planar differential operator are written as a sum of a mean or isotropic part plus a deviation. The edges of the circular electrodes are taken to end at the junction between the outer and inner spherical contours.

The radial dependence of the eigensolution in the inner electroded region is obtained by the method of Froebinius and in the outer unelectroded region is obtained by means of an asymptotic expansion about a point just outside the edge of the electrode[2]. For this nearby isotropic case the transcendental frequency equation is obtained by satisfying the edge conditions between the inner and outer regions. The resonant frequencies for the actual anisotropic case are obtained from an equation for the perturbation in eigenfrequency from the isotropic solution. As in earlier work, and for the same reasons, it turns out that the first perturbation in eigenfrequency vanishes for all modes except those with one angular nodal plane. A lumped parameter representation of the admittance, which is valid in the vicinity of a resonance, is obtained. Calculated results are presented for a number of contoured, beveled SC-cut resonators and the influence of the curvature of the inner and outer contours on the mode shapes, resonant frequencies and motional capacitances is exhibited.

2. Basic Equations

A plan view and cross-section of the contoured, beveled cylindrical resonator is shown in Fig.1 along with the associated coordinate systems. The x_2-axis points up. The larger inner radius of curvature is denoted $\bar{R}$ and the smaller outer radius of curvature is denoted R. The dotted line denotes the surface that would exist if the outer radius of curvature existed over the entire plate. In Fig.1 the cylindrical coordinates r and θ are shown in addition to the particular Cartesian coordinates x_1' and x_3', which are defined in Eqs.(97) and (98) of Ref.1 and will be discussed further later on in this section. Note that θ is measured from the x_3'-axis.

It has been shown[3] that the differential equation for the dominant displacement $u_1^{(n)}$ of the nth harmonic family of transversely varying thickness modes in doubly-rotated quartz plates may be written in the form

$$M_n' \frac{\partial^2 u^n}{\partial x_1'^2} + P_n' \frac{\partial^2 u^n}{\partial x_3'^2} - \frac{n^2\pi^2}{4h^2}\hat{c}^{(1)}u^n - \rho\ddot{u}^n = \frac{\rho\omega^2(-1)^{\frac{n-1}{2}} e_{26} 4V e^{i\omega t}}{c^{(1)} n^2\pi^2}, \qquad (2.1)$$

where V is the driving voltage applied across surface electrodes and M_n' and P_n' are involved lengthy expressions that are defined in Sec.III of Ref.1 along with x_1' and x_3', which are orthogonal directions in the plane of the plate for which the scalar differential equation (2.1) does not contain mixed derivatives. In (2.1) we have employed the relations

0-7803-2500-1/95 $4.00 © 1995 IEEE

$$u_1^n = u^n(x_1', x_3', t)\sin\frac{n\pi x_2}{2h},$$
$$c^{(1)} = \bar{c}^{(1)}(1-k_1^2), \qquad (2.2)$$

$$\hat{c}^{(1)} = \bar{c}^{(1)}\left[1-\frac{8k_1^2}{n^2\pi^2}-2\hat{R}\right],$$

$$k_1^2 = \frac{e_{26}^2}{\bar{c}^{(1)}\varepsilon_{22}}, \quad \hat{R} = \frac{2\rho' h'}{\rho h}, \qquad (2.3)$$

where $\bar{c}^{(1)}$ is the eigenvalue for the pure thickness mode of the unelectroded plate with nonzero displacement $u_1^{(n)}, e_{26}$ is the piezoelectric constant referred to the thickness eigendisplacement direction $u_1^{(n)}, 2h, 2h'$ and ρ, ρ' are the thicknesses and mass densities of the resonator and electrodes, respectively. The pure thickness frequencies ω_n and $\bar{\omega}_n$ for the unelectroded and electroded regions of the plate, respectively, are given by

$$\omega_n^2 = n^2\pi^2\bar{c}^{(1)}/4h^2\rho, \quad \bar{\omega}_n^2 = n^2\pi^2\hat{c}^{(1)}/4\bar{h}^2\rho, \qquad (2.4)$$

where $2\bar{h}$ denotes the variable thickness of the contoured inner electroded region and 2h denotes the variable thickness in the contoured outer region near the edges.

Since the contoured electroded central region and contoured outer region are circular, we write Eq.(2.1) in the form

$$c_n\nabla^2 u^n + A_n - \rho\bar{\omega}_n^2 u^n + \rho\omega^2 u^n = F_n, \qquad (2.5)$$

where

$$A_n = a_n\left[\frac{\partial^2 u^n}{\partial x_1'^2} - \frac{\partial^2 u^n}{\partial x_3'^2}\right], \quad c_n = \frac{M_n' + P_n'}{2},$$
$$a_n = \frac{M_n' - P_n'}{2}, \qquad (2.6)$$

in which c_n is the mean or isotropic constant, a_n is the anisotropy constant and A_n is the anisotropy term. In the electroded central region we have

$$F_n = \frac{\rho\omega^2(-1)^{\frac{n-1}{2}}e_{26}4Ve^{i\omega t}}{c^{(1)}n^2\pi^2}, \quad \tilde{\omega} = \bar{\omega}_n, \qquad (2.7)$$

and in the unelectroded outer region we have

$$F_n = 0, \quad \tilde{\omega} = \omega_n. \qquad (2.8)$$

In cylindrical coordinates r, θ, ∇^2 takes the form

$$\nabla^2 = \frac{\partial^2}{\partial r} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}, \qquad (2.9)$$

and at the edge of the electrode we have the continuity conditions

$$\bar{u}^n = u^n, \quad \frac{\partial\bar{u}^n}{\partial r} = \frac{\partial u^n}{\partial r} \text{ at } r = \ell. \qquad (2.10)$$

Equations (2.5) - (2.10) define the problem for the anisotropic beveled resonator. However, before we discuss the solution, we note for later use that the variable thicknesses in the contoured inner and outer regions, respectively, are given by

$$2\bar{h} = 2\bar{h}_o\left[1-\frac{r^2}{4\bar{R}\bar{h}_o}\right], \quad 2h = 2h_o\left[1-\frac{r^2}{4Rh_o}\right]. \qquad (2.11)$$

From $(2.11)_2$ it is clear that at r = d, which is shown in Fig.1 and will be defined later, the constant thickness $2h_d$ is given by

$$2h_d = 2h_o\left[1-\frac{d^2}{4Rh_o}\right], \qquad (2.12)$$

and we now note that when R is the radius of a single contour, the radius of a double contour R' is given by R = $R'/2$.

In the next section we obtain the natural frequencies of the anisotropic contoured beveled cylindrical resonator by finding the perturbation in eigenfrequencies from the nearby isotropic plate as a result of the anisotropy. To this end we omit A_n in (2.5) for the case of free vibrations and write

$$c_n\nabla^2\bar{v}^\nu + \rho(\omega_\nu^2 - \bar{\omega}_n^2)\bar{v}^\nu = 0, \quad 0 < r < \ell,$$
$$c_n\nabla^2 v^\nu + \rho(\omega_\nu^2 - \omega_n^2)v^\nu = 0, \quad \ell < r, \qquad (2.13)$$

as the equations for the nearby unperturbed isotropic problem, in which $\bar{v}^\nu$ and v^ν are the unperturbed eigendisplacements in the electroded central region and outer region, respectively, and ω_ν is the unperturbed eigenfrequency of the nearby isotropic problem. At the interface between the two regions from (2.10), we have the continuity conditions

$$\bar{v} = v, \quad \frac{\partial\bar{v}}{\partial r} = \frac{\partial v}{\partial r}, \quad \text{at } r = \ell. \qquad (2.14)$$

The equation for the perturbation in eigenfrequency due to the anisotropy is derived at the end of Section II of Ref.2 and is given by

$$\omega = \omega_\nu - \Delta_\nu, \tag{2.15}$$

where ω is the perturbed frequency and

$$\Delta_\nu = H_\nu / 2\omega_\nu N_\nu^2, \quad H_\nu = \int_i \bar{A}_n \bar{v}^\nu dS + \int_o A_n v^\nu dS, \quad N_\nu^2 = \int_i \rho(\bar{v}^\nu)^2 dS + \int_o \rho(v^\nu)^2 dS, \quad A_n = a_n(v^\nu_{,11} - v^\nu_{,33}), \tag{2.16}$$

in which i represents the inner, and o the outer region.

3. Free Vibrations of Contoured and Beveled Cylindrical Plate

We first obtain the unperturbed solution to the isotropic equations (2.13) subject to the continuity conditions (2.14). To this end we substitute from $(2.11)_1$ and $(2.4)_2$ into $(2.13)_1$ and neglect terms higher than r^2 to obtain

$$\frac{\partial^2 \bar{v}}{\partial r^2} + \frac{1}{r}\frac{\partial \bar{v}}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \bar{v}}{\partial \theta^2} + \alpha_n^2 \bar{v} - \bar{\gamma}_n^2 r^2 \bar{v} = 0, \tag{3.1}$$

where

$$\alpha_n^2 = \frac{\rho}{c_n}(\omega_\nu^2 - \bar{\omega}_{no}^2), \quad \bar{\gamma}_n^2 = \frac{\rho \bar{\omega}_{no}^2}{c_n}\frac{1}{2\bar{R}\bar{h}_o}, \tag{3.2}$$

and

$$\varpi_{no}^2 = n^2\pi^2 \hat{c}^{(1)} / 4\bar{h}_o^2 \rho. \tag{3.3}$$

The solution satisfying (3.1) may be written in the form

$$\bar{v} = e^{im\theta} f(r), \quad m = 0,1,2,3,\ldots \tag{3.4}$$

the substitution of which in (3.1) yields

$$f'' + \frac{1}{r}f + (\alpha_n^2 - \frac{m^2}{r^2} - \bar{\gamma}_n^2 r^2) f = 0. \tag{3.5}$$

Although Eq.(3.5) may readily be transformed to a form satisfied by a Whittaker function[4], since we are interested in the solution which does not diverge at 0 for values of $r \le \ell$, we prefer to use the method of Froebinius directly because it yields the form we need in the most direct way. To this end we write

$$f(r) = \sum_{j=0}^{\infty} G_j r^{\mu+j}, \tag{3.6}$$

the substitution of which in (3.5) for $G_o \neq 0$ yields

$$\mu = m. \tag{3.7}$$

Equating coefficients of like powers of r to zero, we obtain

$$G_1 = 0, \quad G_2 = -\frac{\alpha_n^2}{4(1+m)} G_o, \quad G_3 = 0,$$
$$G_{j+4} = -\frac{\alpha_n^2 G_{j+2} + \bar{\gamma}_n^2 G_j}{(j+4)(j+4+2m)}, \quad j = 0,1,2,\ldots\ldots \tag{3.8}$$

For $G_o = 1$ let us denote the series by $\delta_m(r)$, then the solution of (3.5) that we need may be written

$$f = \bar{B}\delta_{(m)}(r), \tag{3.9}$$

where $\bar{B}$ is an arbitrary constant.

In the contoured outer region the solution decays very rapidly with distance from the edge of the electrode. Consequently, it is possible to make an expansion in $(2.13)_2$ about some distance d slightly larger than ℓ, say 1.1ℓ, by writing[2]

$$r = d + \varsigma, \tag{3.10}$$

and retaining terms of order ς / d but not $(\varsigma / d)^2$ in $(2.13)_2$ and satisfying the resulting equation[5] to order ς / d. To this end we substitute (3.10) in $(2.11)_2$ and employ (2.12) to obtain

$$2h = 2h_d(1 - \varsigma d / 2h_d R), \tag{3.11}$$

which yields

$$\frac{1}{4h^2} = \frac{1}{4h_d^2}\left[1 + \frac{\varsigma d}{h_d R}\right]. \tag{3.12}$$

Substituting from (3.12) into $(2.13)_2$, we obtain

$$c_n \nabla^2 v - \rho\omega_{nd}^2\left[1 + \frac{\varsigma d}{h_d R}\right] v + \rho\omega_\nu^2 v = 0, \tag{3.13}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial \varsigma^2} + \frac{1}{d}\left[1 - \frac{\varsigma}{d}\right]\frac{\partial}{\partial \varsigma} + \frac{1}{d^2}\left[1 - \frac{2\varsigma}{d}\right]\frac{\partial^2}{\partial \theta^2},$$

$$\omega_{nd}^2 = \frac{n^2\pi^2 \bar{c}^{(1)}}{4h_d^2 \rho}, \tag{3.14}$$

in which $(3.14)_1$ is obtained by employing (3.10) in (2.9) and retaining terms of order ς / d. The solution satisfying (3.13)

may be written in the form

$$v = w(\varsigma)e^{im\theta}.\, m = 0,1,2,3\ldots, \qquad (3.15)$$

the substitution of which in (3.13) yields

$$\frac{\partial^2 w}{\partial \varsigma^2} + \frac{1}{d}\left[1 - \frac{\varsigma}{d}\right]\frac{\partial w}{\partial \varsigma} - \frac{m^2}{d^2}\left[1 - \frac{2\varsigma}{d}\right]w$$

$$-\gamma_n^2 \frac{\varsigma d}{h_d R}w + \beta_n^2 w = 0, \qquad (3.16)$$

where

$$\gamma_n^2 = \rho w_{nd}^2 / c_n, \quad \beta_n^2 = (\rho / c_n)(\omega_v^2 - \omega_{nd}^2). \qquad (3.17)$$

We write the asymptotic solution of (3.16) in the form

$$w = Be^{-k_o\varsigma - k_1\varsigma^2} \qquad (3.18)$$

in which the real parts of k_o and k_1 must be positive. Substituting from (3.18) into (3.16), we obtain

$$\left[k_o^2 - 2k_1 - \frac{k_o}{d} - \frac{m^2}{d^2} + \beta_n^2 + \left(4k_o k_1 - \frac{2k_1}{d} + \frac{k_o}{d^2} + \frac{2m^2}{d^3}\right.\right.$$

$$\left.\left. - \frac{\gamma_n^2}{h_d R}\right)\varsigma + \left(4k_1^2 + \frac{2k_1}{d^2}\right)\varsigma^2\theta - \right]Be^{-k_o\varsigma - k_1\varsigma^2} = 0, \qquad (3.19)$$

and we now note, in accordance with the earlier discussion, that ς^2/d^2 is negligible when $|w|$ is not negligible and that $|w| \approx 0$ when ς^2/d^2 is large. Accordingly, we ignore[6] the coefficient of the ς^2 term in (3.19) and take the constant term and the coefficient of the ς term to vanish, which after some manipulation yields

$$4k_o^3 - \frac{6k_o^2}{d} + 4\left(\frac{1}{d^2} - N\right)k_o - 2\left(S - \frac{N}{d}\right) = 0, \qquad (3.20)$$

where

$$N = \frac{m^2}{d^2} - \beta_n^2, \quad S = \frac{\gamma_n^2 d}{h_d R} - \frac{2m^2}{d^3}. \qquad (3.21)$$

Equation (3.20) is a cubic, which always yields one real positive root k_o and a corresponding positive k_1, from either of the two equations which led to (3.20).

At this point the solution consists of (3.4) with (3.9) in the inner electroded region and (3.15) with (3.18) in the outer region. The unperturbed isotropic solution must satisfy the edge conditions (2.14), which with (3.4), (3.15) and (3.10) enables us to write

$$w = f, \quad \partial w/\partial\varsigma = \partial f/\partial r \text{ at } r = \ell \quad (\varsigma = -g). \qquad (3.22)$$

The substitution of (3.9) and (3.18) in (3.22) yields

$$\overline{B}\delta_m(\ell) - Be^{k_o g - k_1 g^2} = 0,$$

$$\overline{B}\delta_m'(\ell) + B(k_o - 2k_1 g)e^{k_o g - k_1 g^2} = 0, \qquad (3.23)$$

which constitute two homogeneous linear algebraic equations in B and $\overline{B}$. For a non-trivial solution the determinant of the coefficients must vanish, which yields the transcendental equation

$$(k_o - 2k_1 g)\delta_m(\ell) + \delta_m'(\ell) = 0, \qquad (3.24)$$

since $e^{k_o g - k_1 g^2} \neq 0$. Satisfaction of (3.24) yields the unperturbed eigenfrequencies ω_{nmp}, where p represents the order of the root of (3.24).

In order to determine the perturbation in eigenfrequency resulting from the anisotropy we must evaluate H_v, A_n and N_v^2 in (2.16) as shown in Sec.III of Ref.2. In that way as for Eq.(3.25) of Ref.2, we obtain

$$H_{nmp} = a_n \int_o^{2\pi} d\theta \int_o^{\infty} r\,dr\left[\left(-\frac{\partial^2 w}{\partial r^2} + \frac{1}{r}\frac{\partial w}{\partial r} - m^2\frac{w}{r^2}\right)\right.$$

$$\left.\times \cos 2\theta\left[\begin{matrix}\cos & m\theta \\ \sin & m\theta\end{matrix}\right]\right]$$

$$\pm 2m\left(\frac{w}{r^2} - \frac{1}{r}\frac{\partial w}{\partial r}\right)\sin 2\theta\left[\begin{matrix}\sin & m\theta \\ \cos & m\theta\end{matrix}\right]w\left[\left[\begin{matrix}\cos & m\theta \\ \sin & m\theta\end{matrix}\right]\right], \qquad (3.25)$$

where from the solution in (3.9) and (3.18), we have

$$w = \overline{B}_{nmp}\delta_{nmp}(r), \quad 0 < r < \ell,$$

$$w = B_{nmp}e^{-k_o^{nmp}\varsigma - k_1^{nmp}\varsigma^2}, \ell < r, \qquad (3.26)$$

in which the upper terms in (3.25) correspond to the real part of $e^{im\theta}$ and the lower terms, to the imaginary part of $e^{im\theta}$ in the solution functions in (3.4) and (3.15). As a consequence of the periodicity of the trigonometric functions, evaluation of the angular integrals in (3.25) reveals that

$$H_{nmp} = 0 \text{ unless } m = 1, \qquad (3.27)$$

while for m = 1, we obtain

$$H_{nlp} = \pm\frac{\pi}{2}a_n\left[\int_o^{\ell}\left(-\frac{d^2f}{dr^2}-\frac{1}{r}\frac{df}{dr}+\frac{f}{r^2}\right)frdr+\right.$$

$$\left.\int_o^{\infty}\left(-\frac{d^2w}{dr^2}-\frac{1}{r}\frac{dw}{dr}+\frac{w}{r^2}\right)wrdr\right], \quad (3.28)$$

in which we substitute from (3.5) and (3.16) for the second derivatives of f and w, respectively, before performing the integration numerically. Clearly, the normalization integral N_ν^2 in (2.16) can be written in the form

$$N_{nmp}^2 = \rho L_{nmp}, \quad (3.29)$$

where

$$L_{nmp} = \int_o^{2\pi} d\theta\left[\int_o^{\ell} r(\bar{v}^{nmp})^2 dr + \int_{\ell}^{\infty} r(v^{nmp})^2 dr\right]. \quad (3.30)$$

Since only m = 1 is of interest for perturbation purposes, substitution of (3.4) and (3.15) for m = 1 and evaluation of the angular integrals yields

$$L_{nlp} = \pi\left[\int_o^{\ell}(f^{nlp})^2 rdr + \int_{\ell}^{\infty}(w^{nlp})^2 rdr\right], \quad (3.31)$$

which is integrated numerically. The perturbed eigenfrequencies for m = 1 may now be obtained from (2.16), which enables us to write

$$\Delta_{nlp} = \frac{H_{nlp}}{2\omega_{nlp}\rho L_{nlp}}, \quad \omega = \omega_{nlp} - \Delta_{nlp}, \quad (3.32)$$

while for $m \neq 1$ there is no perturbation in frequency and we have

$$\omega = \omega_{nmp} \quad m \neq 1. \quad (3.33)$$

4. Forced Vibrations of Contoured and Beveled Cylindrical Plate

Now that we have the free vibrational modes of the contoured and beveled cylindrical plate, we can obtain the forced vibrational solution by expanding in a series of the free vibrational solutions. However, before we do this we must make an important observation. Although we obtained the natural frequencies of free vibration of the anisotropic beveled cylindrical plate, we did not obtain the mode shape for the actual anisotropic case[7]. As a consequence, we neglect the influence of A_n in the forced vibration analysis[8]. Accordingly, from (2.5) we write the equation for forced vibrations in the electroded region in the isotropic form

$$c_n\nabla^2\bar{u}^n - \rho\bar{\omega}_n^2\bar{u}^n + \rho\omega^2\bar{u}^n = \bar{F}_n e^{i\omega t}, \quad (4.1)$$

where from (2.7), we have

$$\bar{F}_n = \rho\omega^2(-1)^{(n-1)/2}e_{26}4V/c^{(1)}n^2\pi^2 \quad (4.2)$$

and from (2.5), (2.8), (3.12) and $(3.14)_2$ we may write the equation for forced vibrations in the unelectroded region in the form

$$c_n\nabla^2u^n - \rho\omega_{nd}^2(1+\varsigma d/h_d R)u^n + \rho\omega^2u^n = 0. \quad (4.3)$$

We now write the steady-state solution of (4.1) and (4.3) as a sum of the eigensolutions obtained in Section 3, thus

$$\bar{u}^n = e^{i\omega t}\sum_{mp} K^{nmp}\bar{v}^{nmp}, u^n = e^{i\omega t}\sum_{mp} K^{nmp}v^{nmp} \quad (4.4)$$

where

$$\bar{v}^{nmp} = f^{nmp}e^{im\theta}, v^{nmp} = w^{nmp}e^{im\theta}, \quad (4.5)$$

and f^{nmp} and w^{nmp} denote the r and ζ dependent solution functions obtained in Section 3. Since no energy is lost from the eigensolutions for the beveled resonator obtained in Section 3, they satisfy an orthogonality condition, which has been derived in Sec.IV of Ref.2 and is given by

$$\int_o^{2\pi} d\theta\left[\int_o^{\ell} r\bar{v}^{nmp}\bar{v}^{rst}dr + \int_{\ell}^{\infty} rv^{nmp}v^{nst}dr\right] =$$

$$L_{n(m)(p)}\delta_{ms}\delta_{pt}. \quad (4.6)$$

Substituting from (4.4) into (4.1) and (4.3) and employing the homogeneous forms in (3.1) with (3.2) and (3.13), we obtain

$$\sum_{mp} K^{nmp}\rho(\omega^2 - \omega_{nmp}^2)\bar{v}^{nmp} = \bar{F}_n,$$

$$\sum_{mp} K^{nmp}\rho(\omega^2 - \omega_{nmp}^2)v^{nmp} = 0, \quad (4.7)$$

in the electroded and beveled regions, respectively. From (4.7) we form

$$\sum_{mp} K^{nmp}\rho(\omega^2 - \omega_{nmp}^2)\int_o^{2\pi} d\theta\left[\int_o^{\ell} r\bar{v}^{nmp}\bar{v}^{nst}dr + \right.$$

$$\left.\int_{\ell}^{\infty} rv^{nmp}v^{nst}dr\right] = \int_o^{2\pi} d\theta\int_o^{\ell}\bar{F}_n\bar{v}^{nst}r\,dr, \quad (4.8)$$

and with the aid of the orthogonality condition in (4.6), we obtain

$$K^{nst} = \bar{F}_n \mathcal{J}_{\text{nst}} / \rho(\omega^2 - \omega_{nst}^2) L_{nst}, \qquad (4.9)$$

$$\mathcal{J}_{\text{nmp}} = \int_o^{2\pi} d\theta \int_o^{\ell} r \bar{v}^{nmp} dr, \qquad (4.10)$$

where

$$L_{nmp} = \int_o^{2\pi} d\theta \left[\int_o^{\ell} r(\bar{v}^{nmp})^2 dr + \int_{\ell}^{\infty} r(v^{nmp})^2 dr \right]. \qquad (4.11)$$

The substitution of (4.2) into (4.9) and (3.4) with (3.9) for $\bar{B} = 1$ into (4.10) enables us to write

$$K^{nmp} = \frac{(-1)^{\frac{n-1}{2}} e_{264} V \mathcal{J}_{\text{nmp}}}{c^{(1)} n^2 \pi^2 (1 - \omega_{nmp}^2 / \omega^2) L_{nmp}}, \qquad (4.12)$$

$$\mathcal{J}_{\text{nmp}} = 0 \text{ for } m \neq 0, \quad \mathcal{J}_{\text{nop}} = 2\pi \int_o^{\ell} r f^{nop} dr, \qquad (4.13)$$

where the integrals in (4.11) and (4.13) are performed numerically.

On account of the resonance denominator in (4.12), in the vicinity of a resonance one term in the sums in (4.4) dominates the others, and in view of the transformations for electroded regions in Eqs.(87) and (88) of Ref.1, the solution in the electroded inner region may be written in the form

$$\bar{u} = e^{i\omega t} K^{nmp} \sin \frac{n\pi x_2}{2\bar{h}} \bar{v}^{nmp} - \frac{e_{26} V x_2}{c^{(1)} 2\bar{h}} e^{i\omega t},$$

$$\bar{\varphi} = \frac{V x_2}{2\bar{h}} e^{i\omega t} + \frac{e_{26}}{\varepsilon_{22}} \left[\sin \frac{n\pi x_2}{2\bar{h}} - (-1)^{\frac{n-1}{2}} \frac{x_2}{\bar{h}} \right] K^{nmp} v^{nmp} e^{i\omega t}, \qquad (4.14)$$

where, as usual, ω_{nmp} in Eqs.(4.14) with (4.12) is to be replaced by

$$\hat{\omega}_{nmp} = \omega_{nmp} + i\omega_{nmp} / 2Q_{nmp}, \qquad (4.15)$$

in which Q_{nmp} is the unloaded quality factor in the nmpth mode. The admittance Y of the nopth mode of the beveled resonator is obtained by first substituting from (4.14) into[9]

$$\bar{D}_2 = e_{26} \bar{u}_{,2} - \varepsilon_{22} \bar{\varphi}_{,2} \qquad (4.16)$$

and then substituting from (4.16) into

$$I = -\int_o^{2\pi} d\theta \int_o^{\ell} \dot{\bar{D}}_2 r \, dr, \qquad (4.17)$$

with the result

$$Y = \frac{I}{V} = i\omega \frac{\pi \ell^2}{2\bar{h}} \varepsilon^{22} (1 + \hat{k}_1^2) - \frac{i\omega \varepsilon_{22} \hat{k}_1^2 4 \mathcal{J}_{\text{nop}}^2}{\bar{h} n^2 \pi^2 (1 - \hat{\omega}_{nop}^2 / \omega^2) L_{nop}}, \qquad (4.18)$$

which by defining the static and motional capacitances C_o and $\hat{C}_{nop}$, respectively, may be written in the form

$$Y = i\omega [C_o + \hat{C}_{nop} / [(\hat{\omega}_{nop}^2 / \omega^2) - 1], \qquad (4.19)$$

where the expressions for C_o and $\hat{C}_{nop}$ are clear from (4.18) and (4.19).

Finally, substituting from (3.4) and (3.15) into (4.11) and performing the integration with respect to θ for m = 0, we obtain

$$L_{nop} = 2\pi \left[\int_o^{\ell} r(f^{nop})^2 dr + \int_{\ell}^{\infty} r(w^{nop})^2 dr \right], \qquad (4.20)$$

in which the integrals in the brackets are performed numerically.

5. Calculated Results

The results of calculations based on the analyses presented in Secs.3 and 4 are presented in this section. However, we first indicate the degree of anisotropy present in the families of anharmonics for the first, third and fifth harmonic for the SC cut. In Table I we show the values of c_n, a_n and a_n / c_n for SC-cut quartz. The table shows that for the third and the fifth families of anharmonics the SC-cut is very nearly isotropic in the $x_1 - x_3$ plane, whereas for the fundamental the cut is quite anisotropic. Since the perturbation analysis in Sec.3 shows that for m ≠ 1 there is no perturbation in eigenfrequency linear in the anisotropy a_n / c_n, the results should be very accurate even for the quite anisotropic fundamentals for $m \neq 1$ because the first correction to the eigenfrequency for the nearby isotropic plate would be proportional to $(a_n / c_n)^2$, which from the last column in Table I is always quite small. Since for m = 1 the

perturbation in frequency is calculated to order a_n / c_n, the results should be accurate even for m = 1 in the case of the fundamental families. However, since the anisotropy is so small for the third and fifth families of anharmonics, the perturbation in frequency for m = 1 should be essentially negligible for these families. The calculations indicate that this is indeed the case.

Calculations have been performed for a fundamental, 2.5 MHz SC-cut quartz resonator with an inner radius of curvature $\overline{R}$ = 10 cm and an outer radius of curvature R = 4 cm and an electrode diameter $2\ell = 6.8\,mm$. The results are shown in Table II for m = 0 and 1 and p = 0 - 3. The perturbation in frequency labeled delta is shown for m = 1 and it can be seen that it is considerably smaller than the spacings in frequency between the anharmonics even in this highly anisotropic case. The calculated motional capacitance in Farads is shown for m = 0. It is not shown for m = 1 because, as already noted, in this description of the mode shape it vanishes for $m \neq 0$.

Table III shows the calculated results for a third overtone, 10 MHz SC-cut resonator for $\overline{R} = 40\,cm$, R = 6 cm and $2\ell = 6.8\,mm$. It can be seen that the calculated perturbation in frequency for m = 1 is considerably smaller for the third overtone family than for the fundamental family shown in Table II and is, in fact, negligible. As already noted, this is a consequence of the near isotropy in the $x_1' - x_3'$ plane of the third overtone family of modes. Table IV shows the same type of calculated results for the fundamental of a third overtone, 10 MHz TCXO with $\overline{R} = 20\,cm$, R = 4 cm and $2\ell = 6.8\,mm$.

Table V shows calculated results for the harmonic (m =0, p = 0) of a fundamental, 2.5 MHz SC-cut resonator with fixed inner radii of curvature $\overline{R}$ and electrode diameters 2ℓ for different outer radii of curvature R. Four sets of calculations are shown in the table. The three sets labeled (a), (b) and (c) were calculated using the analysis presented in this work. The set labeled (a') is for the same resonator as the set labeled (a) but is calculted from an analysis for contoured resonators with beveled edges in which the circular junction between the inner and outer regions is approximated by the circumscribed square[10]. The table shows that the calculation of the resonant frequencies and motional capacitances for the harmonic (N = 1) using the rectangular approximation are quite accurate, but that the energy ratio is too high. This discrepancy is due to the fact that in cylindrical coordinates the energy is spreading over a larger distance with r, while in rectangular coordinates the energy does not spread. The sets labeled (b) and (c) behave essentially as expected with the changed geometry in comparison with (a). Table VI shows the same type of calculated results for the harmonic (m = 0, p = 0) of a third overtone, 10 MHz SC-cut resonator. Again the set labeled (a') is for the same resonator as the set labeled (a) but is calculated from the analysis presented in Ref.9. This table also shows that the calculations for the third harmonic based on either analysis are in substantially good agreement. Table VII shows the same type of calculated results for the fundamental of a third overtone, 10 MHz TCXO.

Figure 2 shows the radial variation in mode shape for a fundamental 2.5 MHz, SC-cut resonator for the same inner radius of curvature $\overline{R}$ and different outer radii of curvature. The vertical dotted line denotes the demarcation point between the inner and outer radii as well as the edge of the electrode. The energy ratio de, which is defined as the ratio of the energy in the calculated mode from the edge of the plate to ∞ divided by the total energy in the calculated mode to ∞, is also shown in the figure. The curves are relatively close together because $\overline{R}$ exists over such a large portion of the plate. Figure 3 shows similar curves for the same $\overline{R}$ and R, but in which the junction between $\overline{R}$ and R, is much closer to the center. The curves are further apart than in Fig.2 because the variable outer radius R extends over such a large region of the plate. Figure 4 shows the variation of the first radial anharmonic of the fundamental family for the same resonators treated in Fig.2. Figure 5 shows the variation of the first radial anharmonic of the fundamental family for the same resonators treated in Fig.3. The curves are virtually identical because the outer radius R extends over such a large portion of the plate and the largest value of R is adequate to obtain almost complete trapping for p = 1.

Acknowledgements

This work was supported in part by the Army Research Office under Grant No.DAAH04-G-0345.

REFERENCES

1. D.S. Stevens and H.F. Tiersten, "An Analysis of Doubly-Rotated Quartz Resonators Utilizing Essentially Thickness Modes with Transverse Variation," J. Acoust. Soc. Am., 79, 1811 (1986).

2. H.F. Tiersten and Y.S. Zhou, "Transversely Varying Thickness Modes in Quartz Resonators with Beveled Cylindrical Edges," J. Appl. Phys., 76, 7201 (1994).

3. Ref.1, Eq.(99).

4. J. Detaint, B. Capelle, A. Zarka and Y. Zhang, "A New Model for the Thickness Shear Resonators with Spherical Contours: Application to Beveled Devices," Proceedings of the 1995 International Frequency Control Symposium, These Proceedings.

5. We could take d = ℓ in $(2.12)_2$, but a small error is made by neglecting terms of order ς^2/d^2. By taking d slightly larger than ℓ, we can make this error smaller. The solution is not sensitive to the exact value of d in this small range.

6. The procedure could be carried to higher order, i.e., ς^2/d^2, by including $e^{-k_2\varsigma^3}$ in the solution and retaining terms of order ς^2/d^2 everywhere. This would increase the order of Eq.(3.20) and is not deemed to be purposeful in this work.

7. We found the eigenfrequencies by obtaining the perturbation in frequency from the nearby isotropic solution resulting from the anisotropy. However, since the perturbation in frequency vanishes for all $m \neq 1$, the mean (or isotropic) mode shape is quite accurate for all m ≠ 1. Moreover, for n = 3 and 5, the perturbation in frequency is very small for m = 1, as shown in Tables II-V. Consequently, for n = 3 and 5 the mean mode shape is accurate even for m = 1. However, on account of this approximation, although the motional capacitances for all nop modes, which are the largest ones and of most interest, can be calculated very accurately, the very small motional capacitances of all the modes with m ≠ 0 resulting from the small anisotropy of the actual mode shape cannot be calculated from this description.

8. If A_n were included in the forced vibration analysis, the orthogonality of the isotropic solution functions would be lost and the solution would become complicated and not purposeful.

9. Equation (135) of Ref.1. As already noted all other $Y_{nmp}(m \neq 0)$ vanish by virtue of (4.13).

10. H.F. Tiersten, B.J. Lwo and B. Dulmet, "An Analysis of Transversely Varying Thickness Modes in Trapped Energy Resonators with Shallow Contours," Proceedings of the 1994 IEEE International Frequency Control Symposium, IEEE Cat. No.94CH3446-2, Institute of Electrical and Electronics Engineers, New York,172 (1994).

TABLE I

PLANAR ISOTROPIC AND ANISOTROPIC COEFFICIENTS FOR TRANSVERSELY VARYING THICKNESS MODES IN SC-QUARTZ RESONATORS

n	c_n 10^{10} N/m^2	a_n 10^{10} N/m^2	a_n/c_n
1	7.716	2.043	0.265
3	5.877	0.516	0.088
5	7.495	0.411	0.055

TABLE II

CALCULATED RESONANT FREQUENCIES AND MOTIONAL CAPACITANCES OF 2.5 MHz, FUNDAMENTAL, SC-CUT RESONATOR

SC-cut, N = 1, f_{100} = 2.5 MHz

$\bar{R}$ = 10 *cm*, R = 4 cm, 2l = 6.8 mm, Resonator Diameter = 12 mm

h_0 = 0.372755 mm, Edge Thickness = 0.191105 mm, de = 2.32139 × 10^{-9}

Nodal Axis	Indices m p	Frequency kHz	Delta kHz	Motional Capacitance Farads
	0 0	2500.001		2.94×10^{-15}
	0 1	2720.883		5.76×10^{-16}
	0 2	3020.791		6.11×10^{-21}
	0 3	3457.353		3.54×10^{-17}
x_1	1 0	2587.356	14.879	
	1 1	2815.081	38.433	
	1 2	3137.020	77.491	
	1 3	3585.867	129.361	
x_3	1 0	2617.052	-14.879	
	1 1	2891.948	-38.433	
	1 2	3292.002	-77.491	
	1 3	3844.590	-129.361	

TABLE III

CALCULATED RESONANT FREQUENCIES AND MOTIONAL CAPACITANCESOF 10 MHz, THIRD OVERTONE, SC-CUT RESONATOR

SC-Cut, N = 3, f_{300} = 10 MHz
$\overline{R}$ = 40 cm, R = 6 cm, 2l = 6.8 mm, Resonator Diameter = 12 mm
h_0 = 0.2694008 mm, Edge Thickness = 0.160342 mm, de = 2.23973×10^{-29}

Nodal Axis	Indices m p	Frequency kHz	Delta kHz	Motional Capacitance Farads
	0 0	10000.000		2.27×10^{-16}
	0 1	10101.519		1.97×10^{-16}
	0 2	10207.941		1.11×10^{-16}
	0 3	10334.908		3.53×10^{-17}
x_1	1 0	10048.489	2.227	
	1 1	10148.438	4.708	
	1 2	10258.795	8.518	
	1 3	10395.034	14.013	
x_3	1 0	10052.943	-2.227	
	1 1	10157.853	-4.708	
	1 2	10275.831	-8.518	
	1 3	10423.061	-14.013	

TABLE IV

CALCULATED RESONANT FREQUENCIES AND MOTIONAL CAPACITANCESOF 3.3 MHz, FUNDAMENTAL, SC-CUT RESONATOR (TCXO, WHEN IN USE, 10 MHz. THIRD OVERTONE)

SC-Cut, N = 1, f_{100} = 3.3 MHz
$\overline{R}$ = 20 cm, R = 4 cm, 21 = 6.8 mm, Resonator Diameter = 12 mm
h_o 0.2777155 mm, Edge Thickness = 0.10516 mm,
de = 2.00945×10^{-12}

Nodal Axis	Indices m p	Frequency kHz	Delta kHz	Motional Capacitance Farads
	0 0	3300.004		3.78×10^{-15}
	0 1	3488.835		1.00×10^{-15}
	0 2	3750.379		4.11×10^{-17}
	0 3	4103.270		2.09×10^{-17}
x_1	1 0	3373.259	12.829	
	1 1	3571.031	34.927	
	1 2	3842.850	66.184	
	1 3	4206.992	109.608	
x_3	1 0	3398.917	-12.829	
	1 1	3640.885	-34.927	
	1 2	3975.219	-66.184	
	1 3	4426.208	-109.608	

TABLE V

CALCULATED RESONANT FREQUENCIES, ENERGY RATIOS, AND MOTIONAL CAPACITANCES OF FUNDAMENTAL HARMONIC, 2.5 MHz SC-CUT RESONATOR

SC-Cut, 2.5 MHz, N = 1, m = 0, p = 0, Resonator Diameter = 12 mm

(a) $\overline{R}$ = 10 cm, electrode diameter 21 = 6.8 mm, $2h_o$ = 0.74522 mm

R(cm)	Edge Thickness (mm)	Energy Ratio	Motional Capacitance Farads	Frequency kHz
8	0.53467	1.2238×10^{-7}	2.97×10^{-15}	2500.005
6	0.48375	2.9429×10^{-8}	2.96×10^{-15}	2500.366
4	0.38192	2.2946×10^{-9}	2.94×10^{-15}	2500.952

(*a'*) Same as in (a) but with rectilinear approximation

R(cm)	Edge Thickness (mm)	Energy Ratio	Motional Capacitance Farads	Frequency kHz
8	0.53467	1.4394×10^{-6}	2.95×10^{-15}	2498.619
6	0.48375	2.9482×10^{-7}	2.94×10^{-15}	2499.372
4	0.38192	2.2620×10^{-8}	2.91×10^{-15}	2500.369

(*b*) $\overline{R}$ = 10*cm*, electrode diameter 21 = 3.4 mm, $2h_o$ = 0.74799 mm

R(cm)	Edge Thickness (mm)	Energy Ratio	Motional Capacitance (Farads)	Frequency kHz
8	0.52660	1.9512×10^{-7}	6.31×10^{-16}	2500.005
6	0.45762	1.7389×10^{-8}	6.65×10^{-16}	2506.048
4	0.31966	3.0059×10^{-10}	7.07×10^{-16}	2515.042

(c) $\overline{R}$ = 15 cm, electrode diameter 21 = 6.8 mm, $2h_o$ = 0.74021 mm.

R(cm)	Edge Thickness (mm)	Energy Ratio	Motional Capacitance (Farads)	Frequency kHz
8	0.54892	3.8539×10^{-7}	3.14×10^{-15}	2500.006
6	0.49801	8.8644×10^{-8}	3.13×10^{-15}	2500.630
4	0.39617	6.6946×10^{-9}	3.12×10^{-15}	2501.597

TABLE VI

CALCULATED RESONANT FREQUENCIES, ENERGY RATIOS, AND MOTIONAL CAPACITANCES OF THIRD HARMONIC, 10 MHz SC-CUT RESONATOR
SC-Cut, 10 MHz, N = 3, m = 0, p = 0

(a) $\overline{R}$ = 30 cm, Resonator Diameter = 12 mm, Electrode Diameter 21 = 6.8 mm, $2h_o$ = 0.53923 mm.

R(cm)	Edge Thickness (mm)	Energy Ratio	Motional Capacitance (Farads)	Frequency kHz
15	0.43850	0.	1.98×10^{-16}	9999.995
10	0.39776	0.	1.98×10^{-16}	9999.995
6	0.31630	0.	$1,98 \times 10^{-16}$	9999.995

(*a'*) Same as in (a) but with rectilinear approximation

R(cm)	Edge Thickness (mm)	Energy Ratio	Motional Capacitance (Farads)	Frequency kHz
15	0.43850	0.	1.98×10^{-16}	10008.960
10	0.39776	0.	1.98×10^{-16}	10008.960
6	0.31630	0.	1.98×10^{-16}	10008.960

(b) $\overline{R}$ = 40 cm, Resonator Diameter = 8 mm, Electrode Diameter 21 = 4.5 mm, $2h_o$ = 0.53884 mm.

R(cm)	Edge Thickness (mm)	Energy Ratio	Motional Capacitance (Farads)	Frequency kHz
15	0.49605	1.2783×10^{-10}	1.82×10^{-16}	10000.001
10	0.47783	6.4546×10^{-12}	1.81×10^{-16}	10000.118
6	0.44137	4.3016×10^{-14}	1.80×10^{-16}	10000.294

TABLE VII

CALCULATED RESONANT FREQUENCIES, ENERGY RATIOS, AND MOTIONAL CAPACITANCES OF FUNDAMENTAL HARMONIC, 3.3 Mhz SC-CUT RESONATOR (TCXO, WHEN IN USE, 10 MHz, THIRD HARMONIC)
TCXO, SC-Cut, 3.3 MHz, N = 1, m = 0, p = 0, Resonant Diameter = 12 mm

(a) $\overline{R}$ = 15 cm, Electrode Diameter 21 = 6.8 mm, $2h_o$ = 0.55733 mm.

R (cm)	Edge Thickness (mm)	Energy Ratio	Motional Capacitance (Farads)	Frequency kHz
8	0.36605	3.4763×10^{-10}	3.59×10^{-15}	3300.005
6	0.31513	3.8907×10^{-11}	3.58×10^{-15}	3300.161
4	0.21330	7.1015×10^{-13}	3.56×10^{-15}	3300.405

(b) $\overline{R}$ = 15 cm, Electrode Diameter 21 = 3.4 mm, $2h_o$ = 0.55990 mm.

R(cm)	Edge Thickness (mm)	Energy Ratio	Motional Capacitance (Farads)	Frequency kHz
8	0.34333	8.1890×10^{-11}	9.85×10^{-16}	3300.005
6	0.27435	2.1459×10^{-12}	1.01×10^{-15}	3304.468
4	0.13649	4.1543×10^{-15}	1.05×10^{-15}	3311.099

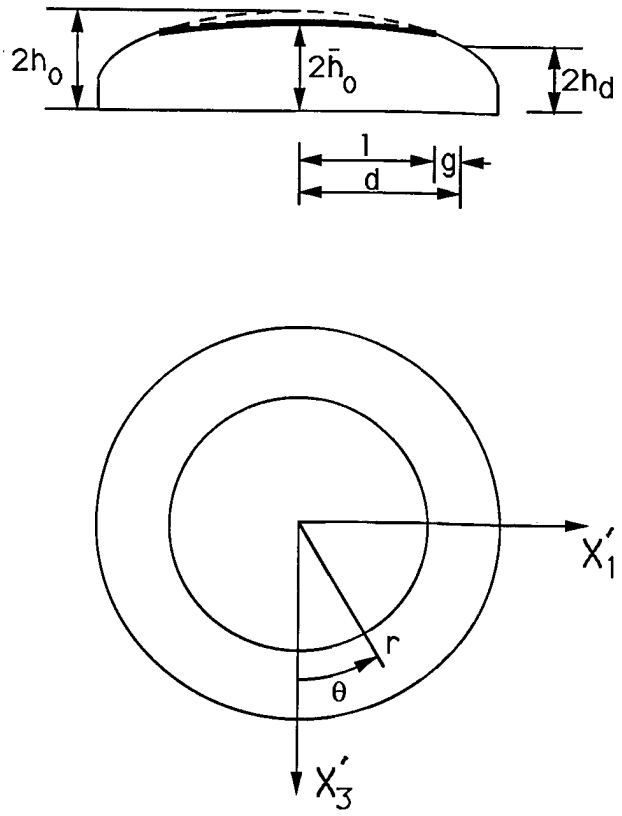


Figure 1 Cross-Section and Plan View of Contoured and Beveled Resonator

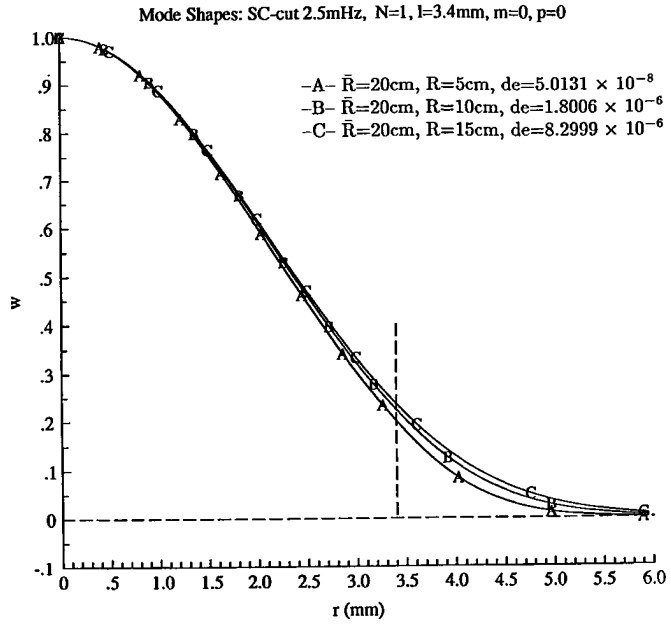


Figure 2 Radial Dependence of Mode Shape for the Fundamental, 2.5 MHz, SC-Cut Resonator for Fixed Inner Radius of Contour and Different Outer Radii of Contour with Large Inner Region

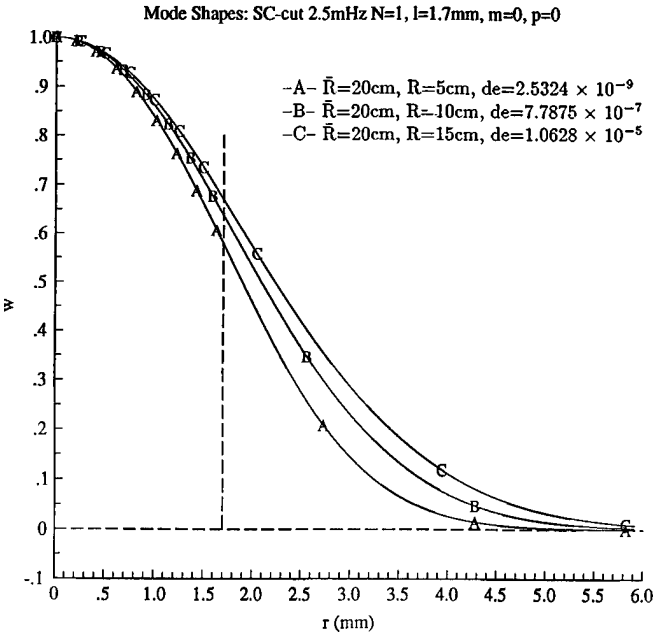


Figure 3 Radial Dependence of Mode Shape for the Fundamental, 2.5 MHz SC-Cut Resonator for Fixed Inner Radius of Contour and Different Outer Radii of Contour with Small Inner Region

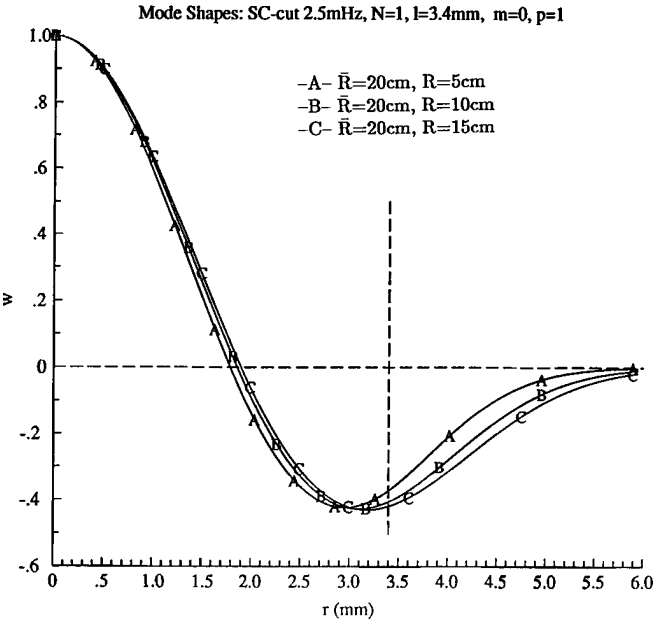


Figure 4 Radial Dependence of Mode Shape for the First Radial Anharmonic of the Fundamental for the Resonator in Figure 2

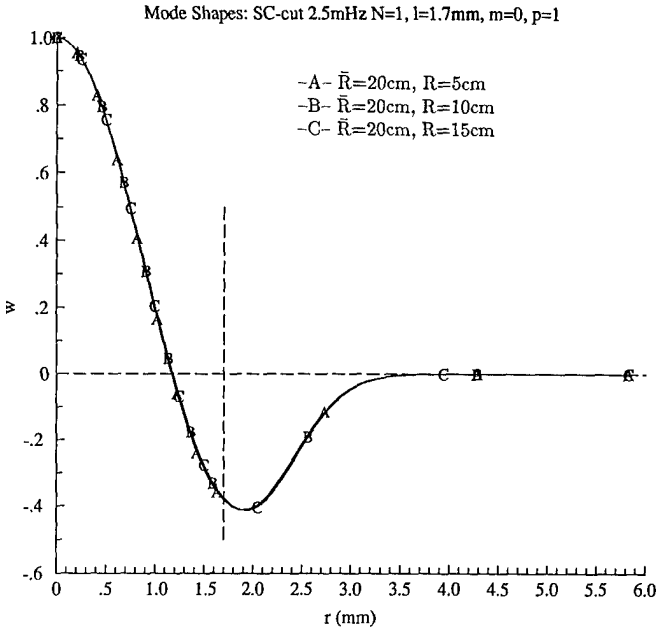


Figure 5 Radial Dependence of Mode Shape for the First Radial Anharmonic of the Fundamental for the Resonator in Figure 3

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THE INFLUENCE OF AN ASYMMETRIC AIR GAP ON THE VIBRATION CHARACTERISTICS OF QUARTZ TRAPPED ENERGY RESONATORS

H.F. Tiersten

Department of Mechanical Engineering, Aeronautical Engineering and Mechanics
Rensselaer Polytechnic Institute, Troy, New York 12180-3590

Abstract

An analysis of the influence of an asymmetric air gap on transversely varying thickness modes in quartz trapped energy resonators is performed. The thickness-dependent coupled electrostatic-piezoelectric problem that transforms the inhomogeneities from the electrodes into the equation for transversely varying thickness modes is identified and solved. It turns out that the asymmetry of the gap has no influence on any of the coefficients in the thickness solution that influence either the resonant frequencies or motional and static capacitances and only the mean value of the gap does. The equation for transversely varying thickness modes modified to account for the gap is presented. The solution for the harmonics and anharmonics of a contoured resonator when a gap is present is given along with the expression for the asssociated resonant frequencies.

1. Introduction

In obtaining the equation for transversely varying thickness modes in quartz resonators with attached electrodes, a simple transformation of the driving voltage from the electrodes to the relevant differential equation is performed[1,2]. If an air gap is present, either symmetric or not, the transformation is invalid.

In this paper the thickness-dependent problem that transforms the driving voltage from the electrodes into the equation for transversely varying thickness modes when an asymmetric air gap is present is identified and solved. The analysis yields the dependence of the thickness resonant frequencies and the thickness coefficient that enables the determination of the motional capacitances of the transversely varying thickness modes on the possibly variable thickness of the plate and the air gaps. The results show that the asymmetry of the gap has no influence on either of these quantities and only the mean value of the gap does. The modifications that arise in the equation for transversely varying thickness modes on account of the gaps are presented.

It is noted that if the trapped energy resonator consists of flat terraced regions, the plate thickness and the gaps are different constants in each region and the trapped energy solution functions satisfy the usual continuity conditions between regions. If, on the other hand, the resonator is contoured, both the plate thickness and the gaps are functions of position along the plate. The modifications in the solution functions for the contoured resonator when the edge of the electrode is ignored[3,2] in the analysis, which arise as a result of the gaps, are presented along with the modifications in the expressions for the resonant frequencies of the transversely varying thickness modes

2. Transversely Varying Thickness Modes

It has been shown[2] that the differential equation for the dominant displacement $u_1^{(1)}$ of the nth harmonic family of transversely varying thickness modes in doubly-rotated quartz plates may be written in the form

$$M_n' \frac{\partial^2 \tilde{u}^n}{\partial x_1'^2} + P_n' \frac{\partial^2 \tilde{u}_n}{\partial x_3'^2}$$

$$-\frac{n^2\pi^2}{4h^2}\tilde{c}^{(1)}\tilde{u}^n - \rho\ddot{\tilde{u}}_n = F_n e^{i\omega t}, \qquad (2.1)$$

where the upper cycle and F_n are used in a generic sense so that Eq.(2.1) can refer to electroded as well as unelectroded regions and M_n' and P_n' are involved lengthy expressions that are defined in Sec.III of Ref.2 along with x_1' and x_3', which are orthogonal directions in the plane of the plate for which the scalar differential equation (2.1) does not contain mixed derivatives. In (2.1) we have employed the relation

$$u_1^n = u^n(x_1', x_3', t)\sin\frac{n\pi x_2}{2h}, \qquad (2.2)$$

where $u_1^{(n)}$ is the pure thickness eigendisplacement of the unelectroded plate with eigenvalue $\bar{c}^{(1)}$ when $u^{(n)} = u^{(n)}(t)$ and 2h and ρ denote the possibly variable thickness and mass density of the resonator, respectively.

In an electroded region we have

0-7803-2500-1/95 $4.00 © 1995 IEEE

$$F_n = \frac{\rho\omega^2(-1)^{\frac{n-1}{2}} e_{26}4V}{c^{(1)}n^2\pi^2}, \quad \tilde{c}^{(1)} = \hat{c}^{(1)},$$

$$\tilde{u}^n = \bar{u}^n, \quad \tilde{\omega}_n = \bar{\omega}_n, \tag{2.3}$$

and in an unelectroded region we have

$$F_n = 0, \ \tilde{c}^{(1)} = \bar{c}^{(1)}, \ \tilde{u}^n = u^n,$$

$$\tilde{\omega}_n = \omega_n, \tag{2.4}$$

where V is the driving voltage, ω is the driving frequency, e_{26} is the piezoelectric constant referred to the thickness eigendisplacemement direction $u_1^{(n)}$ and $\bar{\omega}_n$ and ω_n are the respective pure thickness eigenfrequencies for completely electroded and unelectroded flat plates of thickness $2h_o$, which are given by

$$\bar{\omega}_n^2 = \frac{n^2\pi^2\hat{c}^{(1)}}{4h_o^2\rho}, \quad \omega_n^2 = \frac{n^2\pi^2\bar{c}^{(1)}}{4h_o^2\rho}. \tag{2.5}$$

In (2.3) and (2.5) we have employed the relations

$$\hat{c}^{(1)} = \bar{c}^{(1)}(1-\Delta), \quad c^{(1)} = \bar{c}^{(1)}(1-k_1^2), \tag{2.6}$$

where

$$\Delta = \frac{8k_1^2}{n^2\pi^2} + 2\hat{R}, \quad k_1^2 = \frac{e_{26}^2}{\bar{c}^{(1)}\varepsilon_{22}},$$

$$\hat{R} = \frac{2\rho'h'}{\rho h}, \tag{2.7}$$

and $2h'$ and ρ' denote the thickness and mass density of each electrode. When the plate is contoured rather than flat it has been shown[4,3] that to lowest order the variable thickness of a plano-convex resonator may be written in the form

$$2h = 2h_o\left[1 - \frac{(x_1'^2 + x_3'^2}{4Rh_o}\right]. \tag{2.8}$$

Equation (2.1) satisfies all mechanical differential equations and boundary conditions to lowest order for slow variation along the plate compared with the variation through the thickness. However, only the thickness dependence of all electrical variables is included in the treatment, which is an accurate approximation because of the slow variation along the plate and the small piezoelectric coupling.

When the driving electrodes are attached to the plate the inhomogeneous voltage term is transformed from the electrode into the relevant mechanical differential equation by the transformation[1,2]

$$\bar{u}_1 = u_1 + Kx_2e^{i\omega t}, \quad \bar{\varphi} = \frac{Vx_2}{2h}e^{i\omega t} + \frac{e_{26}}{\varepsilon_{22}}u_1 + Cx_2, \tag{2.9}$$

where

$$K = -e_{26}V / c^{(1)}2h, \quad C = -(e_{26} / \varepsilon_{22})u_1(h) / h. \tag{2.10}$$

Since C influences the resonant frequencies of the modes and C and K influence the motional capacitances of the modes, in this work we seek the equivalent of C and K when there is an asymmetric air gap between the electrodes and the vibrating plate.

3. Pure Thickness Vibrations with Asymmetric Air Gap

A cross-section of the plate showing the driving electrodes, the asymmetric air gap and the coordinate system is shown in Fig.1. Since Eq.(2.1), which is obtained in Ref.2, holds when the thickness eigendisplacement u_1 dominates the other two, we proceed as in Ref.2 and find the dominant pure thickness eigensolution for u_1 which satisfies the vanishing of the one dominant surface traction condition in the direction of the dominant thickness eigendisplacement u_1 on each surface[5]. We now note that in the formulation all variables have $e^{i\omega t}$ dependence, which is suppressed. In terms of the aforementioned dominant thickness eigendisplacement u_1, the relevant thickness dependent differential equations in the piezoelectric plate and the air gap may be written in the form

$$c^{(1)}u_{1,22} + e_{26}\varphi_{,22} + \rho\omega^2u_1 = 0,$$
$$e_{26}u_{1,22} - \varepsilon_{22}\varphi_{,22} = 0, \ \varepsilon_o\varphi^T_{,22} = 0, \ \varepsilon_o\varphi^B_{,22} = 0, \tag{3.1}$$

where $c^{(1)}$ is given in $(2.6)_2$, ε_o is the permittivity of free-space and φ^T and φ^B are the electric potentials in the top and bottom gaps, respectively, and φ is the electric potential in the piezoelectric plate. The relevant constitutive equations in the piezoelectric plate and the top and bottom gaps may be written in the form

$$T_{21} = c^{(1)}u_{1,2} + e_{26}\phi_{,2}, \quad D_2 = e_{26}u_{1,2} - \varepsilon_{22}\varphi_{,2}$$
$$D_2^T = -\varepsilon_o\varphi^T_{,2}, \quad D_2^B = -\varepsilon_o\varphi^B_{,2}, \tag{3.2}$$

where T_{21} is the component of traction on planes normal to x_2 in the dominant eigendisplacement direction and D_2 represents the component of electric displacement in the x_2-direction. These relevant boundary and continuity conditions are given by

$$\varphi^T = V \ \text{ at } x_2 = h + g^T,$$
$$\varphi^B = 0 \ \text{ at } x_2 = -h - g^B, \tag{3.3}$$
$$T_{21} = 0, \ \varphi = \varphi^T, \ D_2 = D_2^T \ \text{at } x_2 = h, \tag{3.4}$$
$$T_{21} = 0, \ \varphi = \varphi^B, \ D_2 = D_2^B \ \text{ at } x_2 = -h. \tag{3.5}$$

In order to transfer the inhomogeneity from the boundary condition $(3.3)_1$ into the differential equation (3.1), we write

$$u_1 = \bar{u} + Kx_2, \quad \varphi = \bar{\varphi} + Lx_2 + G,$$
$$\varphi^T = \bar{\varphi}^T + L^T x_2 + G^T, \quad \varphi^B = \bar{\varphi}^B + L^B x_2 + G^B, \quad (3.6)$$

where the constants are to be chosen so that $\bar{u}, \bar{\varphi}, \bar{\varphi}^T$ *and* $\bar{\varphi}^B$ satisfy homogeneous boundary conditions. Substitution of (3.6) into the differential equations (3.1) yields

$$\bar{c}^{(1)}\bar{u}_{,22} + \rho\omega^2\bar{u} + \rho\omega^2 Kx_2 = 0,$$

$$\bar{\varphi} = \frac{e_{26}}{\varepsilon_{22}}\bar{u} + Cx_2 + F, \quad \varepsilon_o\bar{\varphi}^T_{,22} = 0, \varepsilon_o\bar{\varphi}^B_{,22} = 0, \quad (3.7)$$

where we have integrated $(3.1)_2$ to obtain $(3.7)_2$. Substitution of (3.6) into the boundary conditions (3.3) - 3.5) yields

$$\bar{\varphi}^T + L^T(h+g^T) + G^T = V, \quad \bar{\varphi}^B - L^B(h+g^B) + G^B = 0,$$
$$\bar{\varphi} + Lh + G = \bar{\varphi}^T + L^T h + G^T,$$
$$\bar{\varphi} - Lh + G = \bar{\varphi}^B - L^B h + G^B,$$
$$\bar{c}^{(1)}\bar{u}_{,2} + c^{(1)}K + eC + eL = 0 \text{ at } x_2 = \pm h,$$
$$e_{26}K - \varepsilon_{22}C - \varepsilon_{22}L = -\varepsilon_o\bar{\varphi}^T_{,2} - \varepsilon_o L^T,$$
$$e_{26}K - \varepsilon_{22}C - \varepsilon_{22}L = -\varepsilon_o\bar{\varphi}^B_{,L} - \varepsilon_o L^B, \quad (3.8)$$

where the constants $K, L, L^T, L^B, G, G^T, G^B$ are to be selected so that

$$\bar{\varphi}^T = 0 \text{ at } x_2 h + g^T, \quad \bar{\varphi}^B = 0 \text{ at } x_2 = -h - g^B,$$
$$\bar{\varphi} = \bar{\varphi}^T, \quad \bar{c}^{(1)}\bar{u}_{,2} + e_{26}C = 0,$$
$$-\varepsilon_{22}C = -\varepsilon_o\bar{\varphi}^T_{,2} \text{ at } x_2 = h,$$
$$\bar{\varphi} = \bar{\varphi}^B, \quad \bar{c}^{(1)}\bar{u}_{,2} + e_{26}C = 0,$$
$$-\varepsilon_{22}C = -\varepsilon_o\bar{\varphi}^B_{,2} \text{ at } x_2 = -h. \quad (3.9)$$

From (3.8) and (3.9) we find that the constants are given by

$$L = \frac{V}{2h(1+\alpha)}, \quad G = \frac{V}{2}\left[1 + \frac{s}{(1+\alpha)}\frac{(g^B - g^T)}{2h}\right],$$

$$L^T = L^B = sL, \quad G^T = V - sL(h+g^T),$$
$$G^B = sL(h+g^B),$$

$$K = -\frac{e_{26}V}{c^{(1)}2h(1+\alpha)}, \quad \alpha = \frac{s(g^T + g^B)}{2h},$$

$$s = \frac{1+\hat{k}_1^2}{r_o}, \quad \hat{k}_1^2 = \frac{e_{26}^2}{c^{(1)}\varepsilon_{22}}, \quad r_o = \frac{\varepsilon_o}{\varepsilon_{22}}. \quad (3.10)$$

Thus, at this point the forced thickness vibration problem is defined by (3.7) and (3.9) with the expression for K in (3.10). Since the expression for K in (3.10) and (2.10) differ by α, it is clear that the influence of the asymmetric gap on the forcing is contained in α.

We must now obtain the solutions to the associated thickness eigenvalue problem, i.e., the eigensolutions satisfying (3.7) with K = 0 and (3.9). To this end we rewrite (3.7) in the form

$$\bar{c}^{(1)}\bar{u}_{,22} + \rho\omega^2\bar{u} = 0, \quad \bar{\varphi} = \frac{e_{26}}{\varepsilon_{22}}\bar{u} + Cx_2 + F$$
$$\bar{\varphi}^T_{,22} = 0, \quad \bar{\varphi}^B_{,22} = 0. \quad (3.11)$$

We take the solution satisfying (3.11) in the form

$$\bar{u} = A\sin\eta x_2 + B\cos\eta x_2,$$
$$\bar{\varphi}^T = R^T x_2 + S^T, \quad \bar{\varphi}^B = R^B x_2 + S^B, \quad (3.12)$$

and we note that $(3.12)_1$ satisfies $(3.11)_1$ provided

$$\bar{c}^{(1)}\eta^2 = \rho\omega^2. \quad (3.13)$$

Substituting from $(3.11)_{1,2}$ into $(3.9)_{4,7}$, we obtain

$$\bar{c}^{(1)}A\eta\cos\eta h \mp \bar{c}^{(1)}B\eta\sin\eta + e_{26}C = 0 \text{ at } x_2 = \pm h. \quad (3.14)$$

The successive subtraction and addition of the two equations in (3.14) yields

$$B = 0, \quad \bar{c}^{(1)}A\eta\cos\eta h + e_{26}C = 0 \quad (3.15)$$

which shows that the solutions symmetric in x_2 cannot be driven with an asymmetric air-gap. The substitution of (3.12) into the remaining conditions in (3.9) yields

$$R^T(h+g^T) + S^T = 0, \quad -R^B(h+g^B) + S^B = 0,$$
$$(e_{26}/\varepsilon_{22})A\sin\eta h + Ch + F = R^T h + S^T,$$
$$-(e_{26}/\varepsilon_{22})A\sin\eta h - Ch + F = -R^B h + S^B,$$
$$\varepsilon_{22}C = \varepsilon_o R^T, \quad \varepsilon_{22}C = \varepsilon_o R^B. \quad (3.16)$$

Equations $(3.16)_{5,6}$ immediately show that $R^T = R^B$. The use of this fact in $(3.16)_{1,2}$ along with the difference of $(3.16)_3$ and $(3.16)_4$ enables us to obtain

$$C = -\frac{e_{26}}{\varepsilon_{22}} \frac{A \sin \eta h}{h(1+\beta)}, \tag{3.17}$$

where

$$\beta = (g^T + g^B)/2r_o h. \tag{3.18}$$

Since the expression for C in (3.17) differs from the expression for C in (2.10) by the presence of β, we have shown that the influence of the asymmetric gap on C is contained in β. Furthermore, since the entire influence of the gap on the thickness eigenmodes is contained in the expressions for α and β in $(3.10)_8$ and (3.18), which depend only on the mean gap, we have shown that any asymmetry in the total gap has no influence on any property of the thickness eigensolutions. Substituting from (3.17) into $(3.15)_2$, we obtain

$$A\left[\frac{\eta h(1+\beta)}{k_1^2} - \tan \eta h\right] = 0, \tag{3.19}$$

which for arbitrary A yields

$$\tan \eta h = \eta h(1+\beta)/k_1^2, \tag{3.20}$$

the roots ηh of which determine the eigenfrequencies ω_n from (3.13).

Since the coupling factor k_1^2 is very small, the roots of (3.20) are not far from $\eta\pi/2$, and we write

$$\eta_n h = \frac{n\pi}{2} - \Delta_n, \tag{3.21}$$

the substitution of which in (3.20) and expansion to lowest order yields

$$\Delta_n = \frac{2k_1^2}{n\pi(1+\beta)}. \tag{3.22}$$

Substituting from (3.21) and (3.22) into (3.13), we obtain

$$\omega_n = \frac{n\pi}{2h}\sqrt{\frac{\bar{c}^{(1)}}{\rho}}\left[1 - \frac{4k_1^2}{n^2\pi^2(1+\beta)}\right], \tag{3.23}$$

which gives the pure thickness resonances of the plate in an asymmetric air gap. The variable β defined in (3.18) gives the influence of the air gap. Furthermore, a comparison of the important coefficients, i.e., K, L and C in $(3.10)_{1,7}$ and (3.17), in the solution with the gap obtained in this section with the solution with attached electrodes given in (2.9) and (2.10) reveals that

$$K^g = \frac{K}{1+\alpha}, \quad C^g = \frac{C}{1+\beta}, \quad L^g = \frac{L}{1+\alpha}, \tag{3.24}$$

where the superscript g refers to the solution with the gap obtained in this section.

4. Transverse Variation with Gap

In the pure thickness solution presented in Sec.3 the plate thickness 2h and gap dimensions g^T and g^B can either be constant, as they would for each flat portion of a terraced plate, or they can be slowly varying functions of the in-plane coordinates x_1' and x_3', as in the expression for 2h for the plano-convex plate given in (2.8). Both of these cases will be discussed in this section.

From the derivation of Eq.(2.1) of this paper, which is given in Ref.2 for the case of attached electrodes, and the pure thickness solution when a gap is present, which is given in Sec.3 of this paper, it is clear[6] that Eq.(2.1) could be written in the modified form

$$M_n' \frac{\partial^2 \tilde{u}^n}{\partial x_1'^2} + P_n' \frac{\partial^2 \tilde{u}^n}{\partial x_3'^2} - \rho\tilde{\omega}_n^2 \tilde{u}^n - \rho\ddot{\tilde{u}}^n = F_n e^{i\omega t}, \tag{4.1}$$

where from (3.21), (3.22) and (3.13)

$$\rho\tilde{\omega}_n^2 = \frac{n^2\pi^2}{4h^2}\bar{c}^{(1)}\left[1 - \frac{8k_1^2}{n^2\pi^2[1+\beta]}\right]. \tag{4.2}$$

If we are treating a terraced plate with flat portions of different thickness 2h and different mean gap dimension $g = (g^T + g^B)/2$, then Eq.(4.1) holds in each region with the constant h and β for that region. In addition, at each junction between regions we require the continuity of $u^{(n)}$ and the normal derivative of u_n, as usual.

If on the other hand we are treating a contoured resonator, in which the thickness varies in accordance with Eq.(2.8), 2h and β in the expression for $\rho\tilde{\omega}_n^2$ are functions of x_1' and x_3'. Since the variation of 2h is given in (2.8), we need find only the equivalent expression for $1/(1+\beta)$ as a function of x_1' and x_3'. To this end we first note that since the electrodes are parallel, we have

$$2h + 2g = 2h_o + 2g_o, \tag{4.3}$$

where $2h_o$ and $2g_o$ are the plate thickness and total gap at the center, as shown in Fig.2. If we define ς^2 by

$$\varsigma^2 = x_1'^2 + x_3'^2, \tag{4.4}$$

then (2.8) may be written in the form

$$2h = 2h_o - \frac{\varsigma^2}{2R}, \tag{4.5}$$

the substitution of which in (4.3) yields

$$2g = 2g_o + \frac{\varsigma^2}{2R}. \tag{4.6}$$

From (3.18) and the definition of g we may write

$$\beta = g / r_o h. \tag{4.7}$$

The substitution of (4.5) and (4.6) in (4.7) and expansion to first order in ς^2 yields

$$\beta = \beta_o \left(1 + \frac{\varsigma^2}{4R\hat{g}_o}\right), \tag{4.8}$$

where

$$\beta_o = g_o / h_o r_o, \quad \hat{g}_o = g_o h_o / (g_o + h_o). \tag{4.9}$$

Note that if $h_o >> g_o, \hat{g}_o \approx g_o$. Since in (4.2) β appears in the form $1/(1+\beta)$, we must expand $1/(1+\beta)$ thus

$$\frac{1}{1+\beta} = \frac{1}{1+\beta_o + \beta_o \varsigma^2 / 4R\hat{g}_o}, \tag{4.10}$$

expansion of which to first order in ς^2 yields

$$\frac{1}{1+\beta} = \frac{1}{1+\beta_o} - \frac{\beta_o}{(1+\beta_o)^2} \frac{\varsigma^2}{4R\hat{g}_o}. \tag{4.11}$$

Substituting from (4.5) and (4.11) into (4.2), expanding, retaining terms to first order in ς^2 and recombining terms, after some manipulation, we obtain

$$\rho\tilde{\omega}_n^2 = \frac{n^2\pi^2\hat{c}^{(1)}}{4h_o^2}\left[1 + \frac{(1+\delta_n)\varsigma^2}{2Rh_o}\right], \tag{4.12}$$

where

$$\hat{c}^{(1)} = \bar{c}^{(1)}\left[1 - 8k_1^2 / n^2\pi^2(1+\beta_o)\right], \tag{4.13}$$

$$\delta_n = \frac{\bar{c}^{(1)}}{\hat{c}^{(1)}} \frac{4k_1^2\beta_o}{n^2\pi^2(1+\beta_o)} \frac{h_o}{\hat{g}_o}. \tag{4.14}$$

Substituting from (4.12) and (4.4) into (4.1), we obtain

$$M_n' \frac{\partial^2 u^n}{\partial x_1'^2} + P_n' \frac{\partial^2 u^n}{\partial x_3'^2} - \frac{n^2\pi^2\hat{c}^{(1)}}{4h_o^2}$$

$$\cdot\left[1 + \frac{(1+\delta_n)}{2Rh_o}(x_1'^2 + x_3'^2)\right]u^n - \rho\ddot{u}^n = F_n e^{i\omega t}, \tag{4.15}$$

where for an electroded region from $(2.3)_1$, $(2.10)_1$ and $(3,10)_7$, we have

$$F_n = \frac{\rho\omega^2(-1)^{\frac{n-1}{2}} e_{26} 4V}{c^{(1)} n^2\pi^2(1+\alpha)}, \tag{4.16}$$

and since α is proportional to β, we can substitute from (4.11) into (4.16) to obtain the driving term. However, since the driving term arises in calculating the motional capacitances along with the expression for C in $(2.10)_2$ for attached electrodes and in (3.17) when there is a gap, and in all earlier work[2,3] with attached electrodes on contoured resonators, 2h in (2.10), was replaced by $2h_o$, at this point we recommend that α in (4.16) be replaced by α_o, which seems justified by (4.11). However, in either the case of attached electrodes or when a gap is present this assumption is not necessary. This will not be discussed further in this work.

When the continuity conditions[2,3] at the edge of the electrodes may be left out of account the eigensolutions of (4.15), i.e., with $F_n = 0$ take the form[6]

$$u_{nmp} = e^{-\alpha_n \frac{x_1'^2}{2}} H_m(\sqrt{\alpha_n} x_1') e^{-\beta_n \frac{x_3'^2}{2}} H_p(\sqrt{\beta_n} x_3'), \tag{4.17}$$

where H_n and H_p are Hermite polynomials and

$$\alpha_n^2 = \frac{n^2\pi^2\hat{c}^{(1)}(1+\delta_n)}{8Rh_o^3 M_n'},$$

$$\beta_n^2 = \frac{n^2\pi^2\hat{c}^{(1)}(1+\delta_n)}{8Rh_o^3 P_n'}. \tag{4.18}$$

The associated eigenfrequencies are given by

$$\omega_{nmp}^2 = \frac{n^2\pi^2\hat{c}^{(1)}}{4h_o^2\rho}\left[1 + \frac{(1+\delta_n)^{1/2}}{n\pi}\left(\frac{2h_o}{R}\right)^{1/2} \cdot \left(\sqrt{\frac{M_n'}{\hat{c}^{(1)}}}(2m+1) + \sqrt{\frac{P_n'}{\hat{c}^{(1)}}}(2p+1)\right)\right], \tag{4.19}$$

where $\hat{c}^{(1)}$ and δ_n are defined in (4.13) and (4.14).

Acknowledgements

The author wishes to thank Dr. J. Vig for motivating this investigation. He also wishes to thank Drs. R. Filler and J. Vig for asking an important question immediately after the talk, the answer to which significantly improved the paper and, in fact, resulted in Sec.4.

This work was supported in part by the Army Research Office under Grant No.DAAH04-G-0345.

REFERENCES

1. H.F. Tiersten, "Analysis of Intermodulation in Thickness-Shear and Trapped Energy Resonators," J. Acoust. Soc. Am., 57, 667 (1975).

2. D.S. Stevens and H.F. Tiersten, "An Analysis of Doubly-Rotated Quartz Resonators Utilizing Essentially Thickness Modes with Transverse Variation," J. Acoust. Soc. Am., 79, 1811 (1986).

3. H.F. Tiersten and R.C. Smythe, "An Analysis of Contoured Crystal Resonators Operating in Overtones of Coupled Thickness Shear and Thickness Twist," J. Acoust. Soc. Am., 65, 1455 (1979),

4. C.J. Wilson, "Vibration Modes of AT-Cut Convex Quartz Resonators," J. Phys., D7, 2449 (1974).

5. This approximation is discussed and explained in the sentence after Eq.(86) of Ref.2 along with the discussion of the reduction of Eq.(25) to Eq.(27) of Ref.2.

6. The equation for transversely varying thickness modes, i.e., Eq.(2.1) for attached electrodes, is readily modified to account for the gap simply by employing the $\rho\omega_n^2$ and F_n obtained from the pure thickness solution when there is a gap in place of those quantities obtained from the pure thickness solution for attached electrodes.

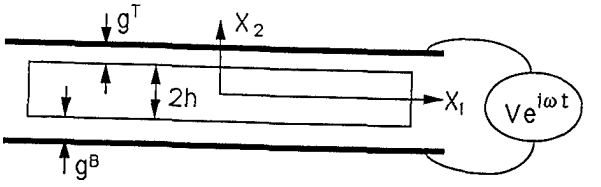


Figure 1 Flat Plate Between Electrodes with Asymmetric Air Gap

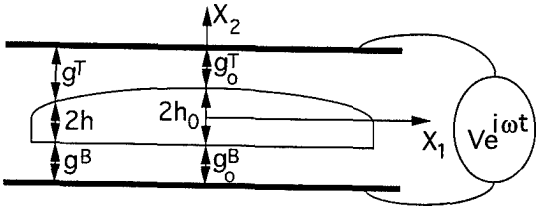


Figure 2 Contoured Plate Between Electrodes with Asymmetric Air Gap

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Lee's Plate Equations with New Correction Factors and Analysis of Spurious Vibrations of Rectangular AT-Cut Quartz Plates

Hitoshi Sekimoto, Daisuke Tajima, Yasuaki Watanabe and Akio Ishizaki

Department of Electronics and Information Engineering, Faculty of Engineering, Tokyo Metropolitan University, Hachioji, 192-03, Tokyo, Japan.

Abstract

Strip AT-cut quartz resonators have been recently developed and widely used in frequency applications because of their small size. Many spurious resonances can be supported in a rectangular AT-cut plate. Therefore, the analysis of the influence of both the X and Z' dimensions of the plate upon the spurious characteristics is necessary to design the strip resonators. However, it is very difficult to exactly solve three dimensional equations of elasticity. The practical analysis of spurious characteristics has been carried out on the basis of approximate two dimensional equations. In this paper, the applicability of Lee's two dimensional equations to AT-cut plates is examined. The resonant frequencies are calculated as a function of X-length, and are compared to Koga's experimental data. The result shows that new correction factors should be introduced in the equations in order to adjust the approximate values of wave numbers for the thickness-flexure waves propagating in X and Z' directions.

1. Introduction

Strip AT-cut quartz resonators operating in the fundamental thickness-shear mode have been recently developed and widely used in frequency applications because of their small size. In addition to the fundamental thickness-shear resonance, many spurious resonances can be supported in a rectangular AT-cut plate. Most of these resonances belong to the modes with phase reversals along both the X and Z' axes, and hence are electrically weak. However, even a weak mode can affect the fundamental thickness-shear mode when the frequencies of two modes approach each other. The mode coupling generally results in the large degradation of the quality of a strip AT-cut resonator. Because of this, there is a need for the theoretical study of spurious vibrations which depend on both the X and Z' dimensions of a rectangular plate.

At present it is still difficult to exactly solve three-dimensional equations of elasticity. It is more practical that the analysis of spurious vibrations of a plate is carried out on the basis of approximate two-dimensional equations. There are two sets of two-dimensional plate equations. One have been derived by Mindlin [1], in which mechanical displacements were expanded in power series of thickness coordinate of the plate. In the other, Lee [2] employed Fourier series expansion. The difference between two sets is small to the first order of approximation which includes the fundamental thickness-shear mode. In this paper, Lee's equations are applied to the analysis of spurious modes with phase reversals along both the X and Z' axes, and the accuracy of these equations is examined.

The resonant frequencies are calculated as a function of X-length, and are compared to Koga's experimental data [4]. The result shows that new correction factors should be introduced in the equations in order to adjust the approximate values of wave numbers for the thickness-flexure waves propagating in X and Z'directions.

2. Lee's Plate Equations

Consider a rectangular AT-cut quartz plate of mass density ρ , length 2a, thickness 2b, width 2c as shown in Fig.1, where the x_1, x_2 and x_3 coordinates are chosen as the X, Y' and Z' axes, respectively. In Lee's plate equations [2], the components of mechanical displacement are expanded in an infinite series with their thickness-dependence expressed by trigonometric functions, as follows.

$$u_j = \sum_{n=0} u_j^{(n)}(x1,x3)\cos[n\pi(b-x2)/2b] \quad (1)$$

Since we are interested in the fundamental thickness-shear mode and its spurious modes, only six components of displacement with n=0 and 1 are kept in the series. Also, we can omit electric terms from the equations because of small piezoelectric coupling of quartz. Then the plate equations of first-order approximation reduce to

$$\begin{aligned} &T_{ik,i}^{(0)} + 2\rho\omega^2 u_k^{(0)} = 0, \\ &T_{i2,i}^{(0)} + 2\rho\omega^2 u_2^{(0)} / \alpha_2^2 = 0, \\ &T_{ij,j}^{(1)} - \left(\frac{\alpha_1}{2b}\right)\overline{T}_{2,j}^{(1)} + \rho\omega^2 u_j^{(1)} = 0 \\ &(k=1,3,\ i,j=1,2,3) \end{aligned} \quad (2)$$

where ω is the radian frequency, α_1 and α_2 are Lee's correction factors, and $T_{ij}^{(n)}$ and $\overline{T}_{ij}^{(n)}$ are the two-dimensional components of stress. For the present approximation, the constitutive relations between stress and strain are

0-7803-2500-1/95 $4.00 © 1995 IEEE

$$T_{ij}^{(0)} = 2c_{ijkl}S_{kl}^{(0)} + \alpha_1\left(\frac{4}{\pi}\right)c_{ijkl}\bar{S}_{kl}^{(1)},$$
$$T_{ij}^{(1)} = \alpha_2^{\mu_{ij}+\mu_{kl}}c_{ijkl}S_{kl}^{(1)}, \quad (3)$$
$$\bar{T}_{ij}^{(1)} = c_{ijkl}\bar{S}_{kl}^{(1)} + \alpha_1\left(\frac{4}{\pi}\right)c_{ijkl}S_{kl}^{(1)}$$

in which c_{ijkl} are elastic constants and $\mu_{ij} = \cos^2\left(\frac{ij\pi}{2}\right)$, (i,j,k,l,=1,2,3, n = 0,1) The two-dimensional strains $S_{ij}^{(n)}$ and $\bar{S}_{ij}^{(n)}$ are expressed in terms of $u_j^{(n)}$ by

$$S_{ij}^{(n)} = (u_{j,i}^{(n)} + u_{i,j}^{(n)})/2,$$
$$\bar{S}_{ij}^{(1)} = \left(\frac{\pi}{4b}\right)(\delta_{2i}u_j^{(1)} + \delta_{2j}u_i^{(1)}) \quad (4)$$

In above equations there are two correction factors which are to compensate for incorrect approximation contained in the dispersion relations of straight-crested thickness-flexural waves. The first factor α_1 is for the flexural branch in the region of small wavenumber and low frequency, and the second α_2 for the flexural branch in large wavenumber and high frequency. In Lee's paper [2], α_1 is set to $\pi/\sqrt{8}$, and α_2 is determined so that the phase velocity of the flexural branch obtained from the approximate solutions approaches the surface wave velocity. For the straight-crested thickness-flexural wave propagating in the X direction, Bechmann's elastic constants [3] yield a value of $\alpha_2 = 0.954$.

To examine the accuracy of the equations with $\alpha_1 = \pi/\sqrt{8}$ and $\alpha_2 = 0.954$, the resonant frequencies of a rectangular AT-cut plate (2b = 1.650mm, c/b = 16.366) were calculated in the vicinity of the fundamental thickness-shear mode, as a function of X-length, and compared to Koga's experimental data [4]. Koga's data have included many spurious modes with phase reversals along both the X and Z' axes. These resonances were detected through a pair of metal plates, with which the main faces of an AT-cut plate were fully covered. In calculation, assuming the impressed electric field was uniform over the AT-cut plate, the components of displacement are taken in the following symmetry forms

$u_1^{(0)}$: even in x_1 and odd in x_3,
$u_2^{(0)}$ and $u_3^{(0)}$: odd in x_1 and even in x_3,
$u_1^{(1)}$: even in both x_1 and x_3,
$u_2^{(1)}$ and $u_3^{(1)}$: odd in both x_1 and x_3.

The functional dependence of displacements on x_1 and x_3 means that only the thickness-flexural, face-shear, face-flexural and antisymmetric face-extentional modes are included in the analysis of spurious vibrations. The calculation of frequencies follows the numerical procedure previously developed by the authors [5].

The results of comparison are shown in Fig.2, where the fundamental thickness-shear frequency is about 1.017 MHz, and the black and gray dots represent the calculated and measured data, respectively. There are many discrepancies between calculated and measured frequencies. We inferred that this disagreement was due to the improperly selected value of α_2 because it could force the frequencies of thickness-flexural modes to move up or down. Fig.3 shows the dispersion curves of thickness-flexure(F) and thickness-shear(TS) branches for straight-crested waves propagating in the X or Z' direction, where Ω is the normalized frequency and ψ_i is the normalized wavenumber along x_i ($\Omega = (2\omega b/\pi\sqrt{(\rho/c_{66})}$, $\psi_i = 2\beta_i b/\pi$, β_i: wavenumber). In the figure, the black and gray lines represent the approximate and exact curves, respectively. It can be seen that the approximate dispersion curves near Ω =1 of thickness-flexural branches greatly diverge from the exact curves. Therefore we judge that since our purpose is to analyze the spurious modes coupled to the fundamental thickness-shear mode, α_2 is to be determined from the wavenumbers of thickness-flexural waves at Ω = 1 rather than at very high frequency. Furthermore, in order to adjust the inaccurate wavenumbers in both cases of X and Z' propagations, additional correction factors should be introduced into the plate equations.

3. Introduction of New Correction Factors

The predominant displacement components are $u_2^{(0)}$ and $u_1^{(1)}$ for the straight-crested thickness-flexural wave propagating in the X direction, and $u_2^{(0)}$ and $u_3^{(1)}$ for the thickness-flexural wave of Z' propagation. Hence it is reasonable that each of the inertia terms of $u_1^{(1)}$ and $u_3^{(1)}$ is multiplied by a new correction factor. The modification of inertia terms is simple and can effectively adjust the wavenumbers of straight-crested thickness-flexural waves.

We introduce not one but two factors into the plate equations. Those are also inserted in the constitutive relations, in a manner similar to that adopted for α_2, so as to unchange the cutoff frequencies of two fundamental thickness-shear modes. On the basis of these considerations, the plate equations may be rewritten in the following form.

$$T_{ik,i}^{(0)} + 2\rho\omega^2 u_k^{(0)} = 0,$$
$$T_{i2,i}^{(0)} + 2\rho\omega^2 u_2^{(0)}/\alpha_2^2 = 0,$$
$$T_{i1,i}^{(0)} - \left(\frac{\alpha_1}{2b}\right)\bar{T}_{12}^{(1)} + \rho\omega^2\kappa_6^2 u_1^{(1)} = 0, \quad (5)$$
$$T_{i2,i}^{(0)} - \left(\frac{\alpha_1}{2b}\right)\bar{T}_{22}^{(1)} + \rho\omega^2 u_2^{(1)} = 0,$$
$$T_{i3,i}^{(0)} - \left(\frac{\alpha_1}{2b}\right)\bar{T}_{23}^{(1)} + \rho\omega^2\kappa_4^2 u_3^{(1)} = 0$$
(k=1,3, i=1,2,3)

where κ_4 and κ_6 are the new correction factors. Accordingly $T_p^{(0)}$ and $\bar{T}_p^{(1)}$ in the constitutive relations are replaced by

$$T_p^{(0)} = 2c_{pq}S_p^{(0)} + \alpha_1\left(\frac{4}{\pi}\right)c_{pq}\bar{S}_q^{(1)},$$
$$\bar{T}_p^{(1)} = \kappa_p^a\kappa_q^b c_{pq}\bar{S}_q^{(1)} + \alpha_1\left(\frac{4}{\pi}\right)c_{pq}S_q^{(0)} \quad (6)$$

(p,q=1~6) in which a=b=1 for p,q=4,6 and a=b=0 for p,q≠4,6.

We need three conditions for unique determination of κ_4, κ_6 and α_2. The wavenumbers of straight-crested thickness-flexural waves at Ω =1 give two conditions on the factors. The remaining condition can be extracted from the thickness-shear branch shown in Fig.3. In this paper, the curvature at Ω =1 for X propagation was chosen as the third condition. After κ_6 and α_2 were determined by matching the exact and

approximate solutions for the straight-crested waves propagating along X, κ_4 was adjusted for the thickness-flexural wave propagating along Z'. The assigned values are κ_4 = 1.046, κ_6 = 1.016 and α_2 =0.918. The approximate dispersion curves with these factors agree quite well with the exact curves on Fig.3.

4. Numerical Example of Mode Chart

The calculation of resonant frequencies was carried out again on the same rectangular AT-cut plate (2b=1.650mm, c/b=16.366) as we treated in Sec.2. The correction factors determined above were used in the calculation. Our present purpose is to see how the earlier mode chart (Fig.2) is improved by the introduction of additional correction factors. Fig.4 shows the comparison between calculated and measured frequencies. The calculation is matched to the experiment [4] well, although a few discrepancies remain still. This close agreement verifies our correction factors. It also means that most of spurious resonances in the mode chart are composed of the vibrations which depend on both X and Z' dimensions, since spurious modes associated with the straight-crested waves are very few.

The vibrational patterns for resonances were calculated to identify their modes. Many of the patterns were very complicated due to the strong coupling of all displacement components. The resonances labeled ① through ⑫ in Fig.4 could be distinguished for the relatively simple patterns as shown in Fig.5. The figure depicts the two-dimensional shapes of predominant displacement components in the x_1-x_3 plane, and the black lines represent the nodal lines. The resonant modes ①-⑫ are designated by the following mode-symbols.

TS-1,0: the fundamental thickness-shear mode,

TF_jT_k-n,m: the twist overtone mode along x_k of thickness-flexure vibration in the x_2-x_j plane,

FF-n,m: the face-flexure and antisymmetric face-extentional modes in the x_1-x_3 plane.

FS-n,m: the face-shear mode in the x_1-x_3 plane,

where n and m represent the number of half waves in x_1 and the number in x_3, respectively. Among these spurious modes, the twist overtones TF_3T_1 along x_1 of thickness-flexural vibrations have not specified before, while the twist overtones TF_1T_3 along x_3 have been discussed by Mindlin and Spencer [1].

5. Conclusions

In this paper, Lee's plate equations were applied to the analysis of spurious modes in rectangular AT-cut plates. In order to improve the accuracy for the thickness-flexure waves propagating in X and Z' directions, two additional correction factors were introduced into the equations. The resonant frequencies calculated with a new set of correction factors were matched to the measured data well. This close agreement verified the new correction factors. It also clarified that most of spurious resonances in AT-cut plates were composed of the thickness-flexure, face-shear, face-flexure and antisymmetric face-extentional vibrations.

The values of correction factors presented here are independent of plate size. We believe that they are applicable to other AT-cut plates with different X and Z' dimensions.

References

[1] R. D. Mindlin and W. J. Spencer:"Anharmonic,thickness-twist overtones of thickness-shear and flexural vibrations of rectangular,AT-cut quartz plates",J.Acoust.Soc.Amer., 42,6,pp.1268-1277(1967)

[2] P. C. Y. Lee and J. P. Hou:"A two-dimensional theory for high-frequency vibrations of piezoelectric crystal plates with or without electrodes",J.Appl.Phys.,vol.61,No.4,pp.1249-1262(1987)

[3] R. Bechmann:"Elastic and piezoelectric constants of alpha-quartz",Phy.Rev.,110,p.1060(1958)

[4] I. Koga:"Radio-frequency vibrations of rectangular AT-cut quartz plates",J.Appl.Phys.,34,8,pp.2357-2365(1963)

[5] H. Sekimoto, Y. watanabe and M. Nakazawa:"Forced vibrations of thickness-flexure, face-shear and face-flexure in rectangular AT-cut quartz plates",Proc.1992 IEEE FCS, pp.532-536(1992).

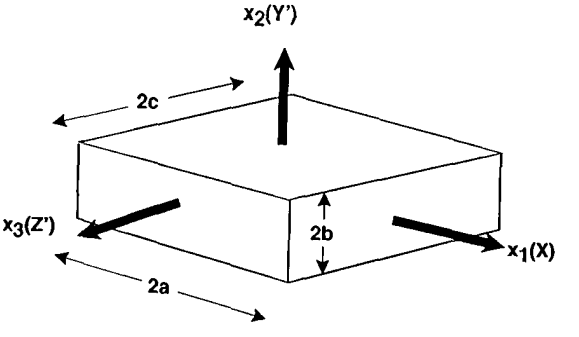


Fig.1 A rectangular AT-cut quartz plate

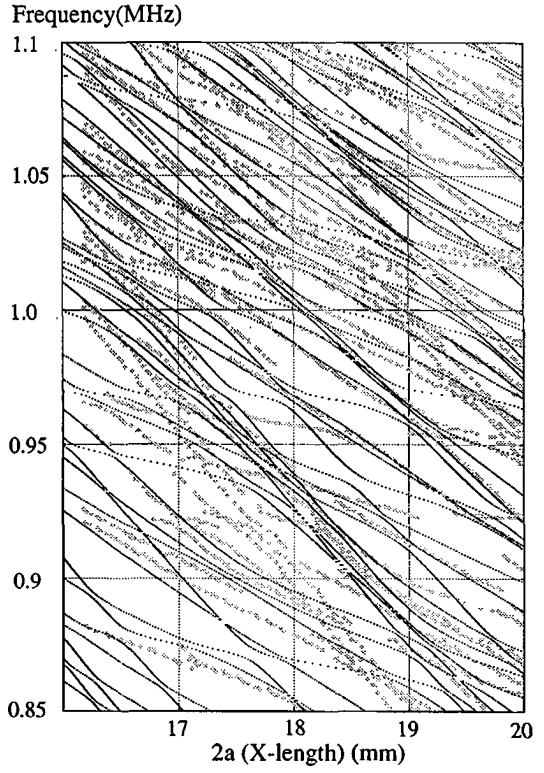


Fig.2 Comparison of calculated and measured resonant frequencies for a rectangular AT-cut quartz plate (thickness:2b=1.650mm, Z'-width:2c=27.004mm). The gray dots represent measured data and the black dots represent calculated data with α_2=0.954.

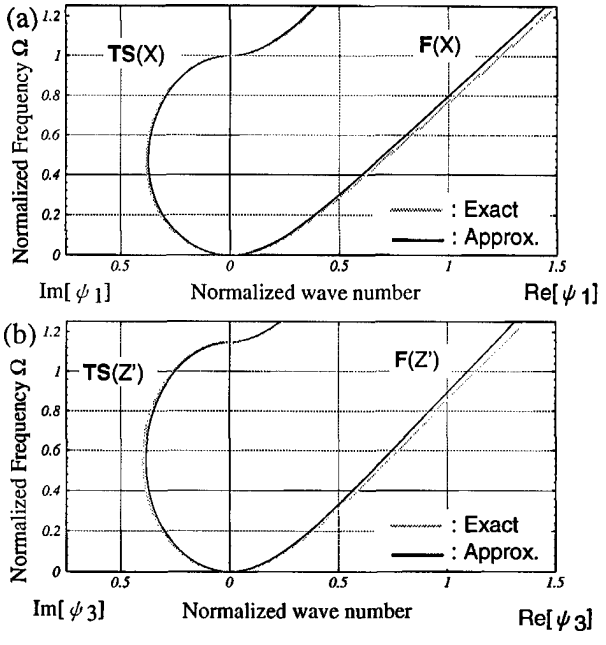


Fig.3 Dispersion curves of thickness-flexure(F) and thickness-shear(TS) branches for straight-crested waves propagating in (a) the X direction and (b) the Z' direction. The gray curves represent exact data and the black curves represent calculated data with α_2=0.954.

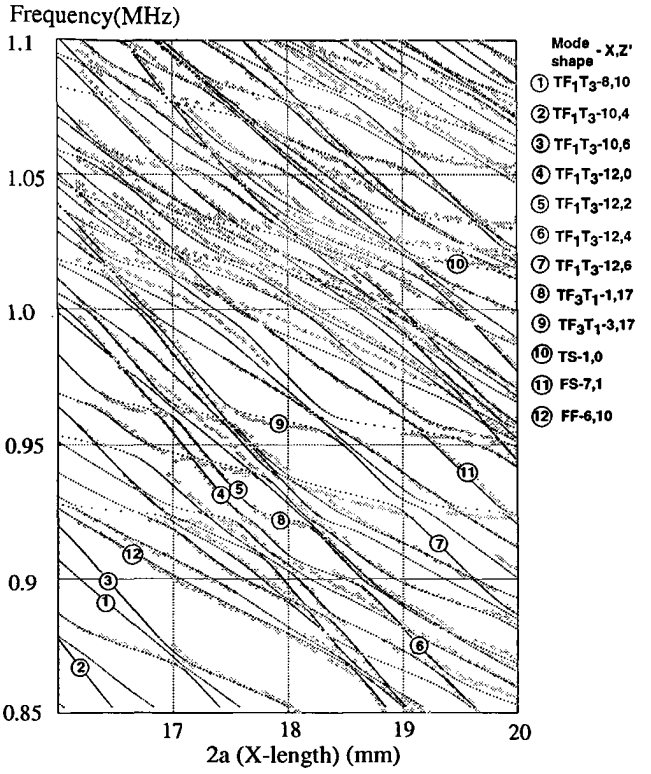


Fig.4 Comparison of calculated and measured resonant frequencies for a rectangular AT-cut quartz plate (thickness:2b=1.650mm, Z'-width:2c=27.004mm). The gray dots represent measured data and the black dots represent calculated data with α_2=0.918,κ_4=1.046 and κ_6=1.016.

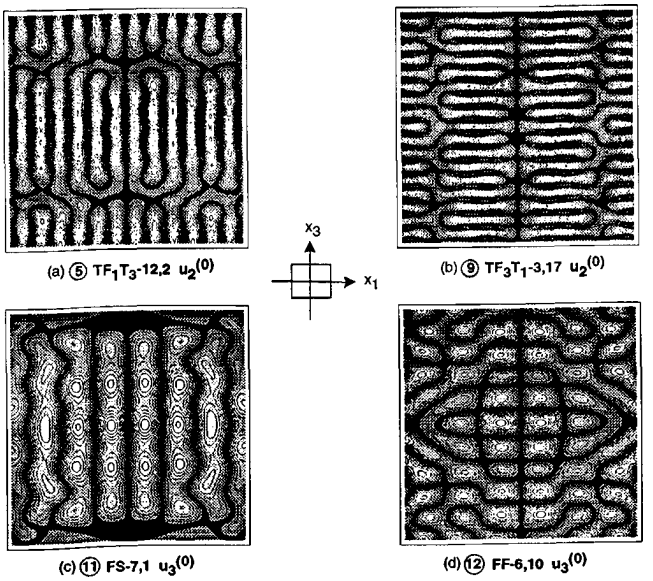


Fig.5 The calculated vibration patterns of predominant displacement components

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

EIGENFUNCTION EXPANSIONS IN STUDIES OF MULTIMODE AND TRANSIENT VIBRATIONS OF PIEZOELECTRIC TRANSDUCERS

Oleg Yu. Zharii

Department of Theoretical and Applied Mechanics, Kiev University,
64 Vladimirskaya Street, Kiev 252017, Ukraine

Abstract — The eigenfunction expansion method in dynamic elestroelasticity suggested earlier by the author, is applied to problems of electric discharge of piezoelectric bodies. Three problems are considered: (i) discharge of a piezoceramic rod, electrically loaded by a static potential difference; (ii) single discharge of the rod excited by harmonic mechanical loadings and (iii) periodic sequence of discharges under harmonic mechanical excitation.

Introduction

The expansion of transient wave fields in a series of eigenfunctions is one of the most elegant general theorems of classical elastodynamics. The simple physics that is revealed in the method of explicit expansions in normal modes, makes its use in transient problems preferable to other methods, for example, to integral transform techniques.

A general scheme of eigenfunction expansions in the theory of vibrations of piezoelectric transducers was developed and thoroughly discussed in the author's work [1]. This scheme provides an alternative to the Laplace integral transform technique in treating transient problems. In addition, eigenfunction expansions are a natural mathematical means for analyzing the electromechanical conversion efficiency using the concept of dynamical electromechanical coupling factor (EMCF) [2]. Several applications of the general technique are discussed in [1].

In this paper, after brief discussion of a refinement of the general theory, we apply eigenfunction expansions to problems of electric discharge of piezoceramic rods. The phenomenon of discharge consists in instantaneous release of a fraction of electric energy stored in the volume of a piezoelectric body. As a result, electric potential and mechanical stresses in the volume change and stress waves arise. Investigation of the state after discharge is important for finding time instants when following discharges may occur (the potential difference becomes high enough for breakdown of the spark gap).

Eigenfunction expansions

According to the eigenfunction expansions method [1], general solutions of initial boundary-value problems for mechanical displacements and electric potential have the form

$$u_i(x,t) = \sum_{m=1}^{\infty} U_i^{(m)}(x) q_m(t),$$

$$\psi(x,t) = \varphi(x,t) + \sum_{m=1}^{\infty} \Psi^{(m)}(x) q_m(t), \qquad (1)$$

where $U_i^{(m)}$ and $\Psi^{(m)}$ are eigenfunctions, x is the three-dimensional coordinate, φ is the uncoupled potential and q_m are the time functions determined from initial and boundary conditions. The general scheme of eigenfunction expansions can be somewhat refined compared to [1]. Namely, instead of cumbersome procedure described in Sec. I.C, for determination of q_m we may use the Betty's identity

$$\int_V \rho\left(\ddot{u}_i^{(1)} u_i^{(2)} - \ddot{u}_i^{(2)} u_i^{(1)}\right) dV = \int_V \left(F_i^{(1)} u_i^{(2)} - F_i^{(2)} u_i^{(1)}\right) dV + \int_A \left(T_{ij}^{(1)} u_i^{(2)} - T_{ij}^{(2)} u_i^{(1)}\right) dA + \int_A n_i \left(D_i^{(1)} \psi^{(2)} - D_i^{(2)} \psi^{(1)}\right) dA, \qquad (2)$$

where $u_i^{(1)}$, $\psi^{(1)}$ and $u_i^{(2)}$, $\psi^{(2)}$ are two arbitrary solutions of elastodynamics equations and F_i are corresponding body forces.

Substituting into this identity $u_i^{(1)} = u_i(x,t)$—the solution looked for, and $u_i^{(2)} = U_i^{(m)}(x) e^{i\Omega_m t}$, we employ the boundary conditions of both the original and the eigenvalue problems. Resulting formulae for q_m and φ are exactly the same as in [1].

Electric discharge of piezoelectric bodies

High voltage generated in piezoelectric samples due to piezoelectric effect, provides an opportunity to use

0-7803-2500-1/95 $4.00 © 1995 IEEE

them as impulse spark generators. To improve characteristics of these generators, it is necessary to estimate upper physical limits for energy that can be released under discharge. It is also necessary to study intensive mechanical waves arising after discharge. Under certain circumstances, these waves can cause fracture of ceramic specimens [3].

Exact modeling of the discharge phenomenon gives rise to a nonlinear model of the spark gap [2] and the resulting boundary-value problem becomes very complicated. However, a precise consideration of the discharge phenomenon has shown that the duration of discharge is so short that the mechanical waves that arise can advance only a small distance in a piezoelectric body. It means that displacements, velocities and deformations in internal points of a body remain unchanged during the discharge.

The fundamentals of the general linear theory of instantaneous discharge of piezoelectric bodies were developed in the paper [4]. From viewpoint of eigenfunction expansions (1), continuity of displacements and velocities during the discharge taking place at $t = t_0$ means that time functions q_m and their derivatives are continuous at $t = t_0$. Therefore, instantaneous changes in the electroelastic field occur thanks to changes in the uncoupled potential φ.

The general formulation of the problems of discharge, as specific initial boundary-value problems of dynamic electroelasticity, includes formulae for determination of unknown currents through discharge electrodes in the time interval containing a discharge. The difference of electric fields after and before discharge is an electrostatic field, and due to that, the relation between currents through electrodes and instantaneous changes of potentials on them ΔV_ν can be written as

$$\frac{d}{dt}\int_{A_\mu} n_i D_i \, dA = -\sum_\nu c_{\mu\nu}\Delta V_\nu \cdot \delta(t - t_0), \quad (3)$$

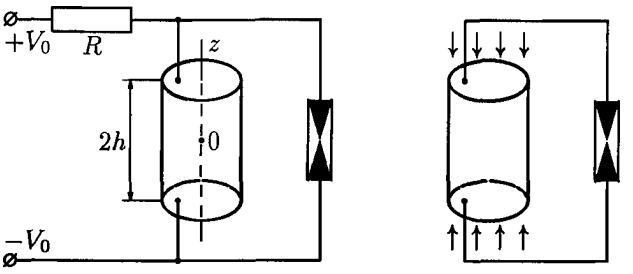
where $c_{\mu\nu}$ is an $N \times N$ static capacitance matrix of an N-electroded body [1, 5] and the sum is taken over all electrodes not connected to voltage generators [1].

Further we use the developed theory in conjunction with eigenfunction expansions to specific problems of discharge.

Discharge of an electrically loaded rod

Consider a longitudinally polarized rod of piezoceramic material having length $2h$ (Fig. 1). To investigate its behavior in a one-dimensional approximation, we use constitutive equations (index z for all variables is omitted)

$$S = s_{33}^E T + d_{33} E,$$
$$D = \epsilon_{33}^T E + d_{33} T, \quad (4)$$



(a) (b)

Fig. 1. Discharge of (a) electrically loaded and (b) mechanically loaded rod.

the stress equation of motion

$$\frac{\partial T}{\partial z} = \rho \frac{\partial^2 u}{\partial t^2}, \quad (5)$$

the electrostatic equations

$$\frac{\partial D}{\partial z} = 0, \quad E = -\frac{\partial \psi}{\partial z}, \quad (6)$$

and the Cauchy relation

$$S = \frac{\partial u}{\partial z}. \quad (7)$$

Eliminating from the above equations all quantities except $u(z,t)$, and $\psi(z,t)$, we obtain the second-order equations

$$\frac{\partial^2 u}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2},$$
$$\frac{\partial^2 \psi}{\partial z^2} = \frac{k_{33}^2}{d_{33}(1 - k_{33}^2)}\frac{\partial^2 u}{\partial z^2}, \quad (8)$$

where $k_{33}^2 = d_{33}^2/(s_{33}^E \epsilon_{33}^T)$ is the longitudinal EMCF and $c = \left[\rho s_{33}^E (1 - k_{33}^2)\right]^{-1/2}$ is the rod velocity [6].

Let the mechanically free rod is quasistatically loaded with an electric circuit containing a very large resistor (Fig. 1,a). Before $t = 0$, inside the rod the distribution of electroelastic field variables agrees with the static variant of equations (8) and boundary conditions $T = 0$, $\psi = \pm V_0$ at $z = \pm h$, namely

$$u_0 = -d_{33} V_0 \frac{z}{h}, \ \psi = V_0 \frac{z}{h}, \ T_0 = 0, \ D_0 = -\epsilon_{33}^T \frac{V_0}{h}. \quad (9)$$

At $t = 0$ the discharge mechanism operates and allows current to pass through the discharger. At the first time instant after discharge ($t = +0$), we have

$$\psi|_{z=h} - \psi|_{z=-h} = 0, \quad (10)$$

consequently $E = 0$. The electric displacement at $t > 0$ is constant (only an infinitesimal current passes through the resistor),

$$D = -k_{33}^2 \epsilon_{33}^T \frac{V_0}{h}. \quad (11)$$

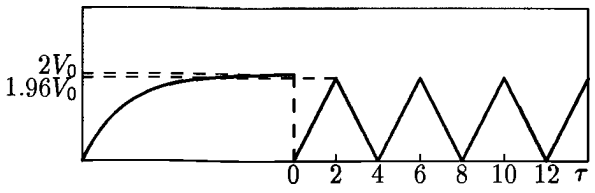


Fig. 2. Potential difference $\Delta\psi$ before and after discharge in an electrically loaded rod.

To investigate the motion of the rod we solve equations (8) at initial conditions

$$u|_{t=0} = u_0, \quad \left.\frac{\partial u}{\partial t}\right|_{t=0} = 0, \tag{12}$$

also using mechanical boundary conditions for free rod $T = 0$ at $z = \pm h$. Electric displacement is given by (11).

According to the general scheme, we first find the corresponding eigenfunctions and eigenfrequencies,

$$U^{(m)} = \frac{\sin \kappa_m z}{\sqrt{\rho h}}, \quad \Psi^{(m)} = \frac{k_{33}^2}{d_{33}(1-k_{33}^2)} U^{(m)},$$
$$\Omega_m = \kappa_m c, \quad \kappa_m = \frac{(2m-1)\pi}{2h}. \tag{13}$$

The uncoupled potential, determined in accordance with the general theory, is

$$\varphi = \frac{k_{33}^2}{1-k_{33}^2} V_0 \frac{z}{h}. \tag{14}$$

Note, that in this problem, the potential eigenfunction $\Psi^{(m)}$ form an orthogonal and complete set in $(0, h)$ together with $U^{(m)}$ (in general, it is not the case). Nevertheless, the uncoupled potential is not zero. So, it arises not due to incompleteness, as was stated in [7, p. 134], but due to another reasons [1].

Calculating the time functions $q_m(t)$ from the initial and boundary conditions on the base of general formulae, we find

$$q_m(t) = d_{33} V_0 h \sqrt{\frac{\rho}{h}} \frac{8}{\pi^2} \frac{(-1)^m}{(2m-1)^2} \times \left[k_{33}^2 + (1-k_{33}^2)\cos \Omega_m t\right] \tag{15}$$

and the solution of the problem is represented by formulae (1), (13), (14) and (15).

The sums of the series obtained can be represented in a closed form. Introducing dimensionless coordinates $\xi = z/h$, $\tau = ct/h$, we have

$$u = -k_{33}^2 d_{33} V_0 \xi - \tfrac{1}{2} d_{33}(1-k_{33}^2) V_0 [\alpha(\xi+\tau) + \alpha(\xi-\tau)],$$
$$\psi = \tfrac{1}{2} k_{33}^2 V_0 [2\xi - \alpha(\xi+\tau) - \alpha(\xi-\tau)]. \tag{16}$$

Here α is the continuous function with period $T = 4$,

$$\alpha(\xi) = \begin{cases} -\xi - 2, & -2 < \xi < -1, \\ \xi, & -1 < \xi < 1, \\ -\xi + 2, & 1 < \xi < 2. \end{cases} \tag{17}$$

In Fig. 2, the potential difference $\Delta\psi = \psi|_{z=h} - \psi|_{z=-h}$ is presented. It is seen that for high-efficient piezoceramic materials (in our example, for PZT-4 with $k_{33}^2 = 0.49$), it can reach almost the value $2V_0$ and a considerable part of the wave energy can be released during the second discharge at $\tau = 2$.

Discharge under harmonic mechanical excitation

Now consider the same rod subjected to harmonic mechanical loading (Fig. 1,b),

$$T|_{z=\pm h} = p_0 \cos \omega_0 t. \tag{18}$$

At $t = 0$, when the potential difference reaches the maximal value, an electric discharge takes place, and immediately after it eq.(10) is satisfied.

It can be shown that the electric displacement after discharge is

$$D = \epsilon_{33}^T (1-k_{33}^2) \frac{V_0}{h} H(t), \tag{19}$$

where V_0 is the amplitude of potential at $z = h$ under steady-state vibrations and H is the Heaviside step function.

The displacement and electric potential are the solutions of electroelastic equations (8) under boundary conditions (18) and (19) and subjected to conditions of continuity of displacements and velocities at $t = 0$. The method of finding the solution is the same as in the preceding section.

The plot of the potential difference for the frequency of excitation that equals to the first resonant frequency of the rod, $\Omega_1 = \pi c/2h$,

$$\Delta\psi = 2V_0 \{\cos \pi\tau - [1 - k_{33}^2 - k_{33}^2 \alpha(1-\tau)] H(\tau)\} \tag{20}$$

is depicted in Fig. 3. Here $2V_0$ is the amplitude of the potential difference under steady-state vibrations.

The main difference with Fig. 2 is that the next time $|\Delta\psi|$ takes the value $2V_0$ when the sign of $\Delta\psi$ it is opposite to that at the first discharge. It occurs at $t = 0.59$ (Fig. 3). It is explained by the fact that constant charges (19) result in a constant addend to electric potential difference.

So, under the mechanical excitation, successful discharges, if they occur under the same value of $|\Delta\psi|$ regardless of sign of $\Delta\psi$, take place with switching signs of the potential difference.

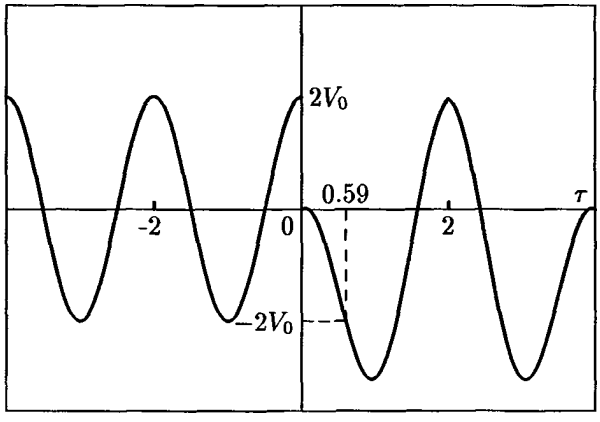


Fig. 3. Potential difference $\Delta\psi$ in a mechanically loaded rod before and after discharge.

Periodic discharges of the rod

If discharges take place each time under the same value of the modulus of the potential difference, after several discharges, a periodic regime is established. In this regime, the time interval between successive discharges is a constant.

Consider the problem for a periodic regime with a period $2\mathcal{T}$, that includes two discharges with opposite (as follows from the previous section) signs of the potential difference $\Delta\psi$ before discharges, $-2V$ at $t = \mathcal{T}$ and $2V$ at $t = 2\mathcal{T}$, and V may be not the same as the maximal value of potential V_0.

To ensure physical grounds for establishment of the periodic regime, we must consider the internal energy losses in the piezoceramic rod.

Mechanical energy losses are taken into account in the following way. Eigenfunctions in the problem under consideration are determined from the solution of the problem of free vibrations of the rod with the electrodes open-circuited, $D \equiv 0$. In this problem, for harmonical excitation at resonant frequencies, we may use complex representations of stresses and deformations,

$$T = \mathrm{Re}\{\hat{T}e^{i\Omega_m t}\}, \quad S = \mathrm{Re}\{\hat{S}e^{i\Omega_m t}\}, \tag{21}$$

and write the equation of state in the form

$$\hat{T} = \frac{1}{s_{33}^D}\left(1 + \frac{i}{Q_m}\right)\hat{S}, \tag{22}$$

where $s_{33}^D = s_{33}^E(1 - k_{33}^2)$ is the elastic compliance under open-circuited electrodes and Q_m is the mechanical quality factor of the mth normal mode.

The same result (22) can be obtained, if instead of the first of the formulae (4) one will use the stress-strain relation in the form

$$T = \left(\frac{1}{s_{33}^E} + \eta_m\frac{\partial}{\partial t}\right)S - \frac{d_{33}}{s_{33}^E}E, \tag{23}$$

where viscosity coefficients η_m are related to Q_m by the equalities

$$\frac{1}{\eta_m} = s_{33}^D\Omega_m Q_m. \tag{24}$$

Using (23) instead of the first of equations (4) and repeating the derivation of the general scheme of eigenfunction expansions, one finds the following equations for normal mode amplitudes $q_m(t)$ $(m = 1, 2, \ldots)$

$$\ddot{q}_m + \frac{\Omega_m}{Q_m}\dot{q}_m + \Omega_m^2 q_m = \Phi_m(t), \tag{25}$$

and $\Phi_m(t)$ are determined in the usual way from boundary conditions on the base of formulae (29) of Ref. 1. An electric boundary condition is that for the electric displacement, namely $D = \pm\epsilon_{33}^T(1 - k_{33}^2)V/(2h)$ in intervals $0 < t < \mathcal{T}$ and $\mathcal{T} < t < 2\mathcal{T}$, respectively. This value was found from the conditions of periodicity which all characteristics of solution are to satisfy,

$$f(t + \mathcal{T}) = -f(t). \tag{26}$$

The functions $\Phi_m(t)$ are thus equal to

$$\Phi_m(t) = \frac{(-1)^m}{\sqrt{\rho h}}\frac{d_{33}V}{s_{33}^E}. \tag{27}$$

Functions q_m are determined from (25) in view of (27) and conditions of periodicity following from (26),

$$q_m(\mathcal{T}) = -q_m(0), \quad \dot{q}_m(\mathcal{T}) = -\dot{q}_m(0). \tag{28}$$

The complete solution of the problem is represented by the sum of eigenfunction expansions (1) and solutions of the steady-state problem. It has been found that the period of the discharge regime $2\mathcal{T}$ can take only discrete values that equal to an odd number of the external excitation period at the first resonant frequency,

$$2\mathcal{T} = (2n - 1)\frac{2\pi}{\Omega_1}, \tag{29}$$

where the integer n is hitherto unknown.

Omitting cumbersome details, we write the final formula for the output potential difference for $0 < t < \mathcal{T}$,

$$\Delta\psi = 2\Big\{V_0\cos[(2n-1)\pi\tau + \theta] - \frac{1}{2}V(1 - k_{33}^2) - k_{33}^2 V\frac{8}{\pi^2}\sum_{m=1}^{\infty}\frac{\cos(\alpha_{mn}\tau - 1/(2Q_m))e^{-\gamma_{mn}\tau}}{(2m-1)^2(1 - e^{-\gamma_{mn}})}\Big\}. \tag{30}$$

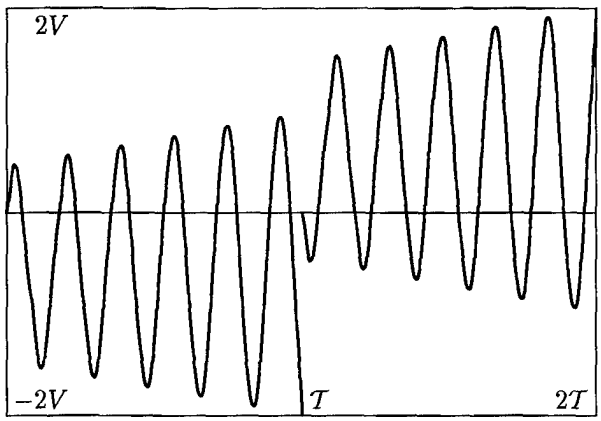


Fig. 4. Potential difference $\Delta\psi$ in the periodic discharge regime.

Here $\alpha_{mn} = (2m-1)(2n-1)\pi$ and $\gamma_{mn} = \alpha_{mn}/(2Q_m)$.

Note that the two unknown quantities in (30), θ and n, are determined from the single equation, $\Delta\psi_{\tau=0} = 0$. This equation can be reduced to

$$\theta = \arccos \frac{V}{V_0}\left\{\frac{1}{2}(1-k_{33}^2) + k_{33}^2\frac{8}{\pi^2}\sum_{m=1}^{\infty}\frac{1}{(2m-1)^2(1-e^{-\gamma_{mn}})}\right\}. \quad (31)$$

An attentive analysis of this equation shows that fortunately, it uniquely determines both θ and n. Quantities Q_m, as a rule, are of order of several dozens or hundreds, so for several first $n = 1, 2, \ldots$, the argument of the arccos function is greater that unity. This means, that for few low values of n the equation has no solution.

Starting from $n = 1$, we increase n successfully by one and the least integer at which the equation is solvable, will determine the period of the discharge regime. This value of n does exist. After that, the value of θ is found. A numerical example has been calculated for values $V/V_0 = 0.9$, $k_{33}^2 = 0.49$ and $Q_1 = Q_2 = \ldots = 25$. Thanks to rapid convergence of series in (30) and (31), it is enough to take only five first terms. So, normal modes of numbers sixth and higher do not affect the periodic discharge regime. This conclusion is important in view of the fact that the frequency applicability of the rod approximation is restricted to few lower eigenfrequencies.

For given input parameters, we found that in this case $n = 7$, i.e., the period of the discharge regime is seven times longer than the period of the external excitation. The plot of the potential difference is depicted in Fig. 4.

The number n is an increasing function of Q_m. For instance, at $Q_1 = Q_2 = \ldots = 50$, we obtain $n = 13$ for the same ratio V/V_0. For materials with higher mechanical quality factors, the amplitude of the maximal potential in the steady-state regime,

$$V_0 = \frac{8}{\pi^2}\frac{hd_{33}}{\epsilon_{33}^T}Q_1 p_0 \quad (32)$$

increases, but the value of energy lost during the discharge simultaneously increases,

$$W = \frac{1}{h}\epsilon_{33}^T(1-k_{33}^2)AV^2 \quad (33)$$

(A is the cross-sectional area). Therefore, it takes more time for the mechanical energy source to fill up the energy lost and due to that the time interval between discharges increases.

Conclusions

In this paper, we applied the eigenfunction expansion method to problems of electric discharge of piezoelectric transducers. The method has demonstrated its flexibility, especially in the problem of periodic discharges, where we used it for materials with internal energy losses and received the result with moderately complicated mathematics.

References

[1] O. Yu. Zharii, "Normal mode expansions in dynamic electroelasticity and their applications to electromechanical energy conversion," *J. Acoust. Soc. Amer.*, vol. 92, pp. 57–68, 1992.

[2] V. T. Grinchenko, A. F. Ulitko and N. A. Shul'ga, *Electroelasticity*, in *Mechanics of Coupled Fields in Elements of Structures*, vol. 5. Kiev, Naukova Dumka, 1989.

[3] G. Goodman, "Electrostrictively generated and modulated fracture in barium titanate ceramics," *J. Appl. Phys.*, vol. 35, pp. 2725–2726, 1964.

[4] O. Yu. Zharii and A. F. Ulitko, "Electric discharge of a piezoceramic rod under stationary mechanical excitation," *Math. Methods and Phys.-Mech. Fields*, No 13. pp. 57–63. Kiev, Naukova Dumka, 1981.

[5] W. R. Smythe, *Static and dynamic electricity*, New York: McGraw-Hill, 1968.

[6] An American National Standard: IEEE Standard on piezoelectricity, *IEEE Trans. Sonics Ultrason.*, vol. SU-31, No. 2, Pt. 2, pp. 1–55, 1984.

[7] R. Holland and E. P. EerNisse, *Design of resonant piezoelectric devices*, Cambridge: MIT, 1969.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ON THE ACCURACY OF PLATE THEORIES FOR THE PREDICTION OF UNWANTED MODES NEAR THE FUNDAMENTAL THICKNESS SHEAR MODE

Y-K YONG[1], Z. ZHANG[1] and J. HOU[2],

[1]Dept. of Civil & Environmental Engineering, Rutgers University, P.O. Box 909, Piscataway, NJ 08855-0909, [2]Allied Signal, Inc. TEL: (908) 445-3219 FAX: (908) 445-0577 EMAIL: yong@jove.rutgers.edu

Abstract

The first order plate theories with correction factors are generally assumed to predict accurately the plate modes which have half wave lengths greater than the plate thickness, and at frequencies up to twenty percent higher than the fundamental thickness shear frequency. This assumption is assessed by comparing the straight crested wave solutions of the plate theories with those of the three-dimensional elastic equations of motion. The frequency spectra for bandwidths of resonant frequencies versus the aspect ratio of length to thickness are compared for three sets of plate equations: The first order Mindlin plate equations, the third order Mindlin plate equations, and the third order Lee and Nikodem plate equations. The finite element results for a quartz SC-cut strip with free edges show that Mindlin's first order plate equations, and Lee and Nikodem's third order plate equations do not yield an accurate frequency spectra of the modes in the vicinity of the fundamental thickness shear mode, although the thickness shear mode itself is predicted accurately. The third order Mindlin plate equations without correction factors, on the other hand, predict well the frequency spectrum in the vicinity. The first order Mindlin plate theory is found to yield accurate frequency spectra for normalized frequencies less than 0.1, which is lower than previously assumed. At normalized frequencies greater than 0.1, deviations are seen in the frequency spectra, starting with the modal branches which are more steeply inclined.

I. Introduction.

The classical method of overcoming the difficulty of solving the three-dimensional elastic equations of motion for a plate with a pair of parallel free major surfaces has been the employment of approximate, two-dimensional equations of motion which already satisfy, at least approximately, the conditions of free faces so that they are analogous to those for free faces in the three-dimensional case. A solution of the approximate equations for the infinite plate contains, usually, only a small number of modes approximating the corresponding ones of the infinity of modes from the three-dimensional equations[1]. These approximate two-dimensional equations of motion are the so called plate equations which are derived using a series expansion of the thickness modes with various functions of the thickness coordinate followed by an integration through the thickness which eliminates one of the three independent spatial variables from the equations of motion. The functions of thickness coordinate that were proposed are: Power series by Mindlin[2], trigonometric functions by Lee and Nikodem[3,4], Legendre polynomials of thickness coordinate by Mindlin and Medick[5], and normal mode expansion by Peach[6]. These plate theories are technically important because they are useful for modeling high frequency vibrations of crystal plate resonators. They are also implemented into finite element codes since they are more economical and easier to use than three-dimensional finite elements. The use of

0-7803-2500-1/95 $4.00 © 1995 IEEE

three-dimensional elements for high frequency vibrations such as those in third overtone thickness shear plate resonators presents an extremely challenging task of identifying the modes of vibration, in addition to its inherently higher computational expense in terms of memory and time. The use of plate elements on the other hand is easier in that the thickness modes are described by a small number of functions of the thickness coordinate. The first order plate theories with correction factors are generally assumed to predict accurately the plate modes which have half wave lengths greater than the thickness of the plate, and at frequencies up to twenty percent higher than the fundamental thickness shear frequency[2]. The aim of this paper is to assess the accuracy of Mindlin's[2], and Lee and Nikodem's[3,4] plate theories in the frequency range of the fundamental thickness shear mode.

II. Plate theories.

The derivations of plate theories using various series expansions of the thickness modes are presented in detail in references 1-7, and will not be repeated here. We consider a quartz plate as shown in fig. 1 with free major surfaces, x_2 axis the normal direction of the major surfaces, and x_1 axis the rotated digonal axis. The thickness of the plate along the x_2 axis is $2b$, the length along the x_1 axis is $2a$, and the width is $2c$.

The series expansions for the displacements $u_i(x_1, x_2, x_3, t)$ $i = 1, 2, 3$ for Mindlin's[2] and Lee and Nikodem[3,4] plate theories are, respectively:

$$u_i(x_1, x_2, x_3, t) = \sum_{n=0}^{n=\infty} x_2^n u_i^{(n)}(x_1, x_3, t) \qquad (1)$$

and

$$u_i(x_1, x_2, x_3, t) = \sum_{n=0}^{n=\infty} u_i^{(n)}(x_1, x_3, t) \cos\frac{n\pi}{2}(1-\eta) \qquad (2)$$

where $\eta = \frac{x_2}{b}$ and the exponent n in the series expansions determines the order of the plate theories. For the first order plate theories, the series is truncated after the first two terms, so that n is set to one, while for the third order plate theories, n is set to three. For example, the series expansion for Mindlin's first order plate theory is from eqn. (1)

$$u_i(x_1, x_2, x_3, t) = u_i^{(0)}(x_1, x_3, t) + u_i^{(1)}(x_1, x_3, t). \qquad (3)$$

In essence, the series expansions yield displacements, strains, stresses, and equations of motion which are functions of only two independent spatial variables x_1 and x_3.

III. Straight Crested Wave Solutions.

The accuracy of plate equations is found by comparing their solutions against those of the three-dimensional elastic equations of motion. In this regard, we use the finite element solutions for straight crested waves in an SC-cut quartz strip with free edges. The straight crested wave solutions form a class of one-dimensional solutions of the two-dimensional plate equations, and therefore are approximate solutions of the three-dimensional problem with free surfaces at $x_2 = \pm b$. and $x_1 = \pm a$. It is reasonable to assume that whatever characteristics found in the straight crested wave solutions would necessarily apply also to the solutions of the two-dimensional plate equations themselves. It is computationally much easier to find the straight crested wave solution of the three-dimensional equations of motion, hence its use in this study to check the accuracy of the plate equations. We consider, in particular, waves straight crested in the x_3 direction, and travelling in the x_1-x_2 plane (length-thickness plane). Hence, the partial derivatives of all terms with respect to x_3 are set to zero:

$$\frac{\partial(\bullet)}{\partial x_3} = 0 \qquad (4)$$

Since the behavior of modes in the immediate vicinity of the operating thickness shear mode is of interest in the design of a thickness shear plate resonator, it is not sufficient to compare only the thickness shear frequency, but also the bandwidth of resonant frequencies in its immediate vicinity as a function of the aspect ratio of length to thickness. The graph of a bandwidth of resonant frequencies versus the aspect ratio of length to thickness is referred to simply as a frequency spectrum. In all the graphs of frequency spectra to be presented, the resonant frequency ω is normalized with respect to $\omega_1^2 = \frac{\pi^2 C_{66}}{4b^2\rho}$ so that $\Omega = \frac{\omega}{\omega_1}$, where Ω is the normalized resonant frequency, and the aspect ratio, *a/b*, is set to range from 31 to 33. The frequency spectra obtained using the plate equations are compared against those obtained using the three-dimensional equations of motion.

Four finite element programs are used in the study. The programs calculate the straight crested wave frequency spectra for an SC-cut quartz strip using, respectively, the first order Mindlin plate equations, the third order Mindlin plate equations, the third order Lee and Nikodem plate equations, and the three dimensional equations of motion. In the first order Mindlin plate equations, three correction factors $\kappa_1 = \kappa_2 = \kappa_3 = \frac{\pi}{\sqrt{12}}$ [2] are used. For calculation of resonant frequencies in the range of the fundamental thickness shear frequency no corrections factors are needed in the third order Mindlin plate equations. For Lee and Nikodem's third order plate equations, two correction factors $\alpha_1 = \frac{\pi}{\sqrt{8}}$ [7], and $\alpha_2 = 0.9395$ [8] are employed.

A one-dimensional cubic element is used in the finite element matrix equations for the plate theories. The element was found to be stable and to converge rapidly for high frequency calculations[9]. A 750 node mesh is used to calculate all the frequency spectra for the plate equations.

A two-dimensional, four-node, bilinear quadrilateral element[10] is used for the straight crested wave solutions of the three-dimensional elastic equations of motion. Figure 2 shows the straight crested wave solutions of the three-dimensional equations of motion for two element meshes with 18 elements in the thickness direction by 400 elements in the length direction, and 18 elements in the thickness direction by 750 elements in the length direction, respectively. One region in which the two frequency spectra appear to differ greater than the others is in the lower left hand corner in the neighborhood of resonant frequency 0.932 and aspect ratio 31.7: For a given aspect ratio, the frequencies from the two meshes differ by about 0.4%, and for a given frequency, the aspect ratios differ by about 0.4%. There are no visual differences found when the frequency spectra from meshes with 18x550 and 18x600 elements are compared with those of the 18x750 element mesh. Hence, the frequency spectra from the 18x750 element mesh is assumed to be the converged solutions to the straight crested wave problem of the three-dimensional elastic equations of motion; and they will be used to assess the accuracy of the plate theories. Another observation regarding the finite element modeling of high frequency vibrations may be made from fig. 2: The frequencies of the predominantly thickness shear mode and its anharmonics, labelled as TS-1, TS-2, and TS-3 on the relatively horizontal branches in the graph, converge earlier and faster than the predominantly flexure, face shear, or extensional modes represented by the steeply inclined branches. This is due to the larger number of predominant half waves in the latter modes of vibrations. These modes are generally classified as the unwanted modes in the quartz plate resonator when a "pure" fundamental thickness shear mode is desired as the operating mode of vibration.

IV. Accuracy of plate equations.

Figure 3 shows the comparison of frequency spectra obtained using Mindlin's first order plate theory and the three-dimensional equations of motion. Although the frequency of the fundamental thickness shear mode, TS-1, is well predicted (about 0.2% lower), the overall pattern of frequency spectrum for unwanted modes from the first order plate theory is not satisfactory. For example, when the aspect ratio is 31.4, Mindlin's first order plate theory predicts a "pure" thickness shear mode, while the three-dimensional equations predicts a thickness shear mode strongly coupled with flexure. Figure 4 is similar to fig. 3 except that Mindlin's third order plate theory is compared. This time the details of the overall pattern of frequency spectrum from the third order plate theory matches well with the three-dimensional equations. The frequency spectrum from the third order plate theory is slightly lower in frequency (less than 0.1%). Figure 5 is again similar to fig. 3, except that Lee and Nikodem's third order plate theory is compared. The fundamental thickness shear frequency is also well predicted, but the overall pattern of the frequency spectrum does not fit well with the three-dimensional equations.

The accuracy of the plate theories at other frequency ranges is investigated. It is found that the first order Mindlin plate theory, for a range of aspect ratios 31 to 33, yield accurate frequency spectra below a normalized frequency of about 0.1. At normalized frequencies greater than 0.1, the frequency spectra begins to deviate from the three-dimensional equations. Figure 6 shows the comparison of frequency spectra in the frequency range of 0.14 to 0.20, and deviations are observed in modal branches which are more inclined, while less deviations are observed in the less inclined modal branches. This phenomenon is also observed in higher frequency ranges.

V. Summary and Conclusion.

The accuracy of Mindlin's first and third order plate equations, and Lee and Nikodem's third order plate equations are assessed in terms of their frequency spectra for straight crested waves in an SC-cut quartz strip with free edges. The frequency spectra for various frequencies ranges, and an aspect ratio of 31 to 33 are compared against those obtained from the three-dimensional elastic equations of motion. The following observations are made:

1. Although Mindlin's first order plate equations, and Lee and Nikodem's third order plate equations predict well the frequency of the fundamental thickness shear mode, the pattern of the frequency spectra in its vicinity is not satisfactory for the unwanted modes. The third order Mindlin plate equations predicts satisfactory frequency spectra in this frequency range. Hence in this frequency range and below, the set of admissible thickness mode functions $\{1, x_2, x_2^2, x_2^3\}$ from the third order Mindlin plate equations approximates the conditions of free faces at $x_2 = \pm b$ better than the set of trigonometric functions
$\{1, \cos\frac{\pi(1-\eta)}{2}, \cos\pi(1-\eta), \cos\frac{3\pi(1-\eta)}{2}\}$
from the third order Lee and Nikodem plate equations. (Mindlin's plate theory is however not suitable for higher harmonic overtones of thickness shear vibrations, while Lee and Nikodem's plate theory have been used in such applications[8,11])
2. The first order Mindlin plate equations with the set of admissible thickness mode functions $\{1, x_2\}$ yield accurate frequency spectra for normalized frequency less than 0.1, which is much lower than the fundamental thickness shear frequency at about 0.94. At normalized frequencies greater 0.1, deviations are seen in the frequency spectra starting with the modal branches which are more steeply inclined. The mode shapes of these modal branches have more half waves than those of modal branches which are less inclined.

3. For finite element modeling of high frequency vibrations, the convergence rate of solution for a given mode of vibration is dependent on the number of half waves in the mode of vibration.

Acknowledgment

Supports by the National Science Foundation grant # NSF CMS 94-12644, and Motorola, Inc. are gratefully acknowledged.

References

1. R. D. Mindlin, "High Frequency Vibrations of Quartz Plates by Expansion in Series of Ekstein Functions," *Int. J. Solids Structures*, Vol. 16, 1980, pp. 785-791.
2. R. D. Mindlin, "An Introduction to the Mathematical Theory of Vibrations of Elastic Plates," monograph, U. S. Army Signal Corps Engineering Laboratories, Fort Monmouth, N.J., 1955.
3. P. C. Y. Lee and Z. Nikodem, "An approximate theory for high-frequency vibrations of elastic plates," *Int. J. Solids Struct.*, Vol. 8, 1972, pp. 581-612.
4. Z. Nikodem and P. C. Y. Lee, "Approximate Theory of Vibration of Crystal Plates at High Frequencies," *Int. J. Solids Struct.*, Vol. 10, 1974, pp. 177-196.
5. R. D. Mindlin and M. A. Medick, "Extensional Vibrations of Elastic Plates," *J. Applied Mechanics, Trans. of the ASME*, Vol. 26, 1959, pp. 561-569.
6. R. C. Peach, "A Normal Mode Expansion for Piezoelectric Plates and Certain of Its Applications," *IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. 35, No. 5, Sept. 1988, pp. 593-611.
7. P.C.Y. Lee, S. Syngellakis, and J.P. Hou, "A Two-Dimensional Theory for High-Frequency Vibrations of Piezoelectric Crystal Plates with or without Electrodes," *J. of Appl. Physics*, 61(4), 1987, pp. 1249-1262.
8. Y-K Yong and Z. Zhang, "Numerical Algorithms and Results for SC-Cut Quartz Plates Vibrating at the Third Harmonic Overtone of Thickness Shear," *IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. 41, No. 5, Sept. 1994, pp. 685-693.
9. T.J.R. Hughes, The Finite Element Method, Linear Static and Dynamic Finite Element Analysis, Prentice-Hall, 1987, Chapter 3.
10. Y-K Yong, "On the Use of 1-D Finite Elements for the Temperature Behavior of a Contoured and Partially Plated SC-Cut Resonator," *Proceedings of the 1987 IEEE Ultrasonics Symposium*, 1987, pp. 353-358.
11. R. D. Mindlin, "Third Overtone Quartz Resonator," *Int. J. Solids Struct.*, Vol. 18, No. 9, 1982, pp. 809-817.

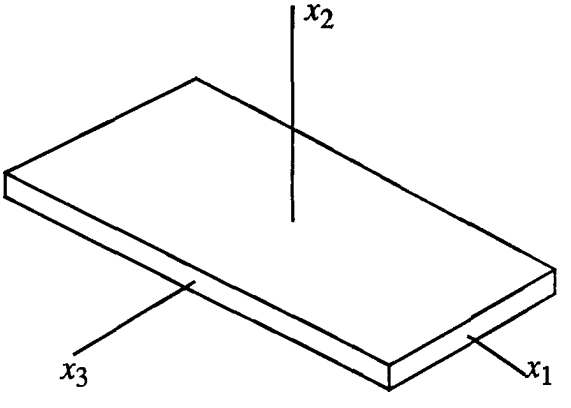


Fig.1 Reference coordinate system of plate.

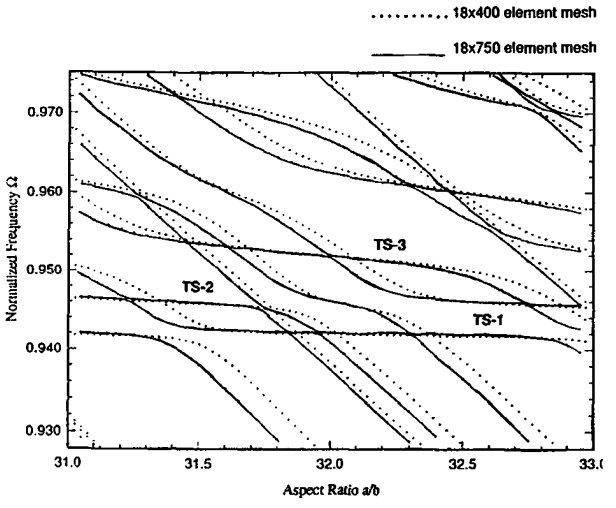


Fig. 2: Numerical convergence of frequency spectra for straight crested waves in an SC-cut quartz strip using the three-dimensional elastic equations of motion.

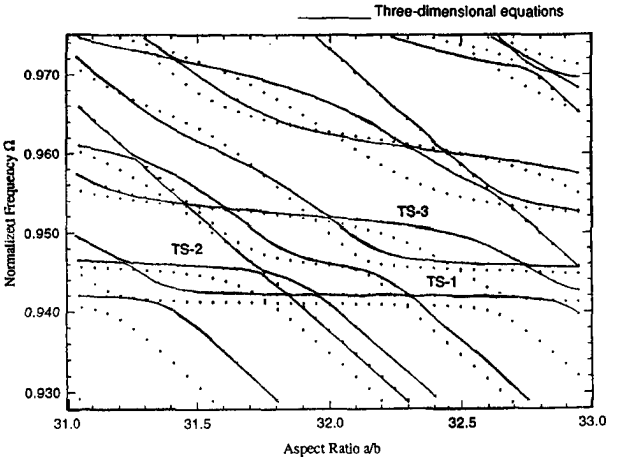


Fig. 3: Comparison of straight crested wave frequency spectra in an SC-cut quartz strip using the first order Mindlin plate equations and the three-dimensional elastic equations of motion.

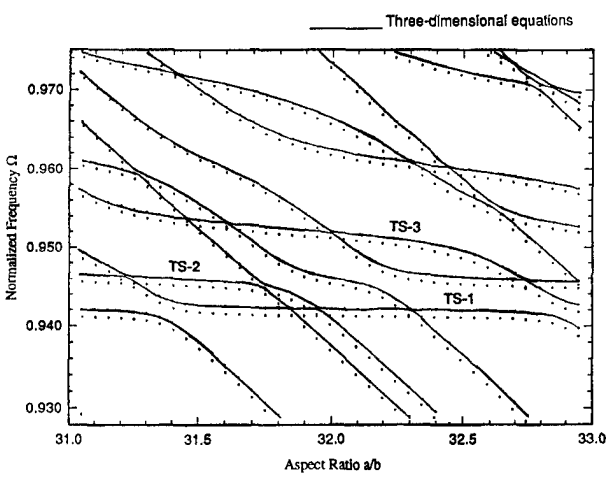


Fig. 4: Comparison of straight crested wave frequency spectra in an SC-cut quartz strip using the third order Mindlin plate equations and the three-dimensional elastic equations of motion.

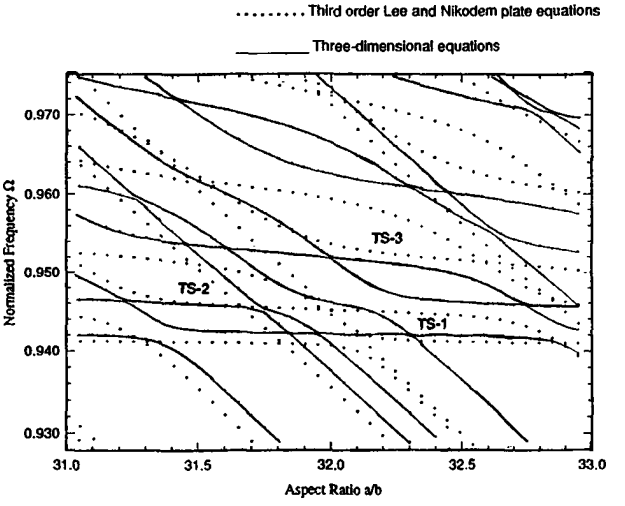


Fig. 5: Comparison of straight crested wave frequency spectra in an SC-cut quartz strip using the third order Lee and Nikodem plate equations and the three-dimensional elastic equations of motion.

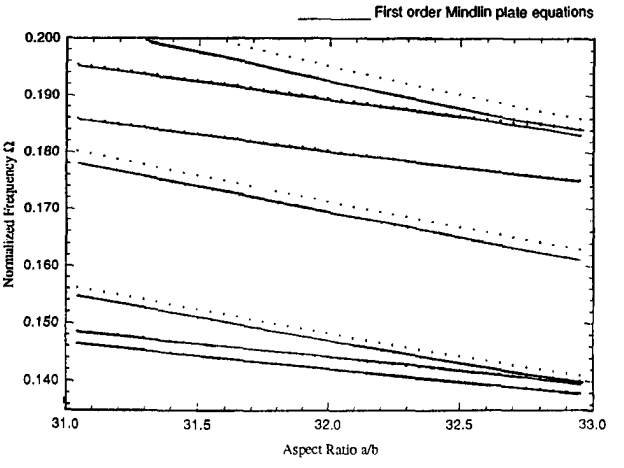


Fig. 6: Comparison of straight crested wave frequency spectra in an SC-cut quartz strip using the first order Mindlin plate equations and the three-dimensional elastic equations of motion.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

TWO-DIMENSIONAL ANALYSIS USING 1D-FEM FOR THE STRAIGHT-CRESTED WAVES IN ARBITRARY ANISOTROPIC CRYSTAL PLATES AND THE AXISYMMETRIC PIEZOELECTRIC VIBRATIONS IN THE CERAMIC DISKS

Akio Ishizaki Hitoshi Sekimoto

Faculty of Engineering, Tokyo Metropolitan University
Hachioji city, Tokyo 192-03, Japan

Abstract: We have developed the two-dimensional analysis using one-dimensional finite element method (1D-FEM) for the straight-crested waves in arbitrary anisotropic crystal plates and the axisymmetric piezoelectric vibrations in the ceramic disks. The solution of two-dimensional equations of motion is described with a linear combination of eigenmodes guided by a pair of parallel edges. 1D-FEM is effectively used to determine the guided eigenmodes and the amplitudes of those. The method attains a high degree of accuracy in spite of small matrix size as compared with two-dimensional (2D) FEM. As examples of the method, the frequency spectra of the straight-crested waves in SC-cut quartz crystal plates and the axisymmetric piezoelectric vibrations in barium titanate ($BaTiO_3$) disks are represented. A convergence study is presented for $BaTiO_3$ disks. The frequency spectra of the straight-crested waves in AT-cut quartz plates are also calculated with the aim of examining the accuracy of the Mindlin's plate equations for those waves.

1 Introduction

The numerical analysis using a computer plays an important role in the design of piezoelectric resonators or transducers. FEM is one of the most powerful techniques for solving piezoelectric problems[1][2]. FEM reduces the analytical boundary-value problem to the numerical eigenvalue problem. However, FEM requires many nodes in order to deal with the vibrational modes having short wavelength or to increase the accuracy of the solution. The analysis of those results in a significant amount of numerical computation because of large matrix size arising from numerous nodes. Hence it is difficult to apply FEM directly to such modes without using a high speed computer with large memory.

On the other hand, the technique which is applicable to two-dimensional vibrational problems in elastic bodies with rectangular free boundaries can greatly reduce the matrix size though it loses the advantage that the resonant frequency is directly obtained from eigenvalue calculation. The method is based on the classical concept[3], in which displacements are represented with a linear combination of eigenmodes guided by a pair of parallel edges in an infinite plate. 1D-FEM is introduced to determine the guided eigenmodes and each amplitude of those. In consequence of employing 1D-FEM, this hybrid approach from numerical and analytic techniques attains a high degree of accuracy in spite of small matrix size.

Our group have already implemented a program to solve Mindlin's plate equations[4] for AT-cut quartz plates[5][6]. In this paper we develop the hybrid approach to analyze the straight-crested waves in arbitrary anisotropic crystal plates and the axisymmetric piezoelectric vibrations in the ceramic disks. As examples of the method, the frequency spectra of the straight-crested waves in SC-cut quartz plates and the axisymmetric piezoelectric vibrations in fully electroded $BaTiO_3$ disks are calculated. A convergence study is also presented for $BaTiO_3$ disks. Numerical results show that our approach has the advantage of giving rapid convergence with small matrix size. The frequency spectra of the straight-crested waves in AT-cut quartz plates are calculated with the aim of examining the accuracy of the Mindlin's plate equations for those waves. The results indicate that the plate equations have to be improved to correct the resonant frequencies of the thickness-flexural mode in two-dimensional analysis.

2 Formulation

2.1 Anisotropic crystal plates

The model: Consider a rectangular anisotropic crystal plate of length $2a$, thickness $2b$ and mass den-

0-7803-2500-1/95 $4.00 © 1995 IEEE

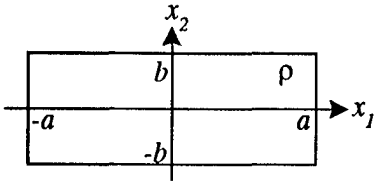


Figure 1: Rectangular plate configuration

sity ρ in the Cartesian coordinates (x_1, x_2) as shown in Fig. 1. Assuming that the field distribution normal to both axes is uniform, the equations of motion and the free boundary conditions are written as

$$T_{j1,1} + T_{j2,2} + \rho\,\omega^2 u_j = 0 \quad (j = 1,2,3) \tag{1}$$

$$T_{j1} = 0 \qquad at\ \ x_1 = \pm a \quad (j = 1,2,3) \tag{2}$$

$$T_{j2} = 0 \qquad at\ \ x_2 = \pm b \quad (j = 1,2,3) \tag{3}$$

where T_{ji} is stress, ω is an angular frequency, u_j is particle displacement and a comma followed by a subscript denotes partial differentiation. Because the configuration of the plate and the boundary conditions are symmetric with respect to the origin, the solutions of (1),(2) and (3) can be separated into two sets which constitute a symmetric and an anti-symmetric displacement field with respect to the origin, respectively. We describe the solutions with a series of N_m eigenmodes guided by the top and bottom surfaces:

$$u_j = \sum_{n=1}^{N_m} p_n [A_j^{(n)}(x_2)\ \cos\beta_n x_1 + S_j^{(n)}(x_2)\ \frac{\sin\beta_n x_1}{\beta_n}] \tag{4}$$

for anti-symmetric modes and

$$u_j = \sum_{n=1}^{N_m} q_n [A_j^{(n)}(x_2)\ \sin\beta_n x_1 - S_j^{(n)}(x_2)\ \frac{\cos\beta_n x_1}{\beta_n}] \tag{5}$$

for symmetric modes, where p_n and q_n are amplitudes, $A_j^{(n)}$ and $S_j^{(n)}$ are eigenfunctions which are odd and even in x_2, respectively, β_n is a wavenumber in x_1 direction, and superscript and subscript n denote that the physical quantity belongs to n-th eigenmode. The solutions (4) or (5) are exact for the straight-crested waves in arbitrary anisotropic crystal plates.

The guided eigenmodes: First step of our method is to determine $A_j^{(n)}$, $S_j^{(n)}$ and β_n by using 1D-FEM. The thickness of the plate is subdivided into line-elements over a range of $0 \le x_2 \le b$. Then we approximate $A_j^{(n)}$ and $S_j^{(n)}$ with their nodal values $A_{jk}^{(n)}$, $S_{jk}^{(n)}$ at $x_2 = x_{2k}$ and the shape functions W_k:

$$A_j^{(n)}(x_2) = \sum_{k=2}^{N_n} W_k(x_2)\ A_{jk}^{(n)}$$

$$S_j^{(n)}(x_2) = \sum_{k=1}^{N_n} W_k(x_2)\ S_{jk}^{(n)} \tag{6}$$

where N_n is a number of nodes. $A_{j1}^{(n)}$ is set to zero and eliminated from the upper equation since $A_j^{(n)}$ is odd in x_2. Substituting (6) into (4), or (5), and according to the procedure of finite element analysis, we obtain the following matrix form equation:

$$\beta_n^2\ \mathbf{B}\ \mathbf{y}_n = \mathbf{C}\ \mathbf{y}_n \tag{7}$$

where $\mathbf{B}$ and $\mathbf{C}$ are $6N_n$–$3(= N_m) \times 6N_n$–3 asymmetric matrices and $\mathbf{y}_n$ is the eigenvector composed of nodal values of the eigenfunctions. (See Appendix A for the derivation of (7).) From (7), we can compute the dispersion relation between a given frequency ,which is included in matrix $\mathbf{C}$, and wavenumbers. Using the eigen-solver for real general matrices, we can easily and successfully determine even the complex valued eigenmodes.

The resonance condition: Second step is to specify the amplitudes of the guided modes which satisfy (1) and (3). This is accomplished by satisfying the remaining boundary condition (2). The solutions (4) or (5) consist of N_m eigenmodes, so those cannot satisfy (2) at all points on the side surfaces. Besides, owing to using 1D-FEM, stresses are discontinuous at the node between neighboring elements. To overcome these difficulties the Galerkin method[7] is employed with W_k used as a weighting function. That is, (2) is multiplied with W_k and integrated with respect to x_2 between the limits $x_2 = 0$ and $x_2 = b$:

$$\sum_{n=1}^{N_m} p_n F_{kn} \cos\beta_n a = 0 \qquad (k = 1,2,\ldots,N_n)$$

$$\sum_{n=1}^{N_m} p_n G_{kn} \frac{\sin\beta_n a}{\beta_n} = 0 \qquad (k = 2,3,\ldots,N_n) \tag{8}$$

for anti-symmetric modes and

$$\sum_{n=1}^{N_m} q_n F_{kn} \sin \beta_n a = 0 \qquad (k = 1, 2, \ldots, N_n)$$

$$\sum_{n=1}^{N_m} q_n G_{kn} \frac{\cos \beta_n a}{\beta_n} = 0 \qquad (k = 2, 3, \ldots, N_n) \quad (9)$$

for symmetric modes where

$$F_{kn} = \int_0^b W_k \left\{ [c_{pp}] \left[S_j^{(n)} \right] + [c_{pq}] \left[A_{j,2}^{(n)} \right] \right\} dx_2$$

$$G_{kn} = \int_0^b W_k \left\{ \beta_n^2 [c_{pp}] \left[A_j^{(n)} \right] - [c_{pq}] \left[S_{j,2}^{(n)} \right] \right\} dx_2 \quad (10)$$

$$[c_{pp}] = \begin{bmatrix} c_{11} & c_{16} & c_{15} \\ c_{16} & c_{66} & c_{56} \\ c_{15} & c_{56} & c_{55} \end{bmatrix} \qquad [c_{pq}] = \begin{bmatrix} c_{16} & c_{12} & c_{14} \\ c_{66} & c_{26} & c_{46} \\ c_{56} & c_{25} & c_{45} \end{bmatrix}$$

$$[c_{qq}] = \begin{bmatrix} c_{66} & c_{26} & c_{46} \\ c_{26} & c_{22} & c_{24} \\ c_{46} & c_{24} & c_{44} \end{bmatrix} \quad (11)$$

$$\left[A_j^{(n)} \right] = \begin{bmatrix} A_1^{(n)} \\ A_2^{(n)} \\ A_3^{(n)} \end{bmatrix} \qquad \left[S_j^{(n)} \right] = \begin{bmatrix} S_1^{(n)} \\ S_2^{(n)} \\ S_3^{(n)} \end{bmatrix} \quad (12)$$

(8) and (9) form a set of linear equations in the amplitudes p_n and q_n, respectively. The resonance condition requires that the determinant of the coefficient matrix is zero. The determinant for a fixed frequency is the function of the length a alone. Consequently the values of the length a corresponding to the frequency ω are determined by searching zeros of the determinant.

The size of the coefficients matrix is $N_m (= 6N_n - 3)$. Our technique can greatly reduce the degree of the matrix though it cannot handle complex geometries.

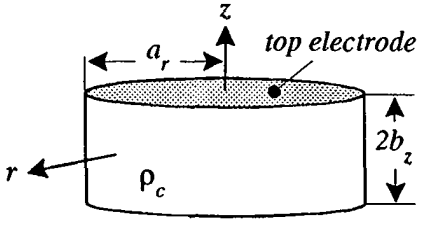


Figure 2: Disk configuration

2.2 The ceramic disks

We also extend the technique to deal with the axisymmetric piezoelectric vibrations in the ceramic disks, since those are essentially two-dimensional and their cross-sectional area is a rectangle. Procedure is principally the same as above. Some differences and additions to the previous analysis are listed bellow:

- Consider axially polarized and fully electroded ceramic disk of diameter $2a_r$, thickness $2b_z$ and mass density ρ_c as illustrated in Fig. 2. Thickness of the electrodes is assumed to be infinitesimal. The analysis is performed in the cylindrical coordinates (r, θ, z). 1D-FEM is applied to determine eigenfunctions of r.
- We use the material constants of *Class 6mm*. The piezoelectric crystals which belong to this class possess an axis of infinite symmetry in respect to elastic and piezoelectric properties.
- The effect of the piezoelectricity is introduced in the system. In addition to the equations of motion, the charge equation of electrostatics is solved under the electrical boundary conditions that the electrical potential is zero on major surfaces and there is no leakage of the electrical displacement on cylindrical surface.

Further details of formulation are given in Appendix B.

3 Numerical results

3.1 SC-cut quartz

The computer program has been developed for SC-cut quartz crystal plates as an example of arbitrary anisotropic crystals. Assuming that the displacements are uniform along Z' crystal axis, the frequency spectrum of anti-symmetric modes was calculated. Using the Bechmann's value[8], the material constants were computed for SC-cut quartz plate which has the rotational symbol (YXwl)22°/33.9° in IRE expression. A three-degree polynomial is selected for W_k. We took $N_m = 93 (N_n = 16)$ so as to attain the 10^{-5} or better accuracy of the resonant frequency and length-to-thickness ratio. The ordinate is the frequency Ω normalized to $\pi/(2b)\sqrt{c_{66}/\rho}$ and the abscissa is the x_1 length-to-x_2 thickness ratio.

There are two horizontal lines labeled TS1-1 and TS3-1 at $\Omega \approx 0.94$ and 1.03. TS1-1 and TS3-1 are fundamental thickness-shear modes which have the

predominant u_1 component and the predominant u_3 component, respectively. Their inharmonics modes are labeled as TS1-k and TS3-k. The spurious modes coupled with TSi-k are classified into the thickness-flexure, the face-shear and the extension which are designated TF-k, FS-k and E-k in Fig. 3, respectively. The number followed by hyphen refers to the number of antinodes across the length $2a$ for TSi-k and TF-k and that of nodes for FS-k and E-k.

The plate theory has been applied in order to get such a figure[9][10]. Our method is helpful to examine the accuracy of the straight-crested wave solutions obtained from the plate equations because it attains a high degree of accuracy and is based on exact equations of motion. In next section, we will discuss it.

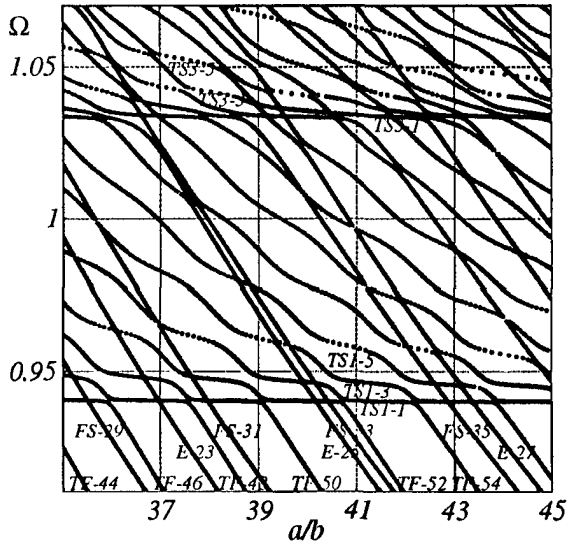


Figure 3: Frequency spectrum of the straight-crested waves in X'-Y' plane of SC-cut quartz plates.

3.2 Barium titanate

The frequency constants, $\omega b/\pi$, are plotted for the first seven modes as a function of the diameter-to-thickness ratio a_r/b_z in Fig. 4. We use the material constants of [2] and take $N_m = 74(N_n = 25)$ for this calculation. Also included in Fig. 4 is the numerical result of 2D-FEM, in which a six-node triangular element is used and the number of degrees of freedom is chosen to be 926, 934 and 942, depending on the diameter-to-thickness ratio. Excellent agreement is confirmed in Fig. 4. This makes it clear that our method is effective to deal with the axisymmetric piezoelectric vibrations in ceramic disks.

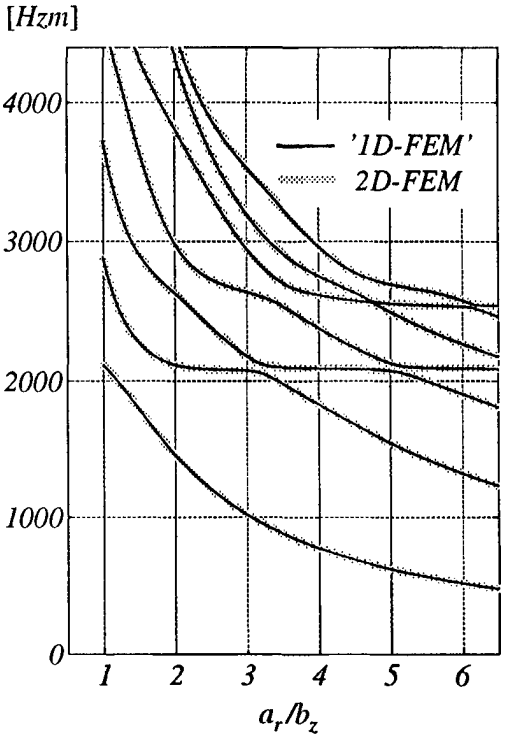


Figure 4: Comparison of resonant frequencies calculated by our technique and those by 2D-FEM, as a function of the diameter-to-thickness ratio in barium titanate disks.

Next let us consider convergence of resonant frequencies against matrix size. We calculate the lowest seven resonant frequencies of BaTiO3 disk of $a_r/b_z = 6$ with varying the number of nodes. The results are shown in Fig. 5 with those of 2D-FEM, and the mode shapes are also shown for convenience. It is found that except the first and second modes, 2D-FEM needs about 10 times as large matrix as our approach. Memory requirements for numerical computation can considerably decrease with employing our technique. The explanation for the advantage of giving rapid convergence is that we do not subdivide the body into finite element directly for two dimensions but for one dimension and then use trigonometric functions as the field distribution along the other dimension.

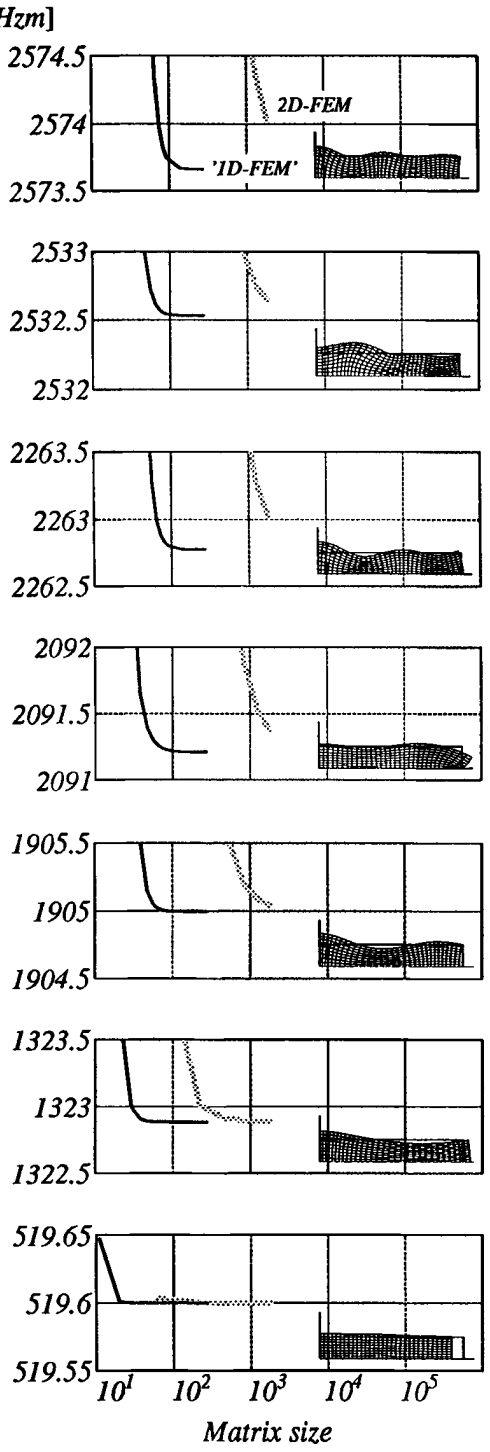


Figure 5: Convergence of the lowest seven frequencies of $BaTiO_3$ disks versus matrix size. Black and gray line show results of our technique and those of 2D-FEM, respectively.

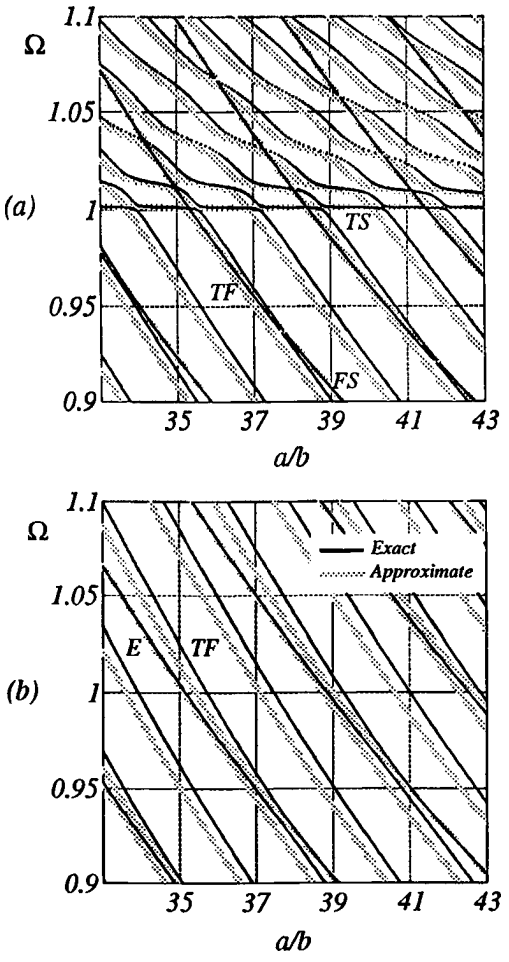


Figure 6: Comparison of resonant frequencies calculated from exact equations of motion and those from Mindlin's plate equations (a) in X-Y' and (b) in Y'-Z' plane of AT-cut quartz plates. The results of the exact and approximated ones are given by black and gray curves, respectively.

4 Discussion

In this section, we examine the accuracy of the straight-crested wave solutions obtained from Mindlin's plate equations[4]. Fig. 6 shows the frequency spectra of the straight-crested waves (a) in X-Y' and (b) in Y'-Z' plane of AT-cut quartz plates. Gray and black curves represent the results of the exact and approximate equations of motion, respectively. In the Fig. 6(a), the modes coupled with the fundamental TS-mode are classified into TF- and FS-modes. In the Fig. 6(b), steeper lines show TF-resonances and the others are E-resonances. Both figures indicate that the plate equations estimate the resonant frequencies of TF-mode lower than exact ones. Relative frequency differences between the two results are about 0.8%.

Much more TF-resonances which belong to the mode with phase reversals along both X and Z' axes are observed in actual AT-cut quartz plates[11]. It is reasonable that the analysis of such a mode is based on the plate equations as three-dimensional FEM is still very costly at present. However, it seems to us that TF-frequencies determined from the plate equations might be seriously different from measured ones if one turns two-dimensional to three-dimensional analysis. Therefore the plate equations have to be improved in order to give the result which completely agree with those of the exact ones at least in two-dimensional analysis.

5 Conclusion

We have developed the two-dimensional analysis using 1D-FEM for the straight-crested waves in arbitrary anisotropic crystal plates and the axisymmetric piezoelectric vibrations in the ceramic disks. Frequency spectra of the straight-crested waves in SC-cut quartz plates and the axisymmetric piezoelectric vibrations in $BaTiO_3$ disks were calculated. The method attains a high degree of accuracy in spite of small matrix size as compared with 2D-FEM. Frequency spectra of the straight-crested waves in AT-cut quartz plates were also calculated. The results show that Mindlin's plate equations of motion evaluate the resonant frequencies of TF-mode lower than exact ones.

References

[1] Y.-K.Yong and Z.Zhang,"Numerical algorithms and results for SC-cut quartz plates vibrating at the third harmonic overtone of thickness-shear," IEEE Trans. Ultrason., Ferroelec., Freq. Contr., vol.41, no.5, pp.685-693, Sep. 1994.

[2] Y.Kagawa and T.Yamabuchi,"Finite element approach for a piezoelectric circular rod," IEEE Trans. Sonics and Ultrason., vol.SU-23, no.6, pp.379-385, Nov. 1976.

[3] H.F.Tiersten, Linear piezoelectric plate vibrations. New York: Plenum 1969, ch.12

[4] R.D.Mindlin and W.J.Spencer,"Anharmonic, thickness-twist overtones of thickness-shear and flexural vibrations of rectangular, AT-cut quartz plates," J. Acoust. Soc. Amer., vol.42, no.6, pp.1268-1277, 1967

[5] H.Sekimoto, Y.Watanabe and M.Nakazawa, "Two-dimensional analysis of thickness-shear and flexural vibrations in rectangular AT-cut quartz plates using a one-dimensional finite element method," in Proc. 44th Annu. Symp. Freq. Contr., 1990, pp.358-362

[6] Y.Watanabe, H.Sekimoto, Y.Kitada and M. Nakazawa,"Two-dimensional charge patterns of spurious modes in rectangular AT-cut quartz plates," in IEEE Ultrason. Symp. Proc., 1993, pp.571-574

[7] R.H.Gallagher, Finite element analysis: fundamentals. Englewood Cliffs, NJ: Prentice-Hall, 1975, ch.5

[8] R.Bechmann,"Elastic and piezoelectric constants of alpha-quartz," Phy. Rev., 110, p.1060, 1958.

[9] Y.-K.Yong, J.T.Stewart, J.Détaint, A.Zarka, B.Capelle and Y.Zheng, "Thickness-shear mode shapes and mass-frequency influence surface of a circular and electroded AT-cut quartz resonator," IEEE Trans. Ultrason., Ferroelec., Freq. Contr., vol.39, no.5, pp.609-617, Sep. 1992.

[10] Y.-K.Yong and Z.Zhang, "A perturbation method for finite element modeling of piezoelectric vibrations in quartz plate resonators," IEEE Trans. Ultrason., Ferroelec., Freq. Contr., vol.40, no.5, pp.551-562, Sep. 1993.

[11] W.J.Spencer,"Observation of resonant vibrations and defect structure in single crystals

by X-ray diffraction topography" in Principles and methods, vol.5, Physical Acoustics, W.P.Mason,Ed. New York: Academic, 1968.pp.143

Appendix A

We derive the characteristic equation which the n-th guided eigenmode satisfies. Substituting (4), or (5), into (1) leads to six equations since each coefficient of $\sin\beta_n x_1$ and $\cos\beta_n x_1$ of three equations of motion must vanish. Now we introduce the Galerkin method to those with the boundary condition (3) taken into account. Since three equations of the six consist of odd functions of x_2 alone and the others are composed of even ones, the integration is carried out between the limits $x_2 = 0$ and $x_2 = b$:

$$\begin{aligned}
&\beta_n^2 \int_0^b \left\{ W_k [c_{pp}][A_j^{(n)}] \right\} dx_2 \\
&= \int_0^b \left\{ W_k [c_{pq}][S_{j,2}^{(n)}] - W_{k,2}[c_{pq}]^T [S_j^{(n)}] \right\} dx_2 \\
&+ \int_0^b \left\{ \rho\omega^2 W_k [A_j^{(n)}] - W_{k,2}[c_{qq}][A_{j,2}^{(n)}] \right\} dx_2 \\
&\qquad\qquad (k = 2, 3, \ldots, N_n) \\
&\beta_n^2 \int_0^b \left\{ W_k [c_{pp}][S_j^{(n)}] + W_k [c_{pq}][A_{j,2}^{(n)}] \right. \\
&\qquad\qquad \left. - W_{k,2}[c_{pq}]^T [A_j^{(n)}] \right\} dx_2 \\
&= \int_0^b \left\{ \rho\omega^2 W_k [S_j^{(n)}] - W_{k,2}[c_{qq}][S_{j,2}^{(n)}] \right\} dx_2 \\
&\qquad\qquad (k = 1, 2, \ldots, N_n) \qquad (13)
\end{aligned}$$

where superscript T denotes transpose of a matrix, and $k \geq 2$ in the upper equation since $A_j^{(n)}(0) = 0$. After substitution of (6) and (11) into (13), it is discretized and arranged in the matrix form (7).

Appendix B

The resonant condition for the ceramic disks with shorted electrodes is formulated here. The equations of motion and the charge equation of electrostatics for a disk are

$$\begin{aligned}
&T_{rr,r} + T_{rz,z} + \frac{T_{rr} - T_{\theta\theta}}{r} + \rho_c \omega^2 u_r = 0 \\
&T_{rz,r} + T_{zz,z} + \frac{T_{rz}}{r} + \rho_c \omega^2 u_z = 0 \\
&D_{r,r} + \frac{D_r}{r} + D_{z,z} = 0 \qquad (14)
\end{aligned}$$

where D_i is the electric displacement. The mechanical and the electric boundary conditions are

$$T_{rr} = T_{rz} = D_r = 0 \qquad at\ r = a_r \qquad (15)$$

$$T_{zz} = T_{rz} = 0 \quad \phi = 0 \qquad at\ z = \pm b_z \qquad (16)$$

where ϕ is electric potential. These means that the disk is mechanically free, the electrodes are shorted and there is no flux leakage on the cylindrical surface. Stress and the electric displacement are related with the displacement and the electric potential through the following equations:

$$\begin{aligned}
T_{rr} &= c_{11}^E u_{r,r} + \frac{c_{12}^E u_r}{r} + c_{13}^E u_{z,z} + e_{31}\phi_{,z} \\
T_{\theta\theta} &= c_{12}^E u_{r,r} + \frac{c_{11}^E u_r}{r} + c_{13}^E u_{z,z} + e_{31}\phi_{,z} \\
T_{zz} &= c_{13}^E u_{r,r} + \frac{c_{13}^E u_r}{r} + c_{33}^E u_{z,z} + e_{33}\phi_{,z} \\
T_{rz} &= c_{44}^E (u_{r,z} + u_{z,r}) + e_{15}\phi_{,r} \\
D_r &= e_{15}(u_{r,z} + u_{z,r}) - \varepsilon_{11}^S \phi_{,r} \\
D_z &= e_{31} u_{r,r} + \frac{e_{31} u_r}{r} + e_{33} u_{z,z} - \varepsilon_{33}^S \phi_{,z} \qquad (17)
\end{aligned}$$

where e_{ij}, ε_{ij}^S and c_{ij}^E are, respectively, the piezoelectric stress constant, the dielectric permittivity when the strain field is zero and the elastic stiffness constant when the electric field is zero. The displacements and the electrical potential can be described as

$$\begin{aligned}
u_r &= \sum_{n=1}^{N_m} p_n A_r^{(n)}(r) \cos\beta_n z \\
u_z &= \sum_{n=1}^{N_m} p_n S_z^{(n)}(r) \frac{\sin\beta_n z}{\beta_n} \\
\phi &= \sum_{n=1}^{N_m} p_n S_\phi^{(n)}(r) \frac{\sin\beta_n z}{\beta_n} \qquad (18)
\end{aligned}$$

Half thickness of the disk is subdivided into line-elements. Then we approximate the eigen functions with their nodal values and the shape functions :

$$\begin{aligned}
A_r^{(n)}(r) &= \sum_{k=2}^{N_n} W_k(r) A_{rk}^{(n)} \\
S_z^{(n)}(r) &= \sum_{k=1}^{N_n} W_k(r) S_{zk}^{(n)}
\end{aligned}$$

$$S_\phi^{(n)}(r) = \sum_{k=1}^{N_n} W_k(r) S_{\phi k}^{(n)} \tag{19}$$

After substitution of (17) and (18) into (14), it is multiplied by rW_k and integrated with respect to r between the limits $r = 0$ and $r = a_r$. Taking the boundary condition (15) into consideration, we have

$$\begin{aligned}
&\beta_n^2 \int_0^{a_r} \left\{ rW_k c_{44}^E A_r^{(n)} \right\} dr \\
&= \int_0^{a_r} \left\{ rW_k [c_{44}^E, e_{15}] \begin{bmatrix} S_{z,r}^{(n)} \\ S_{\phi,r}^{(n)} \end{bmatrix} \right. \\
&\qquad \left. -(rW_k)_{,r} [c_{13}^E, e_{31}] \begin{bmatrix} S_z^{(n)} \\ S_\phi^{(n)} \end{bmatrix} \right\} dr \\
&+ \int_0^{a_r} \left\{ \rho\omega^2 rW_k A_r^{(n)} \right. \\
&\qquad \left. -\frac{1}{r} [rW_{k,r}, W_k] \begin{bmatrix} c_{11}^E & c_{12}^E \\ c_{12}^E & c_{11}^E \end{bmatrix} \begin{bmatrix} rA_{r,r}^{(n)} \\ A_r^{(n)} \end{bmatrix} \right\} dr \\
&\qquad\qquad (k = 2, 3, \ldots, N_n)
\end{aligned}$$

$$\begin{aligned}
&\beta_n^2 \int_0^{a_r} \left\{ W_k \begin{bmatrix} c_{33}^E & e_{33} \\ e_{33} & -\varepsilon_{33}^S \end{bmatrix} \begin{bmatrix} rS_z^{(n)} \\ rS_\phi^{(n)} \end{bmatrix} \right. \\
&\qquad + W_k \begin{bmatrix} c_{13}^E \\ e_{31} \end{bmatrix} (rA_r^{(n)})_{,r} \\
&\qquad \left. - W_{k,r} \begin{bmatrix} c_{44}^E \\ e_{15} \end{bmatrix} rA_r^{(n)} \right\} dr \\
&= \int_0^{a_r} \left\{ \rho\omega^2 W_k \begin{bmatrix} rS_z^{(n)} \\ 0 \end{bmatrix} \right. \\
&\qquad \left. - W_{k,r} \begin{bmatrix} c_{44}^E & e_{15} \\ e_{15} & -\varepsilon_{11}^S \end{bmatrix} \begin{bmatrix} rS_{z,r}^{(n)} \\ rS_{\phi,r}^{(n)} \end{bmatrix} \right\} dr \\
&\qquad\qquad (k = 1, 2, \ldots, N_n)
\end{aligned} \tag{20}$$

Substituting (19) into (20), it is arranged to a matrix form.

The resonance condition is derived from the remaining boundary condition (16):

$$\sum_{n=1}^{N_m} p_n F_{kn} \cos \beta_n b_z = 0 \qquad (k = 1, 2, \ldots, N_n)$$

$$\sum_{n=1}^{N_m} p_n G_{kn} \frac{\sin \beta_n b_z}{\beta_n} = 0 \qquad (k = 2, 3, \ldots, N_n)$$

$$\sum_{n=1}^{N_m} p_n S_\phi^{(n)}(r_k) \frac{\sin \beta_n b_z}{\beta_n} = 0 \quad (k = 1, 2, \ldots, N_n) \tag{21}$$

where

$$\begin{aligned}
F_{kn} &= \int_0^{a_r} W_k \\
&\quad \times \left\{ c_{13}^E (A_{r,r}^{(n)} + \frac{A_r^{(n)}}{r}) + c_{33}^E S_z^{(n)} + e_{33} S_\phi^{(n)} \right\} r\, dr
\end{aligned}$$

$$\begin{aligned}
G_{kn} &= \int_0^{a_r} W_k \\
&\quad \times \left\{ -c_{44}^E \beta_n^2 A_r^{(n)} + c_{44}^E S_{z,r}^{(n)} + e_{15} S_{\phi,r}^{(n)} \right\} r\, dr
\end{aligned} \tag{22}$$

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Electromagnetic Waves in an Infinite Isotropic Dielectric Plate Enclosed by Two Parallel Perfect Conductors

P.C.Y. Lee and J.D. Yu

Department of Civil Engineering & Operations Research
Princeton University, Princeton, N.J. 08544

Abstract

Electromagnetic (EM) plane-wave propagation is analyzed for an infinite isotropic dielectric plate and its surrounding free space enclosed by two parallel perfect conductors which are of equal distance from the faces of the dielectric plate.

Exact solutions are obtained satisfying the three-dimensional Maxwell's equations, the continuity conditions at the interfaces of the dielectric plate and free space, and the electric boundary conditions at the faces of the perfect conductors. Dispersion relations are obtained for both the transverse electric (TE) and transverse magnetic (TM) waves. Dispersion curves are computed and analyzed for frequency branches with both real and imaginary wave numbers.

The partition of energy, i.e. the ratio of EM energy per unit area in the dielectric plate to that in the surrounding free space are calculated as a function of frequency for both TE and TM waves. The effect of the gap between the shielding conductor and the face of the dielectric plate and the effect of the relative dielectric permittivity on the energy partition are examined systematically.

1 Introduction

Exact solutions of straight-crested waves propagating in an infinite and isotropic dielectric plate surrounded by free space can be found in many books, for instance, in Collin and Marcus [1][2]. Dispersion relations and modes for dielectric plate with perfect electric faces, perfect magnetic faces, and faces in contact with free space of infinite extent were studied by Lee and Yang [3]. Exact solutions of straight-crested waves propagating in an infinite and most general anisotropic dielectric plate surrounded by free space were obtained by Lee and Yu [4].

In the present paper, we consider the electromagnetic (EM) waves in an infinite isotropic dielectric plate enclosed by two parallel perfect conductors. Exact solutions are obtained satisfying the three-dimensional Maxwell's equations, the continuity conditions at the interfaces between the plate and free space, and the perfect electric boundary conditions at the faces of the perfect conductors.

The dispersion relations and the field variations along the thickness direction are obtained and discussed for both the transverse electric (TE) and transverse magnetic (TM) waves. Dispersion curves are computed and analyzed for frequency branches with both real and imaginary wave numbers. Two families of frequency bounds, which are governed by two sets of simple algebraic equations, are obtained by a method similar to that employed by Mindlin [5]. The effect of $\bar{b}/b$, the distance from the conductor to the face of the plate divided by the one half of the plate thickness, and the relative dielectric permittivity ϵ/ϵ_0 on the behavior of dispersion curves can be seen clearly in terms of the frequency bounds.

The partition of energy is examined by calculating the ratio of EM energy per unit area in the plate to that in the surrounding free space as a function of frequency for both TE and TM waves and for different values of $\bar{b}/b$ and ϵ/ϵ_0. It is found that the energy ratio increases as the frequency increases. The increase of $\bar{b}/b$ from 0.1 to 1.0 increases R_e ratio for the TM modes, but decreases the R_e ratio for the TE modes. But the energy ratio R_e remains almost constant for $\bar{b}/b$ varying from 1.0 to 20.0. The increase of the ϵ/ϵ_0 ratio increases the energy ratio R_e for both TE and TM waves.

2 Governing Equations and Boundary Conditions

We consider an isotropic dielectric plate of infinite extent referred to a rectangular coordinate system x_i

0-7803-2500-1/95 $4.00 © 1995 IEEE

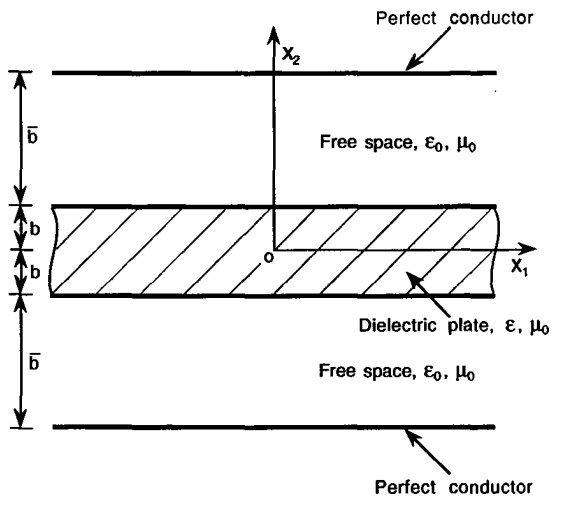


Figure 1: An infinite dielectric plate enclosed by two parallel perfect conductors.

with the faces at $x_2 = \pm b$ and with x_1 and x_3 being the coordinates of the middle plane of the plate. The plate is surrounded by free space which is enclosed by two parallel perfect conductors of equal distance from the faces of the plate as shown in Fig. 1.

The governing equations of the EM field in the plate comprise Maxwell's equations

$$\epsilon_{ijk}E_{k,j} = -B_{i,t} \ , \quad B_{i,i} = 0 \ , \tag{1}$$

$$\epsilon_{ijk}H_{k,j} = D_{i,t} \ , \quad D_{i,i} = 0 \ , \tag{2}$$

and the linear constitutive equations

$$D_i = \epsilon E_i \ , \quad B_i = \mu_0 H_i \ . \tag{3}$$

In these equations, E_i and H_i are the electric and magnetic field intensities, D_i and B_i the electric and magnetic flux densities, ϵ the dielectric permittivity, μ_0 the magnetic permeability of free space, and ϵ_{ijk} the unit alternating tensor. It is assumed that there are no free charges and no current and that the dielectric is nonmagnetic.

Since the region of the plate is source free and the field is time-varying, an electromagnetic vector potential A_i may be introduced from which E_i and B_i can be obtained by [6]

$$E_i = -A_{i,t} \ , \quad B_i = \epsilon_{ijk}A_{k,j} \ . \tag{4}$$

Substitution of (4) into (1) shows that (1) are identically satisfied. Therefore, in terms of A_i, (2), (3), and (4) become the governing equations of the EM field in the dielectric plate.

We designate $\bar{E}_i, \bar{D}_i, \bar{H}_i$ and $\bar{B}_i$ as the field vectors in the free space. The governing equations of EM field in free space are the same as (2)–(4) except that ϵ is replaced by ϵ_0, the permittivity of free space.

The continuity conditions to be satisfied at $x_2 = \pm b$, the interfaces between the plate and surrounding free space, are

$$\begin{aligned} E_a(x_2 = \pm b) &= \bar{E}_a(x_2 = \pm b) \ , \\ B_2(x_2 = \pm b) &= \bar{B}_2(x_2 = \pm b) \ , \end{aligned} \tag{5}$$

$$\begin{aligned} H_a(x_2 = \pm b) &= \bar{H}_a(x_2 = \pm b) \ , \\ D_2(x_2 = \pm b) &= \bar{D}_2(x_2 = \pm b) \ , \end{aligned} \tag{6}$$

where $a = 1, 3$.

The boundary conditions for the perfect conductors at $x_2 = \pm(b + \bar{b})$ are

$$\begin{aligned} E_a\left(x_2 = \pm(b + \bar{b})\right) &= 0 \ , \\ B_2\left(x_2 = \pm(b + \bar{b})\right) &= 0 \ , \end{aligned} \tag{7}$$

where $a = 1, 3$.

By substituting (4) into (5), it may be seen that conditions (5) and (7) can be replaced, respectively, by the following conditions on the tangential components of A_i

$$A_a(x_2 = \pm b) = \bar{A}_a(x_2 = \pm b) \ , \quad a = 1, 3 \ , \tag{8}$$

$$A_a\left[x_2 = \pm(b + \bar{b})\right] = 0 \ , \quad a = 1, 3 \ . \tag{9}$$

3 Guided TE Waves

Substitution of (4) into (3) and, in turn, in (2) results in wave equations on the vector potential A_i in the dielectric plate

$$A_{i,jj} = \frac{1}{c^2}A_{i,tt} \ , \qquad A_{i,i} = 0 \ , \tag{10}$$

where $c = 1/(\epsilon\mu_0)^{\frac{1}{2}}$ is the speed of light in the dielectric plate. The equations for $\bar{A}_i$, the vector potential in the free space, are the same as (10) except that c is replaced by $c_0 = 1/(\epsilon_0\mu_0)^{\frac{1}{2}}$, the speed of light in free space.

For straight-crested TE waves propagating in the x_1 direction in the plate ($|x_2| < b$), we take

$$\begin{aligned} &A_1 = 0 \ , \qquad A_2 = 0 \ , \\ &A_3 = (C_1 \cos\eta x_2 + C_2 \sin\eta x_2)e^{i(\xi x_1 - \omega t)} \ , \end{aligned} \tag{11}$$

where ξ and η are the wave numbers in the x_1 and x_2 directions, respectively, and ω is the angular or radian frequency. The phase velocity v is related to ω and ξ by

$$v = \omega/\xi \ . \tag{12}$$

We note that in (11) the terms associated with C_1 and C_2 are the two independent solutions of (10) and they represent, respectively, the variation of E_3 symmetric and antisymmetric to the x_1-x_3 plane. Equations (10) are satisfied by (11), provided

$$\eta^2 = \frac{\omega^2}{c^2} - \xi^2 = \xi^2 \left(\frac{v^2}{c^2} - 1 \right) , \qquad (13)$$

and insertion of (11) in (4) gives the following non-zero field components:

$$\begin{aligned} E_3 &= i\omega \left(C_1 \cos \eta x_2 + C_2 \sin \eta x_2 \right) e^{i(\xi x_1 - \omega t)}, \\ B_1 &= \eta \left(-C_1 \sin \eta x_2 + C_2 \cos \eta x_2 \right) e^{i(\xi x_1 - \omega t)}, \\ B_2 &= -i\xi \left(C_1 \cos \eta x_2 + C_2 \sin \eta x_2 \right) e^{i(\xi x_1 - \omega t)}. \end{aligned} \qquad (14)$$

It may be seen from (14) that (11) is, indeed, the vector potential for the electric waves transverse to the x_1 and x_2 axes, since $E_1 = E_2 = 0$.

For the surrounding free space, $b < |x_2| < (b + \bar{b})$, we take the wave form which is propagating in the x_1 direction by letting

$$\begin{aligned} &\bar{A}_1 = 0 , \quad \bar{A}_2 = 0 , \\ &\bar{A}_3 = \left(\bar{C}_1 \pm \bar{C}_2 \right) \frac{\sinh \bar{\eta}(b + \bar{b} \mp x_2)}{\sinh \bar{\eta}(b + \bar{b})} e^{i(\xi x_1 - \omega t)}, \end{aligned} \qquad (15)$$

where the upper and lower signs apply to the regions of the upper half $(b \le x_2 \le b + \bar{b})$ and the lower half $(-b - \bar{b} \le x_2 \le -b)$ of the surrounding free space, respectively. We note that in (15) the terms associated with $\bar{C}_1$ and $\bar{C}_2$ are solutions symmetric and antisymmetric with respect to the middle plane of the plate, respectively.

The vector potential (15) satisfies the wave equation (10) for free space provided that

$$\bar{\eta}^2 = \bar{\xi}^2 - \frac{\omega^2}{c_0^2} = \bar{\xi}^2 \left(1 - \frac{\bar{v}^2}{c_0^2} \right) , \qquad (16)$$

where $\bar{v} = \omega / \bar{\xi}$. Substitution of (15) into (4) gives the non-zero field components

$$\begin{aligned} \bar{E}_3 &= i\omega (\bar{C}_1 \pm \bar{C}_2) \frac{\sinh \bar{\eta}(b + \bar{b} \mp x_2)}{\sinh \bar{\eta}(b + \bar{b})} e^{i(\bar{\xi} x_1 - \omega t)}, \\ \bar{B}_1 &= -\bar{\eta} (\pm \bar{C}_1 + \bar{C}_2) \frac{\cosh \bar{\eta}(b + \bar{b} \mp x_2)}{\sinh \bar{\eta}(b + \bar{b})} e^{i(\bar{\xi} x_1 - \omega t)}, \\ \bar{B}_2 &= -\frac{1}{\bar{v}} \bar{E}_3 . \end{aligned} \qquad (17)$$

It may be seen that the perfect electric boundary conditions (9) are satisfied by (17) and the continuity conditions (6) and (8) reduce to

$$A_3 = \bar{A}_3 , \quad H_1 = \bar{H}_1 , \quad \text{at} \quad x_2 = \pm b . \qquad (18)$$

Insertion of $(11)_3$, $(15)_3$, $(14)_2$, and $(17)_2$ in (18) leads to

$$\begin{aligned} C_1 \cos \eta b &= \bar{C}_1 \frac{\sinh \bar{\eta} \bar{b}}{\sinh \bar{\eta}(b + \bar{b})} , \\ C_1 \eta \sin \eta b &= \bar{C}_1 \frac{\bar{\eta} \cosh \bar{\eta} \bar{b}}{\sinh \bar{\eta}(b + \bar{b})} , \end{aligned} \qquad (19)$$

for the symmetric modes ($C_2 = \bar{C}_2 = 0$), and

$$\begin{aligned} C_2 \sin \eta b &= \bar{C}_2 \frac{\sinh \bar{\eta} \bar{b}}{\sinh \bar{\eta}(b + \bar{b})} , \\ C_2 \eta \cos \eta b &= -\bar{C}_2 \frac{\bar{\eta} \cosh \bar{\eta} \bar{b}}{\sinh \bar{\eta}(b + \bar{b})} , \end{aligned} \qquad (20)$$

for the antisymmetric modes ($C_1 = \bar{C}_1 = 0$). Also, we have

$$\xi = \bar{\xi} , \qquad v = \bar{v} . \qquad (21)$$

The frequency equations resulted from (19) and (20) may be combined into one expression

$$\tan \eta b = \pm \left(\frac{\bar{\eta}/\eta}{\tanh \bar{\eta} \bar{b}} \right)^{\pm 1} , \qquad (22)$$

where the upper and lower signs apply to the symmetric and antisymmetric modes, respectively.

By the use of (12), (13), (16), and (21), the above transcendental equation (22) can be expressed in terms of v and ξ

$$\tan \left(\frac{v^2}{c^2} - 1 \right)^{\frac{1}{2}} \xi b = \pm \left[\frac{(1 - v^2/c_0^2)^{\frac{1}{2}} / (v^2/c^2 - 1)^{\frac{1}{2}}}{\tanh (1 - v^2/c_0^2)^{\frac{1}{2}} \xi \bar{b}} \right]^{\pm 1} . \qquad (23)$$

Certain analytical properties of the dispersion relations (22) can be obtained by the following considerations:

(1) When $\bar{b}$, the distance between the plate and the perfect conductors, approaches zero, the continuity conditions (18) are replaced by the perfect elastic face conditions, i.e.

$$A_3(x_2 = \pm b) = 0 . \qquad (24)$$

Therefore (19) and (20) become, respectively,

$$\cos \eta b = 0 \quad \text{or} \quad \eta = \frac{n\pi}{2b}, \; n = 1, 3, 5, \cdots \qquad (25)$$

for the symmetric modes, and

$$\sin \eta b = 0 \quad \text{or} \quad \eta = \frac{n\pi}{2b}, \; n = 2, 4, 6, \cdots \qquad (26)$$

for the antisymmetric modes. These results are the same as those given by eq.(15) in [3].

(2) When v approaches c_0, $\bar{\eta} \to 0$, both the denominator and numerator at the right hand side of (22) vanish. By using

$$\lim_{\bar{\eta} \to 0} \frac{\bar{\eta}}{\tanh \bar{\eta} \bar{b}} = \frac{1}{\bar{b}} , \qquad (27)$$

we obtain

$$\tan \eta b = \pm \left(\frac{b/\bar{b}}{\eta b} \right)^{\pm 1} , \qquad (28)$$

and the amplitude ratios, from (19) and (20), are

$$\frac{\bar{C}_1}{C_1} = \frac{b+\bar{b}}{b} \cos \eta b , \qquad \frac{\bar{C}_2}{C_2} = \frac{b+\bar{b}}{b} \sin \eta b . \qquad (29)$$

(3) When v approaches c and ξ approaches infinity, (23) are satisfied by

$$\eta = \lim_{v \to c,\ \xi \to \infty} \left(\frac{v^2}{c^2} - 1 \right)^{\frac{1}{2}} \xi = \frac{(p+1)\pi}{2b} , \qquad (30)$$
$$p = 0, 1, 2, \cdots ,$$

or

$$\frac{\omega^2}{c^2} - \xi^2 = \left[\frac{(p+1)\pi}{2b} \right]^2 , \qquad (31)$$

where $p = 0, 2, 4, \cdots$ for symmetric modes, and $p = 1, 3, 5, \cdots$ for antisymmetric modes. From $(19)_1$ and $(20)_1$

$$\left| \frac{\bar{C}_a}{C_a} \right| = 0 , \quad a = 1, 2 \quad \text{(no sum)} , \qquad (32)$$

and from $(19)_2$ and $(20)_2$

$$\lim_{v \to c,\ \xi \to \infty} \left| \frac{\bar{C}_a}{C_a} \frac{\bar{\eta} \cosh \bar{\eta} \bar{b}}{\sinh \bar{\eta}(b+\bar{b})} \right| = \frac{(p+1)\pi}{2b} . \qquad (33)$$

Therefore, we see, in $(17)_2$, $\bar{B}_1 = 0$ for $|x_2| > b$, but it has non-zero values on the faces of the plate.

(4) When $v > c_0$, $\bar{\eta}$ is imaginary. From (16), we have

$$\bar{\eta} = i\hat{\eta} = i \left(\frac{\omega^2}{c_0^2} - \xi^2 \right)^{\frac{1}{2}} , \quad \hat{\eta} = \left(\frac{\omega^2}{c_0^2} - \xi^2 \right)^{\frac{1}{2}} . \qquad (34)$$

The field components for the free space can be written as

$$\begin{aligned} \bar{E}_3 &= i\omega(\bar{C}_1 \pm \bar{C}_2) \frac{\sin \hat{\eta}(b + \bar{b} \mp x_2)}{\sin \hat{\eta}(b+\bar{b})} e^{i(\bar{\xi} x_1 - \omega t)} , \\ \bar{B}_1 &= -\hat{\eta}(\pm \bar{C}_1 + \bar{C}_2) \frac{\cos \hat{\eta}(b + \bar{b} \mp x_2)}{\sin \hat{\eta}(b+\bar{b})} e^{i(\bar{\xi} x_1 - \omega t)} , \qquad (35) \\ \bar{B}_2 &= -\frac{1}{\bar{v}} \bar{E}_3 . \end{aligned}$$

And the frequency equations (22) become

$$\tan \eta b = \pm \left(\frac{\hat{\eta}/\eta}{\tan \hat{\eta} \bar{b}} \right)^{\pm 1} . \qquad (36)$$

(5) As $\xi \to 0$, the frequency equations (22) reduce to

$$\tan \frac{\omega b}{c} = \pm \left(\frac{c/c_0}{\tan \frac{\omega \bar{b}}{c_0}} \right)^{\pm 1} . \qquad (37)$$

The roots of (37) or ω are called the cut-off frequencies, since $\xi = 0$.

(6) When the wave number ξ is imaginary, it is seen from (13) and (16) that η is real and $\bar{\eta}$ is imaginary. We let

$$\xi = i\xi' , \qquad \bar{\eta} = i\hat{\eta} , \qquad \xi' , \ \hat{\eta} \text{ real} \qquad (38)$$

and then

$$\eta = \left(\frac{\omega^2}{c^2} + \xi'^2 \right)^{\frac{1}{2}} , \qquad \hat{\eta} = \left(\frac{\omega^2}{c_0^2} + \xi'^2 \right)^{\frac{1}{2}} . \qquad (39)$$

The expressions for the field components in the plate and in the surrounding free space remain the same as those given by (14) and (35), respectively, except that ξ $(=\bar{\xi})$ is replaced by $i\xi'$.

(7) Some of the simple roots of the frequency equation (36) or (22) can be obtained in a manner similar to that employed by Mindlin in studying the dispersion relations of plane waves in an elastic plate with traction-free faces [5].

We let $\bar{\eta} = i\hat{\eta}$ in (19) and (20). It is seen that (19), for symmetric B_3 modes, are satisfied by

$$\sin \eta b = 0, \quad \cos \hat{\eta} \bar{b} = 0, \quad \left| \frac{\bar{C}_1}{C_1} \right| = \sin \hat{\eta}(b + \bar{b})$$

and

$$\cos \eta b = 0, \quad \sin \hat{\eta} \bar{b} = 0, \quad \left| \frac{\bar{C}_1}{C_1} \right| = \frac{\eta}{\hat{\eta}} \sin \hat{\eta}(b + \bar{b}). \qquad (40)$$

For the antisymmetric modes, (20) are similarly satisfied by

$$\sin \eta b = 0, \quad \sin \hat{\eta} \bar{b} = 0, \quad \left| \frac{\bar{C}_2}{C_2} \right| = \frac{\eta}{\bar{\eta}} \sin \hat{\eta}(b + \bar{b})$$

and

$$\cos \eta b = 0, \quad \cos \hat{\eta} \bar{b} = 0, \quad \left| \frac{\bar{C}_2}{C_2} \right| = \sin \hat{\eta}(b + \bar{b}). \qquad (41)$$

We note that the simple transcendental equations in (40) and (41) are satisfied by

$$\eta = \frac{n\pi}{2b} , \qquad \hat{\eta} = \frac{m\pi}{2\bar{b}} , \qquad (42)$$

where n and m are non-zero integers and they take on values as follows.

For symmetric modes: n=odd and m=even, or n=even and m=odd.

For antisymmetric modes: n=even and m=even, or n=odd and m=odd.

By substituting (42) into (12) and $(34)_2$ and introducing the definitions for the normalized wave number, frequency, and phase velocity

$$X = \frac{\xi}{(\pi/2b)}, \quad \Omega = \frac{\omega}{c_0(\pi/2b)}, \quad V = \frac{v}{c_0} = \frac{\Omega}{X}, \qquad (43)$$

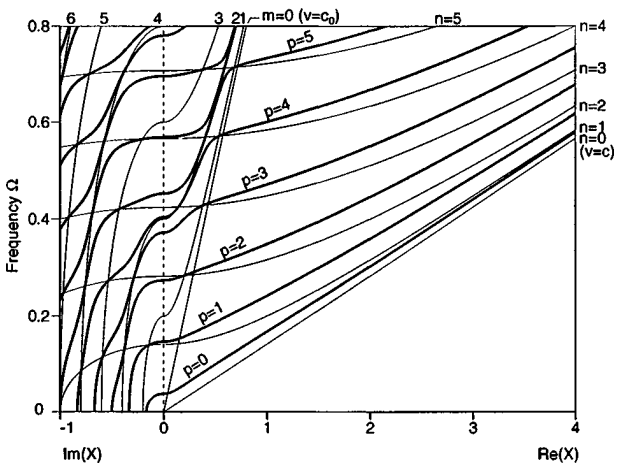


Figure 2: Dispersion curves (p even for symmetric modes; p odd for antisymmetirc modes) and their bounds (thin lines denoted by m and n) for guided TE waves in a dielectric plate enclosed by two parallel perfect conductors ($\epsilon/\epsilon_0 = 50$ and $\bar{b}/b = 5$).

we have

$$\hat{n}^2\Omega^2 - X^2 = n^2\,, \qquad \Omega^2 - X^2 = \left(\frac{b}{\bar{b}}\right)^2 m^2\,, \quad (44)$$

where $\hat{n} = \sqrt{\epsilon/\epsilon_0} = c_0/c$ is the refractive index. The intersections of $(44)_1$ and $(44)_2$ with n odd and m even or with n even and m odd are the simple solutions for the symmetric E_3 modes. And the intersections of the curves for n even and m even or for n odd and m odd are solutions for the antisymmetric E_3 modes.

Dispersion curves are obtained by solving the transcendental equations (22), (36), and (37) numerically for $\epsilon/\epsilon_0 = 50$, $\bar{b}/b = 5$ and they are plotted as thick solid lines denoted by $p = 0, 2, 4, \cdots$ for the symmetric E_3 modes and $p = 1, 3, 5, \cdots$ for the antisymmetric E_3 modes in Fig. 2. The systems of thin lines designated by n=constant and by m=constant are the bounds of the frequency branches. It may be seen in Fig. 2 that the frequency branches, indeed, pass through these bounds only at the intersection points. The points at which the frequency branches cross the line $m = 0$, which is not a bound, can be calculated from (28).

We see from $(44)_2$ that $b/\bar{b}$ ratio affects only the behavior of the m=constant family of curves. In the case $\bar{b} \to 0$ or $b/\bar{b} \to \infty$, i.e. when the perfect conductor is touching the face of the dielectric plate, the intersection points will occur at the region where either Ω or ImX are very large. Hence, as $\bar{b}$ approaches zero the pth frequency branch approaches the $n = p + 1$ bound as shown in Fig. 2 of [3]. In the case $\bar{b} \to \infty$ or $b/\bar{b} \to 0$, i.e. when the dielectric plate is surrounded by free space of infinite extent, the m=constant curves reduce to the $m = 0$ line. There are no intersection of bounds or no frequency branches in the region $v > c_0$. Hence, as $\bar{b} \to \infty$, frequency branches exist in the region $v < c_0$ and they are bounded by the n=constant family of curves as given in Fig. 5 of [3].

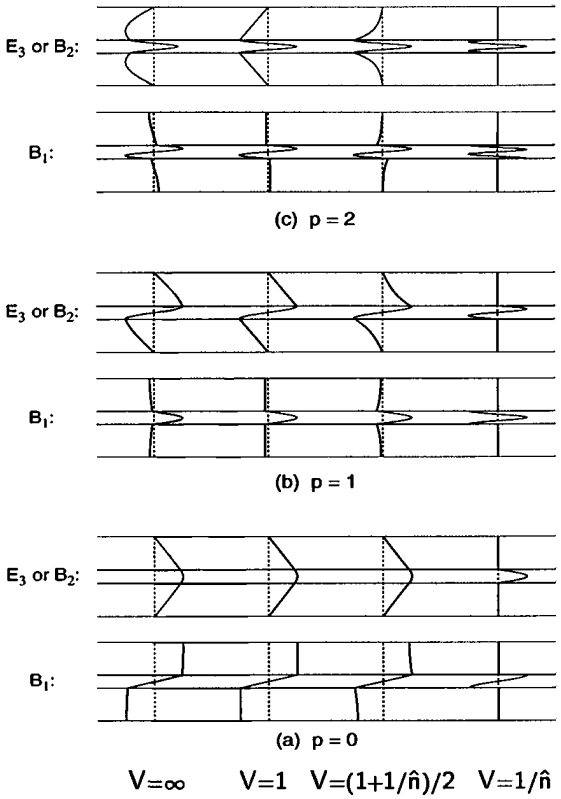


Figure 3: Modes of guided TE waves in a dielectric plate enclosed by parallel perfect conductors on the first three frequency branches ($p = 0, 1, 2$) at $V = \infty$, 1, $(1 + 1/\hat{n})/2$, and $1/\hat{n}$ for $\hat{n}^2 = 50$ and $\bar{b}/b = 5$.

Similarly we see from $(44)_1$ that ϵ/ϵ_0 ratio only affects the behavior of n=constant family of frequency bounds which are hyperbolars for X real and approach asymptotically to the line $\Omega/X = 1/\hat{n}$ (or $v = c$) as $Re(X) \to \infty$ (see Fig. 2). Therefore, we see that the larger the ϵ/ϵ_0 ratio, the smaller the slopes of frequency bounds and, hence, the smaller the slopes of frequency branches for large $Re(X)$. The effect of ϵ/ϵ_0 on the dispersion curves of TE and TM waves for the case $\bar{b}/b = \infty$ has been examined in Ref.[3].

Once the roots ω and ξ are computed from (22), the amplitude ratios can be computed from (19) and (20) and the field components from (11) and (15).

Variation of the field components E_3 (or B_2) and B_1 as functions of x_2 are plotted in Fig. 3 for the first

three frequency branches ($p = 0, 1, 2$) at the normalized phase velocity $V = \infty$ (or $X = 0$), 1.0, $\frac{1}{2}(1+1/\hat{n})$, and $1/\hat{n}$ for $\hat{n}^2 = 50$. We see from Fig. 3 that, for $V < 1$ (or $v < c_0$), modes begin to decay in the free space and at $V = 1/\hat{n}$ (or $v = c$), they become completely confined inside the plate and are identical to TE modes in dielectric plate with perfect electric faces (see Fig. 3 of [3]).

4 Guided TM Waves

For straight-crested TM waves propagating in the x_1 direction in the plate ($|x_2| < b$), we take

$$\begin{aligned} A_1 &= (C_1 \sin \eta x_2 + C_2 \cos \eta x_2) e^{i(\xi x_1 - \omega t)} , \\ A_2 &= (G_1 \cos \eta x_2 + G_2 \sin \eta x_2) e^{i(\xi x_1 - \omega t)} , \\ A_3 &= 0 . \end{aligned} \tag{45}$$

Substitution of (45) into (10) yields (13) and the relations

$$C_1 = -i \left(\frac{\eta}{\xi}\right) G_1 , \quad C_2 = i \left(\frac{\eta}{\xi}\right) G_2 . \tag{46}$$

And by inserting (45) in (4) and eliminating C_a's ($a = 1, 2$) by (46), we have the following non-zero components

$$\begin{aligned} B_3 &= i\frac{\omega v}{c^2}(G_1 \cos \eta x_2 + G_2 \sin \eta x_2) e^{i(\xi x_1 - \omega t)}, \\ E_1 &= v\eta(G_1 \sin \eta x_2 - G_2 \cos \eta x_2) e^{i(\xi x_1 - \omega t)}, \\ E_2 &= \frac{c^2}{v} B_3 . \end{aligned} \tag{47}$$

Again, we see that A_i in (45) is the vector potential of magnetic waves transverse to the x_1 and x_2 axes, since $B_1 = B_2 = 0$.

For the TM waves in the free space ($b \le |x_2| \le b + \bar{b}$), we let

$$\begin{aligned} \bar{A}_1 &= (\pm \bar{C}_1 + \bar{C}_2) \sinh \bar{\eta}(b + \bar{b} \mp x_2) e^{i(\xi x_1 - \omega t)}, \\ \bar{A}_2 &= (\bar{G}_1 \pm \bar{G}_2) \cosh \bar{\eta}(b + \bar{b} \mp x_2) e^{i(\xi x_1 - \omega t)}, \\ \bar{A}_3 &= 0 . \end{aligned} \tag{48}$$

which are solutions of (10), provided (16) is satisfied and

$$\bar{C}_a = -i \left(\frac{\bar{\eta}}{\xi}\right) \bar{G}_a , \quad a = 1, 2 . \tag{49}$$

We note that in (48) the fulfilment of (21) has been assumed.

By substituting (49) into (48), and, in turn, into (4), we have

$$\begin{aligned} \bar{B}_3 &= i\frac{\omega v}{c_0^2}(\bar{G}_1 \pm \bar{G}_2) \cosh \bar{\eta}(b + \bar{b} \mp x_2) e^{i(\xi x_1 - \omega t)}, \\ \bar{E}_1 &= v\bar{\eta}(\pm \bar{G}_1 + \bar{G}_2) \sinh \bar{\eta}(b + \bar{b} \mp x_2) e^{i(\xi x_1 - \omega t)}, \\ \bar{E}_2 &= \frac{c_0^2}{v} \bar{B}_3 . \end{aligned} \tag{50}$$

We see that (48) or (50) satisfy the perfect electric boundary conditions (9) at $x_2 = \pm(b + \bar{b})$.

For the present case, the continuity conditions (6) and (8) reduce to

$$A_1 = \bar{A}_1 , \quad H_3 = \bar{H}_3 , \quad \text{at} \quad x_2 = \pm b . \tag{51}$$

Application of (51) to $(45)_1$, $(48)_1$, $(47)_1$, and $(50)_1$ gives

$$\begin{aligned} \eta G_1 \sin \eta b &= \bar{\eta} \bar{G}_1 \sinh \bar{\eta} \bar{b} , \\ \epsilon G_1 \cos \eta b &= \epsilon_0 \bar{G}_1 \cosh \bar{\eta} \bar{b} , \end{aligned} \tag{52}$$

for the symmetric modes of B_3, and

$$\begin{aligned} \eta G_2 \cos \eta b &= -\bar{\eta} \bar{G}_2 \sinh \bar{\eta} \bar{b} , \\ \epsilon G_2 \sin \eta b &= \epsilon_0 \bar{G}_2 \cosh \bar{\eta} \bar{b} , \end{aligned} \tag{53}$$

for the antisymmetric modes.

The frequency equations obtained from (52) and (53) can be written as

$$\tan \eta b = \pm \left(\frac{\epsilon \bar{\eta} \tanh \bar{\eta} \bar{b}}{\epsilon_0 \eta}\right)^{\pm 1} , \tag{54}$$

where + sign is for the symmetric modes and − sign for the antisymmetric modes.

By the use of relations (13) and (16), (54) is rewritten in the form

$$\tan \left(\frac{v^2}{c^2} - 1\right)^{\frac{1}{2}} \xi b = \pm \left[\frac{\epsilon \tanh (1 - v^2/c_0^2)^{\frac{1}{2}} \xi \bar{b}}{\epsilon_0 (v^2/c^2 - 1)^{\frac{1}{2}} / (1 - v^2/c_0^2)^{\frac{1}{2}}}\right]^{\pm 1} . \tag{55}$$

Certain analytical properties of the dispersion relations (54) for the TM waves are obtained by considerations similar to those in the previous section for the TE waves:

(1) When $\bar{b}$ approaches zero, the continuity conditions (51) reduce to

$$A_1(x_2 = \pm b) = 0 . \tag{56}$$

Then (52) and (53) reduce, respectively, to

$$\sin \eta b = 0 \quad \text{or} \quad \eta = \frac{n\pi}{2b}, \; n = 0, 2, 4, \cdots \tag{57}$$

for the symmetric modes, and

$$\cos \eta b = 0 \quad \text{or} \quad \eta = \frac{n\pi}{2b}, \; n = 1, 3, 5, \cdots \tag{58}$$

for the antisymmetric modes. These results are the same as those given by eq.(48) in [3].

(2) When v approaches c_0, $\bar{\eta} \to 0$. From $(52)_1$ and $(53)_1$, we have

$$\eta = \frac{n\pi}{2b} , \tag{59}$$

where $n = 0, 2, 4, \cdots$ for symmetric modes and $n = 1, 3, 5, \cdots$ for antisymmetric modes. The amplitude ratios, from $(52)_2$ and $(53)_2$, are

$$\left|\frac{\bar{G}_a}{G_a}\right| = \frac{\epsilon}{\epsilon_0}\,, \quad a = 1, 2 \quad \text{(no sum)}\,. \tag{60}$$

(3) When v approaches c and ξ approaches infinity, (55) are satisfied by

$$\eta = \lim_{\substack{v\to c,\ \xi\to\infty \\ p = 0, 1, 2, \cdots}} \left(\frac{v^2}{c^2} - 1\right)^{\frac{1}{2}} \xi = \frac{(p+1)\pi}{2b}\,, \tag{61}$$

or

$$\frac{\omega^2}{c^2} - \xi^2 = \left[\frac{(p+1)\pi}{2b}\right]^2\,, \tag{62}$$

where $p = 0, 2, 4, \cdots$ for symmetric modes, and $p = 1, 3, 5, \cdots$ for antisymmetric modes. From $(52)_2$ and $(53)_2$

$$\left|\frac{\bar{G}_a}{G_a}\right| = 0\,, \quad a = 1, 2 \quad \text{(no sum)}\,, \tag{63}$$

and from $(52)_1$ and $(53)_1$

$$\lim_{\substack{v\to c,\ \xi\to\infty}} \left|\frac{\bar{G}_a}{G_a}\bar{\eta}\sinh\bar{\eta}\bar{b}\right| = \frac{(p+1)\pi}{2b}\,. \tag{64}$$

Therefore, we see, in $(50)_2$, $\bar{E}_1 = 0$ for $b < |x_2| \leq b + \bar{b}$, but it has non-zero values on the faces of the plate.

(4) When $v > c_0$, $\bar{\eta}$ is imaginary. The field components (50) and frequency equations (54) become, respectively,

$$\begin{aligned}
\bar{B}_3 &= i\frac{\omega v}{c_0^2}(\bar{G}_1 \pm \bar{G}_2)\cos\hat{\eta}(b + \bar{b} \mp x_2)e^{i(\xi x_1 - \omega t)}, \\
\bar{E}_1 &= -v\hat{\eta}(\pm\bar{G}_1 + \bar{G}_2)\sin\hat{\eta}(b + \bar{b} \mp x_2)e^{i(\xi x_1 - \omega t)}, \\
\bar{E}_2 &= \frac{c_0^2}{v}\bar{B}_3\,,
\end{aligned} \tag{65}$$

and

$$\tan\eta b = \mp\left(\frac{\epsilon_0\eta}{\epsilon\hat{\eta}\tan\hat{\eta}\bar{b}}\right)^{\mp 1}\,, \tag{66}$$

where the $-$ sign is for the symmetric modes and $+$ sign for the antisymmetric modes, and $\hat{\eta}$ is defined by (34).

(5) When the wave number is imaginary, the field components for the plate and the surrounding free space are still given by (47) and (65) except that ξ is replaced by $i\xi'$ and η, $\hat{\eta}$ are given by (39).

(6) As $\xi \to 0$, the frequency equations (66) reduce to

$$\tan\frac{\omega b}{c} = \mp\left(\frac{c_0/c}{\tan\frac{\omega\bar{b}}{c_0}}\right)^{\mp 1}\,. \tag{67}$$

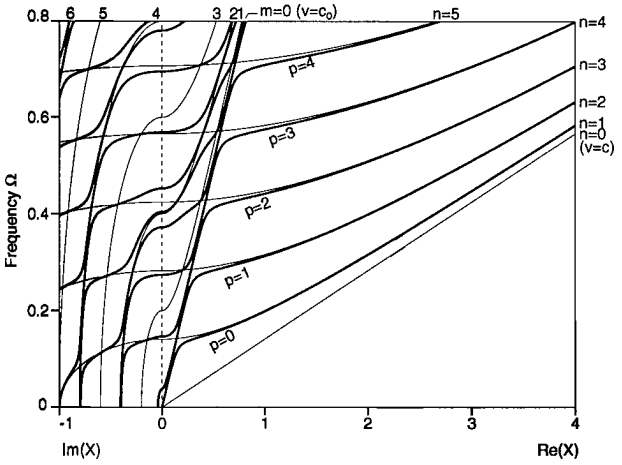


Figure 4: Dispersion curves (p even for symmetric modes; p odd for antisymmetirc modes) and their bounds (thin lines denoted by m and n) for guided TM waves in a dielectric plate enclosed by two parallel perfect conductors ($\epsilon/\epsilon_0 = 50$ and $\bar{b}/b = 5$).

from which the cut–off frequencies can be calculated.

We note that (67) and (37) are of the same form except in that the minus and plus signs in (67) are, respectively, for the symmetric and antisymmetric H_3 modes, while the plus and minus signs in (37) are, respectively, for the symmetric and antisymmetric E_3 modes.

(7) Simple roots of (66) can be obtained by letting $\bar{\eta} = i\hat{\eta}$ in (52) and (53). We see that (52), for the symmetric B_3 modes, are satisfied by

$$\begin{aligned}
&\sin\eta b = 0, \quad \sin\hat{\eta}\bar{b} = 0, \quad \left|\frac{\bar{G}_1}{G_1}\right| = \frac{\epsilon}{\epsilon_0} \\
&\text{and} \\
&\cos\eta b = 0, \quad \cos\hat{\eta}\bar{b} = 0, \quad \left|\frac{\bar{G}_1}{G_1}\right| = \frac{\eta}{\hat{\eta}}\,.
\end{aligned} \tag{68}$$

And (53), for the antisymmetric B_3 modes, are satisfid by

$$\begin{aligned}
&\cos\eta b = 0, \quad \sin\hat{\eta}\bar{b} = 0, \quad \left|\frac{\bar{G}_2}{G_2}\right| = \frac{\epsilon}{\epsilon_0} \\
&\text{and} \\
&\sin\eta b = 0, \quad \cos\hat{\eta}\bar{b} = 0, \quad \left|\frac{\bar{G}_2}{G_2}\right| = \frac{\eta}{\hat{\eta}}\,,
\end{aligned} \tag{69}$$

From (68) and (69), we may similarly conclude that the intersections of $(44)_1$ and $(44)_2$ with n even and m even (including $m = 0$) and n odd and m odd are simple solutions for the symmetric B_3 modes, while the intersections of the curves for n odd and m even

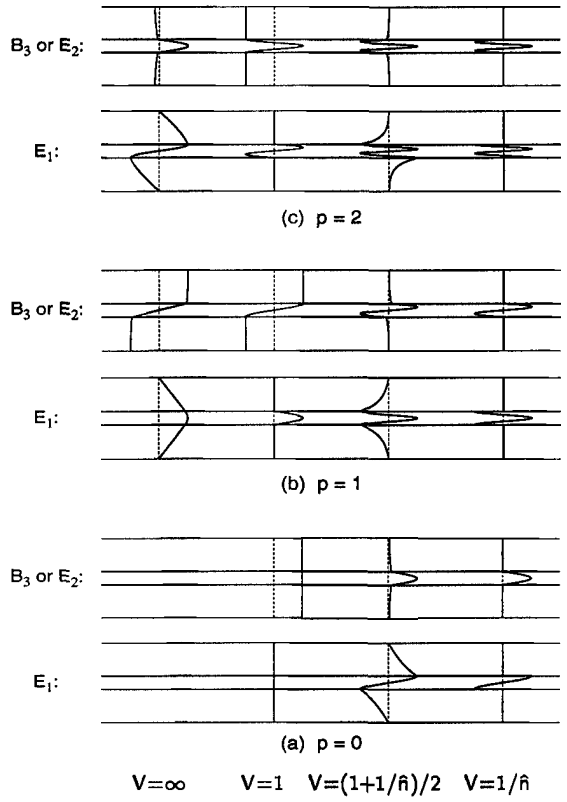


Figure 5: Modes of guided TM waves in a dielectric plate enclosed by parallel perfect conductors on the first three frequency branches ($p = 0, 1, 2$) at $V = \infty$, 1, $(1 + 1/\hat{n})/2$, and $1/\hat{n}$ for $\hat{n}^2 = 50$ and $\bar{b}/b = 5$.

($m = 0$ included) and for n even and m odd are simple solutions for the antisymmetric B_3 modes.

Dispersion curves, i.e. Ω vs. X curves, for TM waves are calculated from (54), (66), and (67) for $\epsilon/\epsilon_0 = 50$ and $\bar{b}/b = 5$ and they are plotted as thick solid lines depicted by $p = 0, 2, 4, \cdots$ for the symmetric B_3 modes and $p = 1, 3, 5, \cdots$ for the antisymmetric B_3 modes in Fig. 4.

Variations of the field components H_3 (or D_2) and E_1 as functions of x_2 are plotted in Fig. 5 for the first three branches ($p = 0, 1$, and 2) and at the normalized phase velocity $V = \infty$ (or $X = 0$), 1.0, $\frac{1}{2}(1+1/\hat{n})$, and $1/\hat{n}$, where $\hat{n}^2 = 50$. We note that in Fig. 5 the modes of the first frequency branch for $p = 0$ at $V = \infty$ (or at $X = 0$) are not given because at this point the frequency is zero. We see in Fig. 5 that, at $V = 1/\hat{n}$ (or $v = c$), the modes are completely confined inside the plate and are identical to TM modes in dielectric plate with perfect magnetic faces (see [3]).

5 Partition of Energy

The electromagnetic energy per unit area (in the x_1x_3 plane) in the plate is defined by

$$U = \frac{1}{T}\int_{t_0}^{t_0+T}\int_{-b}^{b}(\epsilon E_i E_i + \mu_0 H_i H_i)\,dx_2 dt\,, \quad (70)$$

where $T = 2\pi/\omega$ is the period of the EM wave. The EM energy per unit area in the surrouding free space is defined by

$$\bar{U} = \frac{1}{T}\int_{t_0}^{t_0+T}\left[\int_{-b-\bar{b}}^{-b}(\epsilon_0 E_i E_i + \mu_0 H_i H_i)\,dx_2 + \int_{b}^{b+\bar{b}}(\epsilon_0 E_i E_i + \mu_0 H_i H_i)\,dx_2\right]dt\,, \quad (71)$$

We denote the ratio of the energy of the plate per unit area to the energy of the free space per unit area by R_e. Hence

$$R_e = \frac{U}{\bar{U}}\,. \quad (72)$$

Substitution of the non-zero fields for TE and TM waves into (70) and (71), and, in turn, into (72), we obtain the following expressions for the energy ratios: Symmetric TE waves

$$R_e = \left|\frac{C_1}{\bar{C}_1}\right|^2 \sinh^2\bar{\eta}(b+\bar{b})\frac{b\left(\hat{n}^2V^2 + \dfrac{\sin 2\eta b}{2\eta b}\right)}{\bar{b}\left|\dfrac{\sinh 2\bar{\eta}\bar{b}}{2\bar{\eta}\bar{b}} - V^2\right|}\,, \quad (73)$$

Antisymmetric TE waves

$$R_e = \left|\frac{C_2}{\bar{C}_2}\right|^2 \sinh^2\bar{\eta}(b+\bar{b})\frac{b\left(\hat{n}^2V^2 - \dfrac{\sin 2\eta b}{2\eta b}\right)}{\bar{b}\left|\dfrac{\sinh 2\bar{\eta}\bar{b}}{2\bar{\eta}\bar{b}} - V^2\right|}\,, \quad (74)$$

Symmetric TM waves

$$R_e = \left|\frac{G_1}{\bar{G}_1}\right|^2 \hat{n}^2\frac{b\left(\hat{n}^2V^2 + \dfrac{\sin 2\eta b}{2\eta b}\right)}{\bar{b}\left(\dfrac{\sinh 2\bar{\eta}\bar{b}}{2\bar{\eta}\bar{b}} + V^2\right)}\,, \quad (75)$$

Antisymmetric TM waves

$$R_e = \left|\frac{G_2}{\bar{G}_2}\right|^2 \hat{n}^2\frac{b\left(\hat{n}^2V^2 - \dfrac{\sin 2\eta b}{2\eta b}\right)}{\bar{b}\left(\dfrac{\sinh 2\bar{\eta}\bar{b}}{2\bar{\eta}\bar{b}} + V^2\right)}\,. \quad (76)$$

In order to see the effect of the free-space gap on the partition of energy, the energy ratios R_e as a function of frequency Ω are computed for the fundamental

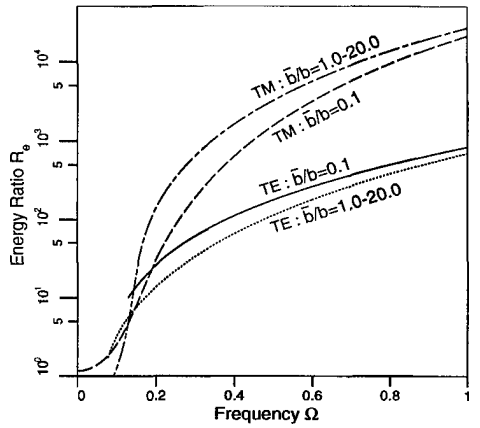


Figure 6: Energy ratio vs frequency for the fundamental mode (p=0) of TE and TM waves and for $\bar{b}/b = 0.1$ and 1.0–20.0 ($\epsilon/\epsilon_0 = 50$).

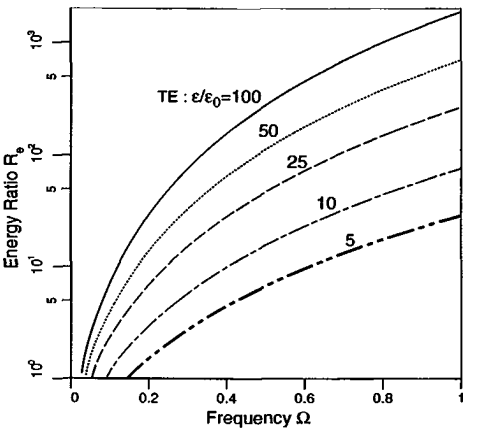


Figure 7: R_e vs Ω curves for the fundamental mode of TE waves (p=0) for $\bar{b}/b = 5$ and $\epsilon/\epsilon_0 = 5, 10, 25, 50,$ and 100.

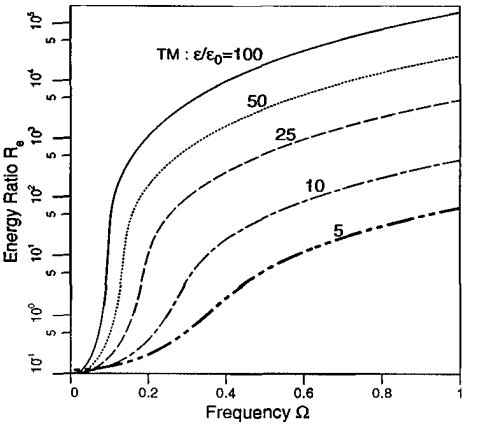


Figure 8: R_e vs Ω curves for the fundamental mode of TE waves (p=0) for $\bar{b}/b = 5$ and $\epsilon/\epsilon_0 = 5, 10, 25, 50,$ and 100.

mode (p=0) of the TE waves with $\epsilon/\epsilon_0 = 50$ and $\bar{b}/b = 1$ and 0.1 as shown in Fig. 6. The energy ratios for the fundamental mode of the TM waves are similarly computed and also shown in Fig. 6.

It may be seen in Fig. 6 that the higher the frequency (or the shorter the wave length of the mode) the higher the energy ratio R_e. The value of R_e for the TM mode is about ten times greater than that for the TE mode for a given frequency $\Omega > 0.2$, since it appears that the TM modes are more confined in the dielectric plate than the TE modes are (see Figs. 2 and 5). Also from Fig. 6 we see that the increase of $\bar{b}/b$ ratio from 0.1 to 1.0 increases R_e ratio for the TM modes, but decreases the R_e ratio for the TE modes. We note that R_e vs Ω curves remain practically constant for a range of values $\bar{b}/b$ varying from 1.0 to 20.0 as indicated in Fig. 6. Computations of R_e vs Ω curves are also made for higher modes (p=1,2,3,...) but not shown here. They exhibit similar behavior as that of the fundamental mode (p=0), but with lower values of R_e for higher modes.

The effect of ϵ/ϵ_0 ratio on the energy partition is examined by computing R_e vs. Ω for the fundamental mode (p=0) for $\bar{b}/b = 5$ and $\epsilon/\epsilon_0 = 5, 10, 25, 50,$ and 100. The computational results are shown in Figs. 7 and 8 for the TE and TM waves, respectively. It is seen from Figs. 7 and 8 that the higher the ϵ/ϵ_0 ratio (or the higher the refractive index) the larger the energy ratio R_e.

References

[1] D. Marcuse, *Light Transmission Optics*, Krieger, Florida, 1989, p. 305.

[2] R. E. Collin, *Field Theory of Guided Waves*, McGraw-Hill, New York, 1960, p. 470.

[3] P. C. Y. Lee and J. S. Yang, "Two-dimensional equations for guided electromagnetic waves in dielectric plates surrounded by free space," *J. Appl. Phys.*, **73** (11), pp. 7069-7082, 1993.

[4] P.C.Y. Lee and J.D. Yu, "Guided Electromagnetic Waves in Anisotropic Dielectric Plates," *J. Appl. Phys.*, **74**(8), pp. 4823-4839, 1993.

[5] E. G. Newman and R. D. Mindlin, "Vibrations of a monoclinic crystal plate," *J. Acou. Soc. Am.*, **29** (11), pp. 1206-1219, 1957.

[6] M. Born and E. Wolf, *Principles of Optics*, 3rd. ed., Pergarmon Press, New York, 1965, pp. 71-73.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

The Shear Coefficient for Quartz Crystal of Rectangular Cross Section in Timoshenko's Beam Theory

Hirofumi KAWASHIMA

Seiko Instruments Inc.
Matsudo-shi,Japan 271

Abstract

This paper describes the shear coefficient K of transverse vibration for quartz crystal that is an anisotropic material. In this paper, we derive the shear coefficient K for quartz crystal of rectangular cross section in Timoshenko's beam theory of three dimensional elasticity. The theoretical result gives K as a function of the transformed elastic stiffness constants $c'_{22}, \hat{c}_{44}, c'_{66}$. For an isotropic material it agrees completely with that derived by G.R.Cowper[2]. In addition, A value of K and its temperature coefficient versus a cut angle is calculated, so that K has a value of 0.796 to 0.867 and its temperature coefficient of the first order varies from -15.3×10^{-6}/℃ to $+13.2\times10^{-6}$/℃, when a Z-plate of quartz crystal is rotated with a cut angle θ of 0 to 180 degrees about the x-axis. The values of K are then compared with those obtained by other authors.

1. Introduction

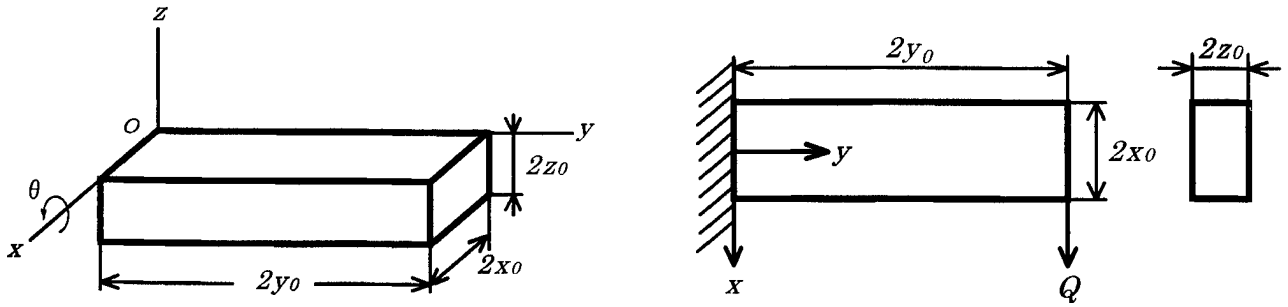
The one-dimensional theory of beams can be extended in the range of applicability by taking account of rotary inertia and transverse shear deformations. The equations with these effects are generally referred to as Timoshenko's beam equations[1]. In these equations the effective transverse shear strain is taken as equal to the average shear stress on a cross section divided by the product of the shear modulus (elastic stiffness constant) and the shear coefficient, K. In addition, the shear coefficient K is commonly defined as the ratio of the average shear strain on a section to the shear strain at the centroid.

However, there occurs a big difference between the analysis results using the commonly defined K and the experimental ones[2]. Leibowitz and Kennard have criticized the definition, pointing out an unwarranted assumption in the underlying analysis, and Cowper has agreed with their criticism and has pointed out that further study of the shear coefficient is needed[2]. In addition, Mindlin, et al[3] and Goodman[4] have pointed out that Timoshenko's beam equations with the customary values of K lead to unsatisfactory results when the high resonance frequencies of vibrating beams are calculated, and have insisted that K should be adjusted arbitrarily in order to obtain better results. However, they do not discuss the basic definition of K. Instead they point out that the distribution of shear strain over a cross section varies with the mode of vibration, i.e., frequency.

According to a way of thinking of Leibowitz and Kennard, Cowper has examined the shear coefficient K for an isotropic material in detail and better results are obtained[2]. It must be noted, however, that K for an anisotropic material such as quartz crystal is not examined at all.

In this paper, the shear coefficient K for quartz crystal which is an anisotropic material is described. In analysis procedure, first, the shear coefficient K is derived from Timoshenko's beam theory and Cowper's definition and is found to be given as a function of stress-component T_6 and residual displacement v_y. Next, T_6 and v_y are derived from the equations of equilibrium and the boundary conditions. Substitution of T_6 and v_y into K gives a function of harmonic function χ'. The shear coefficient K is, therefore, calculated using χ' derived. Finally, the calculated values of K and its first-, second- and third-order temperature coefficients α, β, γ are shown to a cut angle rotated about the x-axis.

0-7803-2500-1/95 $4.00 © 1995 IEEE



(a) (b)

Fig. 1 Flexural quartz crystal resonator of beam and its coordinate system (a), and front view and sectional view when lateral force Q is applied at the end (b).

2. Derivative of Shear Coefficient K

Figure 1 shows a flexural quartz crystal resonator of a beam and its coordinate system(a), and front view and sectional view when lateral force Q is applied at the end (b). Now, as length $2y_0$ is much larger than width $2x_0$ and thickness $2z_0$, stress-components T_1, T_3, T_5 become

$$T_1 = T_3 = T_5 = 0. \tag{1}$$

For the rotation about the x-axis (rotation angle θ) , a relationship between stress T_p and strain S_q is given as

$$\begin{aligned} T_1 &= c_{11}S_1 + c_{12}S_2 + c_{13}S_3 + c_{14}S_4 \\ T_2 &= c_{21}S_1 + c_{22}S_2 + c_{23}S_3 + c_{24}S_4 \\ T_3 &= c_{31}S_1 + c_{32}S_2 + c_{33}S_3 + c_{34}S_4 \\ T_4 &= c_{41}S_1 + c_{42}S_2 + c_{43}S_3 + c_{44}S_4 \\ T_5 &= c_{55}S_5 + c_{56}S_6 \\ T_6 &= c_{65}S_5 + c_{66}S_6. \end{aligned} \tag{2}$$

Accordingly, from the relation of eq.(1), equation (2) is simplified into

$$\begin{aligned} T_2 &= c'_{22}S_2 + c'_{24}S_4 \\ T_4 &= c'_{42}S_2 + c'_{44}S_4 \\ T_6 &= c'_{66}S_6, \end{aligned} \tag{3}$$

where, $c'_{pq} = \bar{c}_{pq} - \bar{c}_{1p}\bar{c}_{q1}/\bar{c}_{11} \quad (p,q = 2,4)$

$\bar{c}_{pq} = c_{pq} - c_{3p}c_{q3}/c_{33} \quad (p,q = 1,2,4)$

$c'_{66} = c_{66} - c_{56}c_{65}/c_{55}.$

The definition of displacement and rotation angle accords to reference(2). The transverse mean deflection U(y) of the cross section and the mean displacement V(y) of the cross section in the y-axis direction are defined as

$$\mathrm{U(y)} = \frac{1}{A}\int_{-z_0}^{z_0}\int_{-x_0}^{x_0} u_x \,\mathrm{d}x\,\mathrm{d}z \tag{4}$$

and

$$\mathrm{V(y)} = \frac{1}{A}\int_{-z_0}^{z_0}\int_{-x_0}^{x_0} u_y \,\mathrm{d}x\,\mathrm{d}z, \tag{5}$$

where A is the area of the cross section, u_x and u_y are, respectively, the x-component and the y-component of displacement of a point of the beam. In addition, the mean angle Φ(y) of rotation of a cross section about the neutral axis is also defined as

$$\Phi\mathrm{(y)} = \frac{1}{I_z}\int_{-z_0}^{z_0}\int_{-x_0}^{x_0} x u_y \,\mathrm{d}x\,\mathrm{d}z, \tag{6}$$

where, $I_z = \int_{-z_0}^{z_0}\int_{-x_0}^{x_0} x^2 \mathrm{d}x\,\mathrm{d}z.$

Further, residual displacements v_x, v_y are defined by the following relation, namely

$$\begin{aligned} u_x &= \mathrm{U(y)} + v_x \\ u_y &= \mathrm{V(y)} + x\Phi\mathrm{(y)} + v_y. \end{aligned} \tag{7}$$

It follows from eq.(7) that

$$\int_{-z_0}^{z_0}\int_{-x_0}^{x_0} v_x \,\mathrm{d}x\,\mathrm{d}z = \int_{-z_0}^{z_0}\int_{-x_0}^{x_0} v_y \,\mathrm{d}x\,\mathrm{d}z$$

$$= \int_{-z_0}^{z_0}\int_{-x_0}^{x_0} x v_y \,\mathrm{d}x\,\mathrm{d}z = 0. \tag{8}$$

The residual displacement v_y essentially represents the warping of the cross section. From eq.(3) the stress-strain relation is

$$T_6 = c'_{66}\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x}\right). \tag{9}$$

Substitution of eq.(7) into eq.(9) gives

$$\frac{\partial U(y)}{\partial y}+\Phi(y)=\frac{T_6}{c'_{66}}-\frac{\partial v_x}{\partial y}-\frac{\partial v_y}{\partial x}. \quad (10)$$

Integration of eq.(10) over the cross section then yields, in view of eq.(8),

$$\frac{\partial U(y)}{\partial y}+\Phi(y) = \frac{1}{Ac'_{66}}\int_{-z_0}^{z_0}\int_{-x_0}^{x_0}\left(T_6-c'_{66}\frac{\partial v_y}{\partial x}\right)dxdz. \quad (11)$$

The right term of eq.(11) can be replaced by

$$\frac{Q}{KAc'_{66}}. \quad (12)$$

The K of eq.(12) is called the shear coefficient. From the relation of eqs.(11) and (12) the shear coefficient K is rewritten as

$$K=\frac{Q}{\int_{-z_0}^{z_0}\int_{-x_0}^{x_0}\left(T_6-c'_{66}\frac{\partial v_y}{\partial x}\right)dxdz}. \quad (13)$$

That is, the shear coefficient per unit force K/Q is given as a function of stress T_6, elastic stiffness constant c'_{66} and residual displacement v_y. The stress T_6 and the residual displacement v_y are derived next.

3. Derivative of Stress T_6 and Residual Displacement v_y

3.1 Equation of equilibrium

The bending moment M at the cross section distant y from the fixed end is

$$M = Q(2y_0-y). \quad (14)$$

Here, we assume that stress-component T_2 on any element of this section is given by the equation

$$T_2 = -Q(2y_0-y)x/I_z. \quad (15)$$

In addition, the equations of equilibrium are given as

$$\begin{aligned}&\frac{\partial T_1}{\partial x}+\frac{\partial T_6}{\partial y}+\frac{\partial T_5}{\partial z}=0\\&\frac{\partial T_6}{\partial x}+\frac{\partial T_2}{\partial y}+\frac{\partial T_4}{\partial z}=0\\&\frac{\partial T_5}{\partial x}+\frac{\partial T_4}{\partial y}+\frac{\partial T_3}{\partial z}=0.\end{aligned} \quad (16)$$

Since two of the equations of equilibrium from the relation of eq.(1) must be satisfied with

$$\frac{\partial T_6}{\partial y}=0 \text{ and } \frac{\partial T_4}{\partial y}=0, \quad (17)$$

it follows that T_6 and T_4 must be independent of y. By use of the relation of eq.(15), the second of the equations of equilibrium reduces to

$$\frac{\partial T_6}{\partial x}+\frac{\partial T_4}{\partial z}+\frac{Q}{I_z}x=0. \quad (18)$$

In addition, there is a relationship with respect to strain S_q (q=3 to 6) as follows:

$$\frac{\partial}{\partial z}\left(\frac{\partial S_1}{\partial x}+\frac{\partial S_5}{\partial y}-\frac{\partial S_6}{\partial z}\right)=2\frac{\partial^2 S_3}{\partial x\partial y}. \quad (19)$$

From the relation of eqs.(2) and (17), it follows that $\partial S_5/\partial y=0$, and since strain S_4 is given as a function of x and z, equation(19) is transformed into

$$\frac{\partial S_4}{\partial x}-\frac{\partial S_6}{\partial z}=2\tau-\frac{2\sigma_z}{c'_{22}}\frac{Q}{I_z}z, \quad (20)$$

where 2τ is a constant of integration and σ_z is Poisson's ratio;and from this equation it follows that S_4 and S_6 which satisfy eq.(20) can be expressed in the forms

$$\begin{aligned}S_4&=\tau x+\frac{\partial\phi_0}{\partial z}\\S_6&=-\tau z+\frac{\partial\phi_0}{\partial x}+\frac{\sigma_z}{c'_{22}}\frac{Q}{I_z}z^2,\end{aligned} \quad (21)$$

where ϕ_0 is a function of x and z.

From eqs.(21) and (3) the equation of equilibrium of eq.(18) reduces to

$$c'_{66}\frac{\partial^2\phi_0}{\partial x^2}+\hat{c}_{44}\frac{\partial^2\phi_0}{\partial z^2}+\frac{Q}{I_z}x=0, \quad (22)$$

where, $\hat{c}_{44}=c'_{44}-c'^2_{24}/c'_{22}$ $(\because c'_{24}=c'_{42})$.

In addition, we change the variables by putting

$$\begin{aligned}&x'=x\sqrt{\frac{\hat{c}_{44}+c'_{66}}{2c'_{66}}},\quad z'=z\sqrt{\frac{\hat{c}_{44}+c'_{66}}{2\hat{c}_{44}}},\\&\phi'_0=\phi_0\frac{\hat{c}_{44}+c'_{66}}{2\sqrt{\hat{c}_{44}c'_{66}}}.\end{aligned} \quad (23)$$

The equation of equilibrium of eq.(22) is then rewritten as

$$\frac{\partial^2\phi'_0}{\partial x'^2}+\frac{\partial^2\phi'_0}{\partial z'^2}+\frac{Q}{c'_{22}I_z}\Xi x'=0, \quad (24)$$

where, $\Xi=\frac{c'_{22}}{\hat{c}_{44}}\sqrt{\frac{2\hat{c}_{44}}{\hat{c}_{44}+c'_{66}}}$.

3.2 Boundary condition

On the other hand, the condition that the bounding surface is free from traction must be

$$T_6\cos(x,\nu)+T_4\cos(z,\nu)=0. \tag{25}$$

Here ν denotes the direction of the normal to a plane, and the general form of $\cos(a,\nu)$ has the following relation

$$\cos(a,\nu)=\begin{cases}1 & |\nu|=a\\ 0 & |\nu|\neq a\end{cases}. \tag{26}$$

From the relations of eqs.(3),(15) and (21), the equation of the boundary condition becomes

$$\begin{aligned}\frac{\partial\phi_0}{\partial\nu}&=\tau\{c'_{66}\,z\cos(x,\nu)-\hat{c}_{44}\,x\cos(z,\nu)\}\\&\quad-\frac{c'_{66}}{c'_{22}}\sigma_z\frac{Q}{I_z}z^2\cos(x,\nu)\\&\quad+\frac{c'_{42}}{c'_{22}}\frac{Q}{I_z}(2y_0-y)x\cos(z,\nu),\end{aligned} \tag{27}$$

which holds at all points of the bounding curve of any cross section. Here $d\nu$ denotes the element of the outward-drawn normal to this curve. In addition, if f(x,z)=0 is the equation of the bounding curve, $\cos(x,\nu)$ and $\cos(z,\nu)$ become equal to $\partial f/\partial x$ and $\partial f/\partial z$, respectively. Now, when a new function F(x',z') is satisfied with

$$\mathrm{F}(\mathrm{x}',\mathrm{z}')\equiv\mathrm{f}\left(\mathrm{x}'\sqrt{\frac{2c'_{66}}{\hat{c}_{44}+c'_{66}}},\mathrm{z}'\sqrt{\frac{2\hat{c}_{44}}{\hat{c}_{44}+c'_{66}}}\right), \tag{28}$$

the equation f(x,z)=0 becomes F(x',z')=0. Therefore, by taking account of eq.(28) and the relation of $\partial F/\partial x'=\cos(x',\nu')$ and $\partial F/\partial z'=\cos(z',\nu')$, the equation of the boundary condition of eq.(27) can be rewritten as

$$\begin{aligned}\frac{\partial\phi'_0}{\partial\nu'}&=\tau\{z'\cos(x',\nu')-x'\cos(z',\nu')\}\\&\quad-\frac{Q}{c'_{22}I_z}\Xi'\sigma_z z'^2\cos(x',\nu')\\&\quad+\frac{Q}{c'_{22}I_z}\frac{c'_{42}}{\hat{c}_{44}}(2y_0-y)x'\cos(z',\nu'),\end{aligned} \tag{29}$$

where, $\Xi'=\sqrt{\dfrac{2\hat{c}_{44}}{\hat{c}_{44}+c'_{66}}}$.

Here $d\nu'$ is the element of the normal to the transformed boundary. Equations (24) and (29) are further simplified by putting

$$\phi'_0=\tau\phi'-\frac{Q}{c'_{22}I_z}\left\{\chi'-\frac{1}{6}\left(1-\frac{1}{2}\Xi\right)x'^3+\frac{1}{2}\left(1+\frac{1}{2}\Xi\right)x'z'^2\right\}, \tag{30}$$

then ϕ' is the torsion function for the cross section. Substitution eq.(30) into eq.(24) gives χ' and ϕ' which are, respectively, satisfied with

$$\frac{\partial^2\chi'}{\partial x'^2}+\frac{\partial^2\chi'}{\partial z'^2}=0 \tag{31}$$

and

$$\frac{\partial^2\phi'}{\partial x'^2}+\frac{\partial^2\phi'}{\partial z'^2}=0. \tag{32}$$

From eq.(30) $\partial\phi'_0/\partial x'$ and $\partial\phi'_0/\partial z'$ are given as

$$\frac{\partial\phi'_0}{\partial x'}=\tau\frac{\partial\phi'}{\partial x'}-\frac{Q}{c'_{22}I_z}\left\{\frac{\partial\chi'}{\partial x'}-\frac{1}{2}\left(1-\frac{1}{2}\Xi\right)x'^2+\frac{1}{2}\left(1+\frac{1}{2}\Xi\right)z'^2\right\}$$

and

$$\frac{\partial\phi'_0}{\partial z'}=\tau\frac{\partial\phi'}{\partial z'}-\frac{Q}{c'_{22}I_z}\left\{\frac{\partial\chi'}{\partial z'}+\left(1+\frac{1}{2}\Xi\right)x'z'\right\}, \tag{33}$$

therefore, $\partial\phi'_0/\partial\nu'$ becomes

$$\begin{aligned}\frac{\partial\phi'_0}{\partial\nu'}&=\frac{\partial\phi'_0}{\partial x'}\frac{\partial x'}{\partial\nu'}+\frac{\partial\phi'_0}{\partial z'}\frac{\partial z'}{\partial\nu'}\\&=\tau\left\{\frac{\partial\phi'}{\partial x'}\cos(x',\nu')+\frac{\partial\phi'}{\partial z'}\cos(z',\nu')\right\}\\&\quad-\frac{Q}{c'_{22}I_z}\left[\left\{\frac{\partial\chi'}{\partial x'}-\frac{1}{2}\left(1-\frac{1}{2}\Xi\right)x'^2+\frac{1}{2}\left(1+\frac{1}{2}\Xi\right)z'^2\right\}\cos(x',\nu')\right.\\&\quad\left.+\left\{\frac{\partial\chi'}{\partial z'}+\left(1+\frac{1}{2}\Xi\right)x'z'\right\}\cos(z',\nu')\right].\end{aligned} \tag{34}$$

Substituting this relation into eq.(29), $\partial\chi'/\partial\nu'$ yields

$$\begin{aligned}\frac{\partial\chi'}{\partial\nu'}&=\left\{\frac{1}{2}\left(1-\frac{1}{2}\Xi\right)x'^2-\frac{1}{2}\left(1+\frac{1}{2}\Xi-2\Xi'\sigma_z\right)z'^2\right\}\cos(x',\nu')\\&\quad-\left\{\left(1+\frac{1}{2}\Xi\right)z'+\frac{c'_{42}}{\hat{c}_{44}}(2y_0-y)\right\}x'\cos(z',\nu').\end{aligned} \tag{35}$$

3.3 Stress T_6

Since strains S_4 and S_6 are ,respectively, given as a function of $\partial\phi_0/\partial x$ and $\partial\phi_0/\partial z$, $\partial\phi_0/\partial x$ is first calculated from eq.(30) and next $\partial\phi_0/\partial z$, so that

$$\frac{\partial\phi_0'}{\partial x'}=\frac{\partial\phi_0'}{\partial x}\frac{\partial x}{\partial x'}+\frac{\partial\phi_0'}{\partial z}\frac{\partial z}{\partial x'}=\frac{1}{\Xi'}\frac{\partial\phi_0}{\partial x}. \tag{36}$$

Substituting this relation into the first equation of eq.(33), $\partial\phi_0/\partial x$ becomes

$$\frac{\partial\phi_0}{\partial x}=\tau\Xi'\frac{\partial\phi'}{\partial x'}-\frac{Q}{c_{22}'I_z}\Xi'\times\left\{\frac{\partial\chi'}{\partial x'}-\frac{1}{2}\left(1-\frac{1}{2}\Xi\right)x'^2+\frac{1}{2}\left(1+\frac{1}{2}\Xi\right)z'^2\right\}. \tag{37}$$

Quite similarly, from the second equation of eq.(33), $\partial\phi_0/\partial z$ becomes

$$\frac{\partial\phi_0}{\partial z}=\tau\Xi''\frac{\partial\phi'}{\partial z'}-\frac{Q}{c_{22}'I_z}\Xi''\left\{\frac{\partial\chi'}{\partial z'}+\left(1+\frac{1}{2}\Xi\right)x'z'\right\}, \tag{38}$$

where, $\Xi''=\sqrt{\dfrac{2c_{66}'}{\hat{c}_{44}+c_{66}'}}$.

Substitution of eqs.(37) and (38) into eq.(21) yields strains S_4 and S_6 as follows:

$$S_4=\tau\left(x+\Xi''\frac{\partial\phi'}{\partial z'}\right)-\frac{Q}{c_{22}'I_z}\left\{\Xi''\frac{\partial\chi'}{\partial z'}+\frac{1}{2}\frac{c_{22}'}{\hat{c}_{44}}\left(1+\frac{2}{\Xi}\right)xz\right\} \tag{39}$$

and

$$S_6=\tau\left(-z+\Xi'\frac{\partial\phi'}{\partial x'}\right)-\frac{Q}{c_{22}'I_z}\left[\Xi'\frac{\partial\chi'}{\partial x'}+\frac{1}{4}\frac{c_{22}'}{\hat{c}_{44}}\times\left\{\frac{\hat{c}_{44}}{c_{66}'}\left(1-\frac{2}{\Xi}\right)x^2+\left(1+\frac{2}{\Xi}-4\frac{\hat{c}_{44}}{c_{22}'}\sigma_z\right)z^2\right\}\right]. \tag{40}$$

Consequently, stress T_6 is given as

$$T_6=c_{66}'\tau\left(-z+\Xi'''\frac{\partial\phi'}{\partial x}\right)-\frac{c_{66}'Q}{c_{22}'I_z}\times\left[\Xi'''\frac{\partial\chi'}{\partial x}+\frac{1}{4}\left\{\frac{c_{22}'}{c_{66}'}\left(1-\frac{2}{\Xi}\right)x^2+\frac{c_{22}'}{\hat{c}_{44}}\left(1+\frac{2}{\Xi}-4\frac{\hat{c}_{44}}{c_{22}'}\sigma_z\right)z^2\right\}\right], \tag{41}$$

where, $\Xi'''=\Xi'\Xi''=\dfrac{2\sqrt{\hat{c}_{44}c_{66}'}}{\hat{c}_{44}+c_{66}'}$.

3.4 Residual displacement v_y

As $\partial S_5/\partial y=0$ is taken into account, strains S_2,S_4 and S_6 have the relation

$$2\frac{\partial^2 S_2}{\partial z\partial x}=\frac{\partial}{\partial y}\left(\frac{\partial S_4}{\partial x}+\frac{\partial S_6}{\partial z}\right). \tag{42}$$

Deriving $\partial S_4/\partial x$ and $\partial S_6/\partial z$ from eqs.(39) and (40), substitution of the derivatives into eq.(42) gives

$$2\frac{\partial^2 S_2}{\partial z\partial x}=\frac{\partial}{\partial y}\left[2\tau\Xi'''\frac{\partial^2\phi'}{\partial x\partial z}-\frac{Q}{c_{22}'I_z}\left\{2\Xi'''\frac{\partial^2\chi'}{\partial x\partial z}+\frac{c_{22}'}{\hat{c}_{44}}\left(1+\frac{2}{\Xi}-2\frac{\hat{c}_{44}}{c_{22}'}\sigma_z\right)z\right\}\right]. \tag{43}$$

Integration of eq.(43) over the cross section and the relation of $S_2=\partial u_y/\partial y$ yield the displacement u_y as

$$u_y=x\,\mathrm{f_z(y)}-\frac{Q}{c_{22}'I_z}\left\{\Xi'''\chi'+\frac{1}{4}\frac{c_{22}'}{\hat{c}_{44}}\times\left(1+\frac{2}{\Xi}-2\frac{\hat{c}_{44}}{c_{22}'}\sigma_z\right)xz^2\right\}, \tag{44}$$

where the function $\mathrm{f_z(y)}$ is a polynomial whose exact form is dependent upon the end conditions of the beam and which has a form of $\alpha\,\mathrm{y}+\beta$ (α,β: constant) to satisfy $\partial S_4/\partial y=0$. The derivative of displacement u_y makes it possible to calculate V(y) and Φ(y) from the relation of eqs.(5) and (6). From the second relation of eq.(7), the residual displacement v_y is transformed into

$$\begin{aligned}v_y&=u_y-\mathrm{V(y)}-x\Phi(\mathrm{y})\\&=x\,\mathrm{f_z(y)}-\frac{Q}{c_{22}'I_z}\left\{\Xi'''\chi'+\frac{1}{4}\frac{c_{22}'}{\hat{c}_{44}}\times\left(1+\frac{2}{\Xi}-2\frac{\hat{c}_{44}}{c_{22}'}\sigma_z\right)xz^2\right\}\\&\quad-\frac{1}{A}\int_{-z_0}^{z_0}\int_{-x_0}^{x_0}u_y\,\mathrm{d}x\,\mathrm{d}z\\&\quad-\frac{x}{I_z}\int_{-z_0}^{z_0}\int_{-x_0}^{x_0}xu_y\,\mathrm{d}x\,\mathrm{d}z.\end{aligned} \tag{45}$$

The differential form, $\partial v_y/\partial x$ of the residual dis-

placement v_y then yields, in view of the displacement u_y of eq.(44),

$$\frac{\partial v_y}{\partial x}=-\frac{Q}{c'_{22}I_z}\left[\Xi'''\frac{\partial\chi'}{\partial x}+\frac{1}{4}\frac{c'_{22}}{\hat{c}_{44}}\times\left(1+\frac{2}{\Xi}-2\frac{\hat{c}_{44}}{c'_{22}}\sigma_z\right)z^2-\frac{1}{I_z}\int_{-z_0}^{z_0}\int_{-x_0}^{x_0}x\left\{\Xi'''\chi'+\frac{1}{4}\frac{c'_{22}}{\hat{c}_{44}}\times\left(1+\frac{2}{\Xi}-2\frac{\hat{c}_{44}}{c'_{22}}\sigma_z\right)xz^2\right\}\mathrm{d}x\,\mathrm{d}z\right]. \quad (46)$$

In eq.(41) we neglect the torsional term, namely

$$\tau\left(-z+\Xi'''\frac{\partial\phi'}{\partial x}\right)=0. \quad (47)$$

Substitution of the T_6 and $\partial v_y/\partial x$ of eq.(46) into the denominator of the right side in eq.(13) gives

$$\int_{-z_0}^{z_0}\int_{-x_0}^{x_0}\left(T_6-c'_{66}\frac{\partial v_y}{\partial x}\right)\mathrm{d}x\,\mathrm{d}z=\frac{Q}{K}=\frac{c'_{66}}{c'_{22}}\frac{Q}{I_z}\left[\sigma_z I_x-\frac{1}{4}\left\{\frac{c'_{22}}{\hat{c}_{44}}\left(1+\frac{2}{\Xi}\right)I_x+\frac{c'_{22}}{c'_{66}}\left(1-\frac{2}{\Xi}\right)I_z\right\}-\frac{A}{I_z}\int_{-z_0}^{z_0}\int_{-x_0}^{x_0}\Xi'''x\chi'\,\mathrm{d}x\,\mathrm{d}z\right], \quad (48)$$

where, $I_x=\int_{-z_0}^{z_0}\int_{-x_0}^{x_0}z^2\,\mathrm{d}x\,\mathrm{d}z$.

Thus, we find that the denominater of the right side in eq.(13) is given as a function of the harmonic function χ'.

4. Derivative of Harmonic Function χ'

The equations of the boundaries are two of $x'=\pm x'_0$ $(x'_0=x_0/\Xi'')$ and $z'=\pm z'_0(z'_0=z_0/\Xi')$. From eq.(35) the boundary condition at $x'=\pm x'_0$ is

$$\frac{\partial\chi'}{\partial x'}=\frac{1}{2}\left(1-\frac{1}{2}\Xi\right)x'^2_0-\frac{1}{2}\left(1+\frac{1}{2}\Xi-2\Xi'\sigma_z\right)z'^2 \quad (49)$$

$$\left(-z'_0\langle z'\langle z'_0\right),$$

while the boundary condition at $z'=\pm z'_0$ is

$$\frac{\partial\chi'}{\partial z'}=\left\{\mp\left(1+\frac{1}{2}\Xi\right)z'_0-\frac{c'_{42}}{\hat{c}_{44}}(2y_0-y)\right\}x' \quad (50)$$

$$\left(-x'_0\langle x'\langle x'_0\right).$$

We introduce a new function $\bar{\chi}$ by the equation

$$\bar{\chi}=\chi'-\frac{1}{6}\left[\left(1+\frac{1}{2}\Xi\right)x'^3-3\left\{\left(1+\frac{1}{2}\Xi\right)x'z'^2+\frac{2c'_{42}}{\hat{c}_{44}}(2y_0-y)x'z'\right\}\right]. \quad (51)$$

Then $\bar{\chi}$ is a plane harmonic function within the rectangle, $\partial\bar{\chi}/\partial z'$ vanishes at $z'=\pm z'_0$, and the condition at $x'=\pm x'_0$ becomes

$$\frac{\partial\bar{\chi}}{\partial x'}=-\frac{1}{2}\Xi x'^2_0+\Xi'\sigma_z z'^2+\frac{c'_{42}}{\hat{c}_{44}}(2y_0-y)z'. \quad (52)$$

Now when $-z'_0\langle z'\langle z'_0$, the function z'^2 can be expanded in a Fourier's series[5] as follows:

$$z'^2=\frac{z'^2_0}{3}+\frac{4z'^2_0}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\cos\frac{n\pi z'}{z'_0}. \quad (53)$$

Hence, $\bar{\chi}$ can be expressed in the form

$$\bar{\chi}=\left(-\frac{1}{2}\Xi x'^2_0+\frac{1}{3}\Xi'\sigma_z z'^2_0\right)x'+\frac{c'_{42}}{\hat{c}_{44}}(2y_0-y)x'z'+\Xi'\sigma_z\frac{4z'^3_0}{\pi^3}\times\sum_{n=1}^{\infty}\frac{(-1)^n}{n^3}\frac{\sinh\frac{n\pi x'}{z'_0}}{\cosh\frac{n\pi x'_0}{z'_0}}\cos\frac{n\pi z'}{z'_0}, \quad (54)$$

and, by means of this relation and eq.(51), χ' can be written down, so that the harmonic function χ' which includes the torsional term becomes

$$\chi'=\left(-\frac{1}{2}\Xi x'^2_0+\frac{1}{3}\Xi'\sigma_z z'^2_0\right)x'+\frac{1}{6}\left(1+\frac{1}{2}\Xi\right)x'^3-\frac{1}{2}\left(1+\frac{1}{2}\Xi\right)x'z'^2+\Xi'\sigma_z\frac{4z'^3_0}{\pi^3}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^3}\frac{\sinh\frac{n\pi x'}{z'_0}}{\cosh\frac{n\pi x'_0}{z'_0}}\cos\frac{n\pi z'}{z'_0}. \quad (55)$$

In addition, when the axes of new coordinates is transformed into the axes of old coordinates, χ' can be rewritten as

$$\chi' = \frac{1}{\Xi'''}\left\{\left(-\frac{1}{2}\frac{c'_{22}}{c'_{66}}x_0^2 + \frac{1}{3}\sigma_z z_0^2\right)x + \frac{1}{6}\left(\frac{1}{2}\frac{c'_{22}}{c'_{66}} + \frac{\hat{c}_{44}}{c'_{66}}\frac{1}{\Xi'}\right)x^3 - \frac{1}{2}\left(\frac{1}{2}\frac{c'_{22}}{\hat{c}_{44}} + \frac{1}{\Xi'}\right)xz^2\right\} + \frac{\sigma_z}{\Xi'^2}\frac{4z_0^3}{\pi^3} \times \sum_{n=1}^{\infty}\frac{(-1)^n}{n^3}\frac{\sinh\left(n\pi\sqrt{\frac{\hat{c}_{44}}{c'_{66}}}\frac{x}{z_0}\right)}{\cosh\left(n\pi\sqrt{\frac{\hat{c}_{44}}{c'_{66}}}\frac{x_0}{z_0}\right)}\cos\frac{n\pi z}{z_0} \quad (56)$$

The derivative of the harmonic function χ' makes it possible to calculate the shear coefficient K .

5. Calculation of Shear Coefficient K and Its Calculated Values

From the relation of eq.(48) the shear coefficient K is given as

$$K = \frac{c'_{22}I_z}{c'_{66}\left[\sigma_z I_x - \frac{1}{4}\left\{\frac{c'_{22}}{\hat{c}_{44}}\left(1+\frac{2}{\Xi}\right)I_x + \frac{c'_{22}}{c'_{66}}\left(1-\frac{2}{\Xi}\right)I_z\right\} - \frac{A}{I_z}\int_{-z_0}^{z_0}\int_{-x_0}^{x_0}\Xi'''x\chi'\,dx\,dz\right]} \quad (57)$$

As a result of calculation, we find that the shear coefficient K for quartz crystal of rectangular cross section which is derived from Timoshenko's beam theory is

$$K = \frac{10c'_{22}}{11c'_{22} + \sqrt{2\hat{c}_{44}(\hat{c}_{44} + c'_{66})}}. \quad (58)$$

As is apparent from eq.(58), K of the rectangular cross section is independent of a thickness-to-width ratio. In case of an isotropic material, when Poisson's ratio is taken as σ, the constants $c'_{22}, \hat{c}_{44}$ and c'_{66} have the relation

$$\hat{c}_{44} = c'_{66} = \frac{c'_{22}}{2(1+\sigma)}. \quad (59)$$

Substituting this into eq.(58), the shear coefficient K for the isotropic material becomes

$$K = \frac{10(1+\sigma)}{12+11\sigma}. \quad (60)$$

This result agrees completly with that obtained by G.R.Cowper[2]. The following shows an example of the calculated results. In the calculation the constants of reference(6) are used.

Table.1 Comparison of our K values and those[2] obtained by other authors.

	K
Timoshenko	0.667
Mindlin	0.822
Goodman (σ =1/3)	0.870
Roark	0.833
Cowper (σ =0)	0.833
(σ =0.3)	0.850
(σ =1/2)	0.870
The author	0.796~0.867

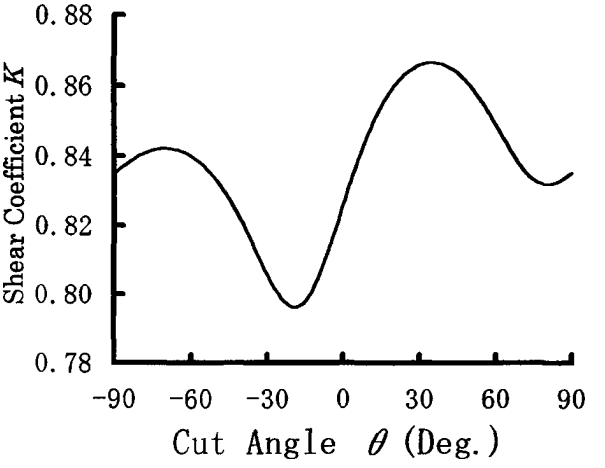


Fig.2 Relationship of the shear coefficient K and a cut angle θ rotated about the x-axis.

Figure 2 shows the relationship of the shear coefficient K and a cut angle θ rotated about the x-axis. K varies with the cut angle θ and has a value of 0.796 to 0.867 when the cut angle θ varies from -90° to +90°.

Figure 3 shows the first-and second-order temperature coefficients K_α, K_β of K versus a cut angle θ. The K_α has a value of -15.3 to $+13.2\times10^{-6}$/℃ and K_β has a value -1.72 to $+1.14\times10^{-8}$/℃^{2} when the cut angle θ varies from -90° to +90°.

Figure 4 shows the third-order temperature coefficient K_γ of K versus θ. The K_γ has a value of -3.28 to $+2.71\times10^{-11}$/℃^{3} versus a variation of the above-mentioned cut angle.

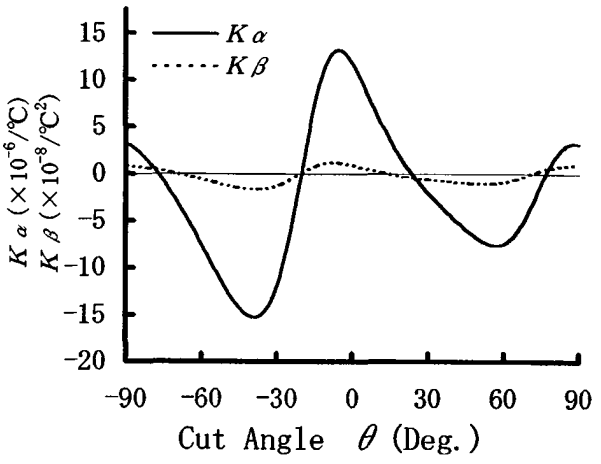


Fig.3 The first- and second-order temperature coefficients K_α, K_β of K versus a cut angle θ.

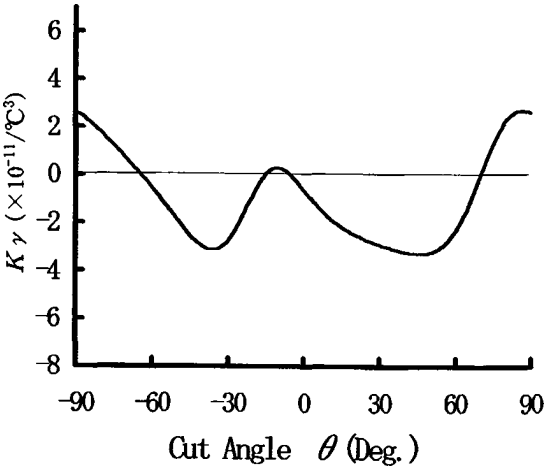


Fig.4 The third-order temperature coefficient K_γ of K versus a cut angle θ.

Table1 compares our values of K with those for the isotropic material obtained by other authors. With the exception of Timoshenko's value, the foregoing results almost include those obtained by other authors and agree fairly well with each other.

6. Conclusions

In this paper, the shear coefficient K for quartz crystal of the rectanglar cross section that is an anisotropic material has been derived from Timoshenko's beam theory. It was shown that the shear coefficient K is given as a function of the transformed elastic stiffness constants $c'_{22}, \hat{c}_{44}, c'_{66}$ only and is independent of a thickness-to-width ratio of the cross section. In addition, K and its temperature coefficients versus a cut angle θ was calculated, so that they were found to be very dependent upon the cut angle θ.

For the coming subject, the influence by K will be clarified on resonance frequency and frequency temperature behavior for flexural mode quartz crystal resonators.

Acknowledgement

The author expresses his thanks to Mr.M.Kato for his help in the numerical calculation by computor.

References

(1) S.P.Timoshenko:"Vibration Problems in Engineering" D.Van Nostrand Co.Inc., New York(1937).
(2) G.R.Cowper:"The Shear Coefficient in Timoshenko's Beam Theory", Journal of Applied Mechanics, Vol.33,pp335-340, 1966.
(3) R.D.Mindlin and H.Deresiewicz:"Timoshenko's Shear Coefficient for Flexural Vibrations of Beams", Technical Report No.10, ONR Project NR 064-388, Department of Civil Engineering, Columbia University, New York, N.Y., 1953.
(4) L.E.Goodman, discussion of paper: "Flexural Vibrations in Uniform Beams According to the Timoshenko Theory", by R.A.Anderson, Journal of Applied Mechanics, Vol.21, pp.202-204, 1954.
(5) A.E.H.Love:"A treatise on the Mathematical Theory of Elasticity", fourth edition, Cambridge University Press,Cambridge, England,1952.
(6) M.Aruga:"On the elastic constants of quartz and their temperature characteristics", Bull. Tokyo Inst. Tech., Vol.A-2, pp.88-182, (1956) (in Japanese).

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

MATHEMATICAL THEORY OF THE FORK-TYPE WAVE GYROSCOPE

Igor A. Ulitko

Department of Theoretical and Applied Mechanics, Kiev University, 64 Vladimirskaya Street, Kiev 252017, Ukraine

Abstract — The work is devoted to mathematical analysis of the fork-type wave gyroscope. It includes the theory of flexural and torsional vibrations of thin bimorph beams and an applied theory of elastic half-ring vibrations both in the plane of fork arrangement and in perpendicular direction. Solution of the problem is expressed in analytic form in terms of hyperbolic and trigonometric functions. Both free and forced vibrations are considered. Analysis of the sensitivity of the gyroscope is carried out.

Introduction

Mathematical analysis of stationary vibrations and wave propagation in rotating elastic bodies is similar to that for motionless bodies [1]. However, there are fundamental differences both in qualitative and in quantitative description of dynamical phenomena in moving elastic bodies. For example, in a uniformly rotating elastic body, travelling plane harmonic waves are dispersive, and unlike classical case, trajectories of particles motion are not straight lines but ellipses or circles instead [2]. Also, it appears that resonance states in a rotating body of finite dimensions are not standing modes. They are travelling in opposite directions waves with different but close to each other phase velocities. When amplitudes of these travelling waves are equal, their superposition is a slowly wobbling wave, discovered by Bryan [3] in studies of free vibrations of a rotating ring.

Under harmonic loading in resonance area, rotating along with a body, we obtain modulated vibrations as result of addition of resonance solutions. So, there is a considerable difference between free and forced vibrations. Therefore, in construction of mathematical models of wave gyroscope it is necessary to take into account the way of outer energy supply used for excitation of vibrations.

In modern constructions of wave gyroscopes, the elecromechanically active materials are used. These are piezoceramics or piezoelectric crystals of specified cuts. Electric field rotates along with the resonator with the same angular velocity. Measuring potential difference on additional electrodes, one obtains information about rotational motion of a resonator.

Main objective of the mathematical theory of wave gyroscopes is to provide an exact estimation of rotational motion of the resonator through output potential difference at the detecting part under strict control of input energy at the exciting (driving) part. Since we deal with resonance vibrations, it is important to take into account both the real mechanical quality factor and temperature stability of resonant elements. The latter depends primarily on the level of mechanical stresses, what causes restrictions on the level of admissible stresses in a resonator.

At last, note that mechanical and electrical wave fields in high-effective piezoelectric crystals and ceramics are coupled. Therefore, calculations of amplitudes of vibrations for driving and detecting parts of gyroscope under specified values of mechanical quality factor and driving voltage as well as another important parameters could be accomplished only from the equations of field theory of electroelasticity [4].

As an example of application of general theory, we consider one of constructions of the wave gyroscope, namely the fork-type wave gyroscope.

Mathematical Model of the Fork-Type Wave Gyroscope

Experimental investigations of main characteristics of miniaturized fork-type wave gyroscope have been described in literature [5, 6]. However, to the best of the author's knowledge, methods of mathematical estimations of main parameters of this devise were not carried out. In our work, the fork is considered as a system consisting of two bimorph piezoceramic beams of square cross-section connected by a non-active elastic half-ring.

Mechanical characteristics of both bimorph beams and the half-ring are identical. The driving arm (on the right in Fig. 1) has remanent thickness polarization in ox-direction. Thin electrodes cover it at planes perpen-

0-7803-2500-1/95 $4.00 © 1995 IEEE

dicular to ox-axis. The detecting arm (on the left in Fig. 1), has remanent polarization in oy-direction, and

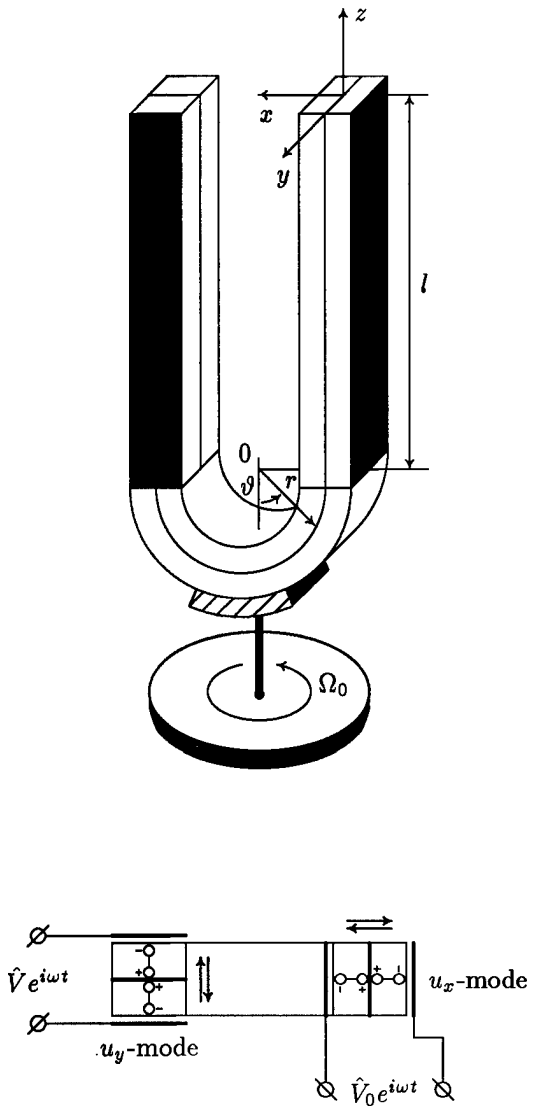


Fig. 1

electroded coverings at planes perpendicular to this axis. The elastic half-ring is connected with the testing object with a thin lap and a rigid rod. The cross-section of the half-ring is a square of side h as well that of beams. The length of each arm is l and radius of the half-ring reference line is r. In the construction of the problem solution for this compound vibratory system we use both the solutions for flexural vibrations of bimorph beams in mutually perpendicular directions [4] and similar solutions for the half-ring. Notice, that the half-ring will undergo a torsional deformation in addition to flexure. It implies a correction of flexural vibrations of arms by some additional equations that will appear owing to their torsional vibrations.

Flexural Vibrations of Bimorph Beams. Governing Equations

Let $u_x(z,t)$ is the flexural displacement of the piezoceramic beam with positive values in ox-direction, and let $u_y(z,t)$ if the same for the perpendicular direction. Using the theory of flexural vibrations of thin piezoceramic beams developed in [4], we can write governing equations in the form

$$\frac{\partial^4}{\partial z^4}\begin{Bmatrix} u_x \\ u_y \end{Bmatrix} + \frac{\rho h^2}{\overline{D}(1-\overline{\nu}^2)}\frac{\partial^2}{\partial t^2}\begin{Bmatrix} u_x \\ u_y \end{Bmatrix} = \frac{q}{\overline{D}(1-\overline{\nu}^2)}, \tag{1}$$

where ρ is the volume density, $\overline{D}$ and $\overline{\nu}$ are the effective cylindrical stiffness and Poisson ratio, respectively. They have values

$$\overline{D} = \frac{h^4}{12 s_{11}^E (1-\nu^2)}\left[1 + \frac{1+\nu}{8}\frac{k_p^2}{1-k_p^2}\right],$$
$$\overline{\nu} = \left(\nu + \frac{1+\nu}{8}\frac{k_p^2}{1-k_p^2}\right)\Big/\left(1 + \frac{1+\nu}{8}\frac{k_p^2}{1-k_p^2}\right). \tag{2}$$

In these formulae, s_{11}^E is the elastic compliance at zero electric field, k_p^2 is the planar electromechanical coupling factor [7, 4]. We neglect the small alteration of cylindrical stiffness $\overline{D}$ at mutually ortogonal displacements $u_x(z,t)$ and $u_y(z,t)$. Therefore, both equations are identical.

But calculating the bending momentum and the shear force, we must use different expressions. So, under the flexure in direction of polarization for driving arm, we have

$$M_x = -\overline{D}(1-\overline{\nu}^2)\frac{\partial^2 u_x}{\partial z^2} + \frac{h^4}{4}\frac{d_{31}}{s_{11}^E}\frac{1-\overline{\nu}}{1-\nu}\frac{V_0(t)}{h},$$
$$Q_x = \frac{\partial M_x}{\partial z} = -\overline{D}(1-\overline{\nu}^2)\frac{\partial^3 u_x}{\partial z^3}, \tag{3}$$

and under the flexure in the perpendicular direction, correspondingly

$$M_y = -\overline{D}(1-\overline{\nu}^2)\frac{\partial^2 u_y}{\partial z^2},$$
$$Q_y = -\overline{D}(1-\overline{\nu}^2)\frac{\partial^3 u_y}{\partial z^3}, \tag{4}$$

where $V_0(t)/h$ is the value of external electrical field.

As for detecting arm, it is clear that in written formulae u_x and u_y will interchange their places. Under the condition of absence of the conductivity current in the external measuring circuit through the open-circuited electrodes of detecting arm we have

$$\frac{V(t)}{h} = -\frac{h}{8}\frac{1-\overline{\nu}}{d_{31}}\frac{k_p^2}{1-k_p^2}\frac{1}{l}\frac{\partial u_y}{\partial z}\Bigg|_0^l \qquad (5)$$

As was noted above, equations for flexural vibrations of the arms must be completed by the equation of torsional vibrations. Let φ is the angle of the beam torsion and $\tau = \partial\varphi/\partial z$ is the same referred to a unit length. Thus we obtain

$$\frac{\partial^2\varphi}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2}, \qquad c = \frac{1}{\sqrt{\rho s_{11}^E}}, \qquad (6)$$

where c is the longitudinal waves velocity in the beam. The torsional couple takes the value

$$M_\varphi = C\frac{\partial\varphi}{\partial z} = C\tau, \qquad (7)$$

where C is the torsional stiffness of the beam.

Flexural Vibrations of Half-Ring

Small flexural deformations of a half-ring in its plane satisfy the equation [8, 9]

$$\frac{\partial^4 u}{\partial\vartheta^4} + 2\frac{\partial^2 u}{\partial\vartheta^2} + u + \frac{\rho h^2 r^4}{EJ}\frac{\partial^2 u}{\partial t^2} = 0. \qquad (8)$$

The bending couple and the shear force are represented by

$$M_\vartheta = -\frac{EJ}{r^2}\left(\frac{\partial^2 u}{\partial\vartheta^2} + u\right),$$
$$Q_\vartheta = \frac{1}{r}\frac{\partial M_\vartheta}{\partial\vartheta} = -\frac{EJ}{r^3}\frac{\partial}{\partial\vartheta}\left(\frac{\partial^2 u}{\partial\vartheta^2} + u\right), \qquad (9)$$

where $J = h^4/12$ is the moment of inertia of square cross-section, and E is the Young modulus. In these formulae $u(\vartheta, t)$ is the radial displacement of half-ring in its plane. Positive values of $u(\vartheta, t)$ are directed to the center. Pozitive direction of angular coordinate ϑ is indicated in Fig. 1.

Governing equations for flexural vibrations of the half-ring in the direction perpendicular to its plane are more complicated. Nevertheless, the kinematics of vibrations can be described using a single component of displacements $v(\vartheta, t)$, which takes pozitive values in oy–direction. After some transformations of known relations for flexural deformations in the direction perpendicular to the plane of half-ring under the torsion, we obtain the equation for $v(\vartheta, t)$

$$\frac{\partial^6 v}{\partial\vartheta^6} + 2\frac{\partial^4 v}{\partial\vartheta^4} + \frac{\partial^2 v}{\partial\vartheta^2} = \rho h^2 r^4 \frac{\partial^2}{\partial t^2}\left(\frac{1}{EJ}\frac{\partial^2 v}{\partial t^2} - \frac{v}{C}\right) = 0. \qquad (10)$$

Here C is the stiffness of the half-ring under torsion, $C = kEJ$, where $k \sim 0.65$ for $\nu = 0.33$. This value of C coincides with that for bimorph beams. The bending couple $M_1(\vartheta, t)$, the shear force $N_2(\vartheta, t)$, the torsional angle $\beta(\vartheta, t)$ and the torsional couple $H(\vartheta, t)$ are expressed through $v(\vartheta, t)$ in such a way

$$M_1 = \frac{k}{1+k}\frac{EJ}{r^2}\left(\frac{\partial^4 v}{\partial\vartheta^4} + \frac{\partial^2 v}{\partial\vartheta^2} + \frac{\rho h^2 r^4}{EJ}\frac{\partial^2 v}{\partial t^2}\right),$$
$$N_2 = k\frac{EJ}{r^2}\frac{\partial}{\partial\vartheta}\left(\frac{\partial^4 v}{\partial\vartheta^4} + 2\frac{\partial^2 v}{\partial\vartheta^2} + v + \frac{\rho h^2 r^4}{EJ}\frac{\partial^2 v}{\partial t^2}\right),$$
$$\beta = \frac{k}{1+k}\frac{1}{r}\left[\frac{\partial^4 v}{\partial\vartheta^4} + \left(2+\frac{1}{k}\right)\frac{\partial^2 v}{\partial\vartheta^2} + \frac{\rho h^2 r^4}{EJ}\frac{\partial^2 v}{\partial t^2}\right],$$
$$H = \frac{k^2}{1+k}\frac{EJ}{r^2}\frac{\partial}{\partial\vartheta}\left[\frac{\partial^4 v}{\partial\vartheta^4} + \left(2+\frac{1}{k}\right)\frac{\partial^2 v}{\partial\vartheta^2} + \left(1+\frac{1}{k}\right)v + \frac{\rho h^2 r^4}{EJ}\frac{\partial^2 v}{\partial t^2}\right], \qquad (11)$$

In the case of a motionless fork, displacements u_x, u_y, u, v, and angles of torsion φ, β can be obtained form corresponding equations separately. If the fork rotates, inertia forces changes. In the case of a uniform rotational motion, a complete expression of inertia forces must include terms originated by relative, translational and Coriolis acceleration components. This circumstance leads to coupled equations of motion in terms of displacement functions of both beams and half-ring.

Magnitudes of coupled cross-terms depend on angular velocity, frequency of vibrations and amplitudes of displacements of different fork parts. But we managed to simplify these complicated equations in the case when the radius of the half-ring is small compared to the length of fork arms ($r/l \ll 1$). Then, we can neglect coupling of equations for u and v. Moreover, in equations for their separate determination one can neglect inertia forces at all. Meanwhile, an analysis of arms vibrations should be performed when inertia force are taken into account.

Equations of Vibrations and Their Solution

Coupled equations for vibrations of fork arms under a uniform rotational motion with angular velocity Ω_0 are

$$\frac{\partial^4 u_x}{\partial z^4} + \frac{\rho h^2}{EJ}\left[\frac{\partial^2 u_x}{\partial t^2} - \Omega_0^2 u_x + 2\Omega_0\frac{\partial u_y}{\partial t}\right] = 0,$$
$$\frac{\partial^4 u_y}{\partial z^4} + \frac{\rho h^2}{EJ}\left[\frac{\partial^2 u_y}{\partial t^2} - \Omega_0^2 u_y - 2\Omega_0\frac{\partial u_x}{\partial t}\right] = 0. \qquad (12)$$

When these equations are compared with those for flexural vibrations of motionless beams (1), it is apparent that

$$\overline{D}(1-\overline{\nu}^2) = EJ\,. \tag{13}$$

The independent equation for torsional vibrations (6) takes the form of purely static equation bacause the first resonance frequency of vibrations of this kind is far exceed the lowest resonance frequency of flexural vibrations.

$$\frac{\partial^2\varphi}{\partial z^2} = 0\,. \tag{14}$$

We seek the solution for harmonic vibrations of arms at frequency ω in the form

$$u_x(z,t) = \hat{u}_x(z)\exp(i\omega t)\,, \quad u_y(z,t) = \hat{u}_y(z)\exp(i\omega t)\,,$$
$$\varphi(z,t) = \hat{\varphi}(z)\exp(i\omega t)\,. \tag{15}$$

Substituting (15) into Eqs. (14) and (13), we obtaine a system of ordinary differential equations for amplitude functions $\hat{u}_x$, $\hat{u}_y$ and $\hat{\varphi}$. The general solution of this a system is

$$\hat{u}_x(z) = \alpha^+(z) + \alpha^-(z)\,, \quad \hat{u}_y(z) = i\alpha^+(z) - \alpha^-(z)\,,$$
$$\hat{\varphi}(z) = \gamma + \delta z, \tag{16}$$

where

$$\alpha^+(z) = a^+\sin(\sqrt{1+\varepsilon}\,\kappa z) + b^+\cos(\sqrt{1+\varepsilon}\,\kappa z)$$
$$+ c^+\,\mathrm{sh}\,(\sqrt{1+\varepsilon}\,\kappa z) + d^+\,\mathrm{ch}\,(\sqrt{1+\varepsilon}\,\kappa z)\,,$$
$$\alpha^-(z) = a^-\sin(\sqrt{1-\varepsilon}\,\kappa z) + b^-\cos(\sqrt{1-\varepsilon}\,\kappa z)$$
$$+ c^-\,\mathrm{sh}\,(\sqrt{1-\varepsilon}\,\kappa z) + d^-\,\mathrm{ch}\,(\sqrt{1-\varepsilon}\,\kappa z)\,. \tag{17}$$

Here

$$\varepsilon = \frac{\Omega_0}{\omega}\,, \qquad \kappa = \frac{12}{h^2}\frac{\omega^2}{c_0^2} \tag{18}$$

are nondimensional parameters. Arbitrary constants $a^\pm$, $b^\pm$, $c^\pm$, $d^\pm$, γ and δ will be determined from conditions of conjunction between bimorph beams and a half-ring at $z = 0$ and from boundary conditions at free ends of beams at $z = l$. A general solution for the amplitude function under in-plane half-ring flexural vibrations is represented by

$$\hat{u}(\vartheta) = \mathcal{A}\sin\vartheta + \mathcal{B}\cos\vartheta + \mathcal{C}\vartheta\sin\vartheta + \mathcal{D}\vartheta\cos\vartheta\,. \tag{19}$$

Here we assume that vibrations have the same circular frequency ω. A similar general solution for the amplitude function under flexure of half-ring in perpendicular direction is written as

$$\hat{v}(\vartheta) = \mathcal{A}_1 + \mathcal{B}_1\vartheta + \mathcal{C}_1\sin\vartheta + \mathcal{D}_1\cos\vartheta$$
$$+ \mathcal{E}_1\vartheta\sin\vartheta + \mathcal{F}_1\vartheta\cos\vartheta\,. \tag{20}$$

At the first symmetrical mode of vibrations of the fork attached to a rigid holder in the central cross-section of the half-ring at $\vartheta = 0$, the following conditions must be fulfilled

$$\hat{u}(\vartheta)|_{\vartheta=0} = \hat{v}(\vartheta)|_{\vartheta=0} = 0\,,$$
$$\left.\frac{\partial\hat{u}}{\partial\vartheta}\right|_{\vartheta=0} = \left.\frac{\partial\hat{v}}{\partial\vartheta}\right|_{\vartheta=0} = 0\,, \qquad \hat{\beta}\Big|_{\vartheta=0} = 0\,. \tag{21}$$

Final expressions for flexure amplitude functions, shear forces, torsional amd bending couples, and angles of torsion are derived from conditions (21).

1. In-plane flexure:

$$\hat{u}(\vartheta) = \mathcal{C}\vartheta\sin\vartheta - \mathcal{D}(\sin\vartheta - \vartheta\cos\vartheta)\,,$$
$$\frac{\partial\hat{u}}{\partial\vartheta} = \mathcal{C}(\sin\vartheta + \vartheta\cos\vartheta) - \mathcal{D}\vartheta\sin\vartheta\,,$$
$$\hat{M}_\vartheta = -EJ\frac{2}{r^2}(\mathcal{C}\cos\vartheta - \mathcal{D}\sin\vartheta)\,,$$
$$\hat{Q}_\vartheta = EJ\frac{2}{r^3}(\mathcal{C}\sin\vartheta + \mathcal{D}\cos\vartheta)\,. \tag{22}$$

2. Flexure in the perpedicular direction

$$\hat{v}(\vartheta) = \mathcal{B}(\vartheta - \sin\vartheta) - \mathcal{E}\left[\frac{2}{1+k}(1-\cos\vartheta) - \vartheta\sin\vartheta\right]$$
$$- \mathcal{F}(\sin\vartheta - \vartheta\cos\vartheta)\,,$$
$$\frac{\partial\hat{v}}{\partial\vartheta} = \mathcal{B}(1-\cos\vartheta) - \mathcal{E}\left(\frac{1-k}{1+k}\sin\vartheta - \vartheta\cos\vartheta\right)$$
$$- \mathcal{F}\vartheta\sin\vartheta\,,$$
$$\hat{\beta}(\vartheta) = \frac{1}{r}\left[\mathcal{B}\sin\vartheta - \mathcal{E}\vartheta\sin\vartheta\right.$$
$$\left. - \mathcal{F}\left(\frac{1-k}{1+k}\sin\vartheta - \vartheta\cos\vartheta\right)\right]$$
$$\hat{M}_1(\vartheta) = -EJ\frac{k}{1+k}\frac{2}{r^2}(\mathcal{E}\cos\vartheta - \mathcal{F}\sin\vartheta)\,,$$
$$\hat{N}_2(\vartheta) = kEJ\frac{1}{r^3}\mathcal{B}\,,$$
$$\hat{H}(\vartheta) = EJ\frac{k^2}{1+k^2}\frac{1}{r^2}$$
$$\times\left[\mathcal{B} - \frac{2}{1+k}(\mathcal{E}\sin\vartheta + \mathcal{F}\cos\vartheta)\right]\,. \tag{23}$$

The five arbitrary constants in these equations are eliminated using conditions of conjunction between the half-ring and fork arms

$$\hat{u}|_{\vartheta=\frac{\pi}{2}} = \hat{u}_x|_{z=0}\,, \quad \left.\frac{1}{r}\frac{\partial\hat{u}}{\partial\vartheta}\right|_{\vartheta=\frac{\pi}{2}} = \left.\frac{\partial\hat{u}_x}{\partial z}\right|_{z=0}\,,$$
$$\hat{M}_\vartheta\Big|_{\vartheta=\frac{\pi}{2}} = \hat{M}_x\Big|_{z=0}\,, \quad \hat{Q}_\vartheta\Big|_{\vartheta=\frac{\pi}{2}} = \hat{Q}_x\Big|_{z=0}\,,$$
$$\hat{v}|_{\vartheta=\frac{\pi}{2}} = \hat{u}_y|_{z=0}\,, \quad \left.\frac{1}{r}\frac{\partial\hat{v}}{\partial\vartheta}\right|_{\vartheta=\frac{\pi}{2}} = \left.\frac{\partial\hat{u}_y}{\partial z}\right|_{z=0}\,,$$

$$\hat{\beta}\Big|_{\vartheta=\frac{\pi}{2}} = \hat{\varphi}|_{z=0}\,, \qquad \hat{M}_1\Big|_{\vartheta=\frac{\pi}{2}} = \hat{M}_y\Big|_{z=0}\,,$$
$$\hat{N}_2\Big|_{\vartheta=\frac{\pi}{2}} = \hat{Q}_y\Big|_{z=0}\,, \qquad \hat{H}\Big|_{\vartheta=\frac{\pi}{2}} = \hat{M}_\varphi\Big|_{z=0}\,, \tag{24}$$

These conditions must be reformulated in order to obtain relations for functions describing vibrations of bimorph beams. The solutions of half-ring vibrations are omitted from this point onward. Therefore, it is necessarily to found arbitrary constants that enter the solution for half-ring, assuming that terms on the right-hand side are known. As a result, we obtain conditions between unknown functions for bimorph beams

$$\begin{aligned}
\hat{u}_x|_{z=0} &= \frac{\pi}{2}\frac{r^3}{2EJ}\hat{Q}_x\Big|_{z=0} - \frac{r^2}{2EJ}\hat{M}_x\Big|_{z=0}\,,\\
r\frac{\partial\hat{u}_x}{\partial z}\Big|_{z=0} &= \frac{r^3}{2EJ}\hat{Q}_x\Big|_{z=0} - \frac{\pi}{2}\frac{r^2}{2EJ}\hat{M}_x\Big|_{z=0}\,,\\
\hat{u}_y|_{z=0} &= \frac{1}{k}\left[\frac{\pi}{2}\left(1+\frac{1+k}{2}\right)-2\right]\frac{r^3}{EJ}\hat{Q}_y\Big|_{z=0}\\
&\quad - \frac{1+k}{k}\frac{r^2}{2EJ}\hat{M}_y\Big|_{z=0}\,,\\
r\frac{\partial\hat{u}_y}{\partial z}\Big|_{z=0} &= \frac{1+k}{2k}\frac{r^3}{EJ}\hat{Q}_y\Big|_{z=0} -\\
&\quad - \frac{\pi}{2}\frac{1+k}{k}\frac{r^2}{2EJ}\hat{M}_y\Big|_{z=0}\,,
\end{aligned} \tag{25}$$

These boundary equtions can be treated as conditions of elastic conjunction between fork arms and the half-ring. The amplitude of torsion angle of the arms is calculated by the formula ($M_\varphi = 0$)

$$\hat{\varphi}(z) = \gamma = \frac{1}{r}\left\{\left[1-\frac{\pi}{4}(1+k)\right]\frac{1}{k}\frac{r^3}{EJ}\hat{Q}_y\Big|_{z=0} - \frac{1-k}{k}\frac{r^2}{2EJ}\hat{M}_y\Big|_{z=0}\right\}, \tag{26}$$

If we complement conditions of conjunction (24) by boundary conditions at the top end ($z = l$) of the arm,

$$\hat{M}_x\Big|_{z=l} = \frac{1}{2}M_0\,,$$
$$\hat{M}_y\Big|_{z=l} = \hat{Q}_x\Big|_{z=l} = \hat{Q}_y\Big|_{z=l} = \hat{M}_\varphi\Big|_{z=l} = 0\,, \tag{27}$$

we obtain eight equations for eight constants included in solutions (16) and (17) for flexural vibrations of bimorph beams. Here

$$M_0 = 3EJd_{31}\frac{1-\overline{\nu}}{1-\nu}\frac{V_0}{h} \tag{28}$$

We diminish the couple M_0 twice at driving arm on the ground of following. In excitation of vibrations at symmetrical mode by means of one-sided boundary conditions a half-value of exciting loading is formally translated at the detecting arm.

In general, solution of the above algebraic system of order eight has rather complicated form. But under assumption $r \ll l$, the constants could be found with a high accuracy in the form

$$\begin{aligned}
d^+ &= -b^+\,, \qquad d^- = -b^-\,,\\
c^+ &= -a^+ - \frac{\pi}{4}\left(\frac{r}{l}\right)\left[\sqrt{1+\varepsilon}\,(\kappa l)\,\frac{2k+1}{k}b^+ + \frac{1-\varepsilon}{\sqrt{1+\varepsilon}}\,(\kappa l)\,b^-\right],\\
c^- &= -a^- - \frac{\pi}{4}\left(\frac{r}{l}\right)\left[\frac{1+\varepsilon}{\sqrt{1-\varepsilon}}\,(\kappa l)\,\frac{1}{k}b^+ + \sqrt{1-\varepsilon}\,(\kappa l)\,\frac{2k+1}{k}b^-\right],\\
b^+ &= -\frac{A^+(\kappa l)}{B^+(\kappa l)}a^+ + \frac{M_0}{4EJ}\frac{\mathrm{ch}\,(\sqrt{1+\varepsilon}\,\kappa l)}{\kappa^2(1+\varepsilon)B^+(\kappa l)}\,,\\
b^- &= -\frac{A^-(\kappa l)}{B^-(\kappa l)}a^- + \frac{M_0}{4EJ}\frac{\mathrm{ch}\,(\sqrt{1-\varepsilon}\,\kappa l)}{\kappa^2(1-\varepsilon)B^-(\kappa l)}\,,
\end{aligned} \tag{29}$$

where

$$\begin{aligned}
A^+(\kappa l) &= \sin(\sqrt{1+\varepsilon}\,\kappa l)\,\mathrm{ch}\,(\sqrt{1+\varepsilon}\,\kappa l) - \cos(\sqrt{1+\varepsilon}\,\kappa l)\,\mathrm{sh}\,(\sqrt{1+\varepsilon}\,\kappa l)\,,\\
B^+(\kappa l) &= 1+\cos(\sqrt{1+\varepsilon}\,\kappa l)\,\mathrm{ch}\,(\sqrt{1+\varepsilon}\,\kappa l) + \sin(\sqrt{1+\varepsilon}\,\kappa l)\,\mathrm{sh}\,(\sqrt{1+\varepsilon}\,\kappa l)\,.
\end{aligned} \tag{30}$$

Values of $A^-(\kappa l)$ and $B^-(\kappa l)$ can be obtained from (30) simply by changing the sign before ε. Remaining unknown constants a^+ and a^- are found from the algebraic system of order two,

$$\begin{aligned}
\frac{a_{11}(\kappa l)}{B^+(\kappa l)}\,a^+ + \frac{a_{12}(\kappa l)}{B^-(\kappa l)}\,a^- &= \frac{M_0}{8EJ\kappa^2}\,C^+(\kappa l)\\
\frac{a_{21}(\kappa l)}{B^+(\kappa l)}\,a^+ + \frac{a_{22}(\kappa l)}{B^-(\kappa l)}\,a^- &= \frac{M_0}{8EJ\kappa^2}\,C^-(\kappa l)
\end{aligned} \tag{31}$$

Here is designated:

$$\begin{aligned}
a_{11}(\kappa l) &= 1+\mathrm{ch}\,(\sqrt{1+\varepsilon}\,\kappa l)\cos(\sqrt{1+\varepsilon}\,\kappa l) - \frac{\pi}{8}\frac{2k+1}{k}\left(\frac{r}{l}\right)(\kappa l)A^+(\kappa l)\\
a_{12}(\kappa l) &= \frac{\pi}{8}\frac{1-\varepsilon}{\sqrt{1+\varepsilon}}\frac{1}{k}\left(\frac{r}{l}\right)(\kappa l)A^-(\kappa l)\\
a_{21}(\kappa l) &= \frac{\pi}{8}\frac{1+\varepsilon}{\sqrt{1-\varepsilon}}\frac{1}{k}\left(\frac{r}{l}\right)(\kappa l)A^+(\kappa l)\\
a_{22}(\kappa l) &= 1+\mathrm{ch}\,(\sqrt{1-\varepsilon}\,\kappa l)\cos(\sqrt{1-\varepsilon}\,\kappa l) - \frac{\pi}{8}\frac{2k+1}{k}\left(\frac{r}{l}\right)(\kappa l)A^-(\kappa l)\\
C^+(\kappa l) &= \frac{1}{(1+\varepsilon)B^+(\kappa l)}
\end{aligned}$$

$$\times\left[\sin(\sqrt{1+\varepsilon}\,\kappa l)-\operatorname{sh}(\sqrt{1+\varepsilon}\,\kappa l)\right.$$
$$\left.-\frac{\pi}{4}\sqrt{1+\varepsilon}\frac{2k+1}{k}\left(\frac{r}{l}\right)(\kappa l)\operatorname{ch}(\sqrt{1+\varepsilon}\,\kappa l)\right]$$
$$+\frac{1}{(1-\varepsilon)B^{-}(\kappa l)}$$
$$\times\frac{\pi}{4}\frac{1-\varepsilon}{\sqrt{1+\varepsilon}}\frac{1}{k}\left(\frac{r}{l}\right)(\kappa l)\operatorname{ch}(\sqrt{1-\varepsilon}\,\kappa l)\,. \qquad (32)$$

The factor $C^{-}(\kappa l)$ on the right-hand side of the second equation of system (32) can be found when signs before ε are reversed. To obtain a resonance equation, we equate the determinant of system (32) to zero

$$\Delta^{+}(\kappa l)\,\Delta^{-}(\kappa l)-\frac{\pi}{8}\frac{2k+1}{k}\left(\frac{r}{l}\right)(\kappa l)$$
$$\times\left[A^{+}(\kappa l)\,\Delta^{-}(\kappa l)+A^{-}(\kappa l)\,\Delta^{+}(\kappa l)\right]=0\,, \qquad (33)$$

where is designated

$$\Delta^{+}(\kappa l)=1+\operatorname{ch}(\sqrt{1+\varepsilon}\,\kappa l)\cos(\sqrt{1+\varepsilon}\,\kappa l) \qquad (34)$$

The value of $\Delta^{-}(\kappa l)$ is calculated by reversing sign in $\Delta^{+}(\kappa l)$. Notice that the latter expression have the same form as the frequency equation for flexural vibrations of the beam with one fixed end. [10] It follows from numerical analysis of equation (33) under $r/l\sim 0.1$ and $\varepsilon\sim 10^{-5}$ that lowest resonance frquencies are $\kappa_0^{-}l\simeq 1.617$ and $\kappa_0^{+}l\simeq 1.751$. Meantime, in study of flexural vibrations of the beam of length l with one fixed end, a single value of resonance frequency $\kappa_0\simeq= 1.875$ was found. These figures show that the elastic half-ring causes slight reduction of resonance frequencies of both in-plane vibrations $\kappa_0^{-}l$ and those in the perpendicular direction $\kappa_0^{+}l$.

Therefore, to clear up principal features mainly concerning the sensitivity of this type of gyroscope one may to ascribe in above formulae $r/l\sim 0$. Then, closely placed resonance frequencies of beam vibrations in mutually ortogonal planes can be obtained from equations

$$\sqrt{1+\varepsilon}\,(\kappa_0^{+}l)=x_0\,,\quad \sqrt{1-\varepsilon}\,(\kappa_0^{-}l)=x_0\,,$$

where $x_0\simeq 1.875$. Assuming that excitation of vibrations is performed at these frequencies, it is necessary to take into account the magnitude of the mechanical quality factor Q. In our case of monoharmonic vibrations it can be introduced using comlex elastic modulus.

As a result, for nondimensional wavenumber κl we have

$$\kappa l=(\kappa' l)(1-\frac{i}{2\mathrm{Q}}) \qquad (35)$$

If we substitute this complex quantity into previous formulae and asymptotically expand trigonometric and hyperbolic functions with respect to the small parameter Q^{-1} with the accoracy of the first order terms, we obtain explicit expressions for arbitrary constants that enter $\alpha^{+}(z)$ and $\alpha^{+}(z)$. So, according to (15), (16) and (5) we may find both displacements of the arms and the output potential difference. For example, amplitudes of displacements at free ends of arms ($z=l$) under excitation of vibrations at resonance frequency corresponding to $(\kappa_0^{+}l)=x_0/\sqrt{1+\varepsilon}$, are calculated in the form

$$\hat{u}_x^{+}(l)=i\,\frac{M_0\mathrm{Q}l^2}{2EJ}\mathrm{N}(x_0)\left(1+\frac{1}{1-2i\varepsilon\mathrm{Q}}\right),$$
$$\hat{u}_y^{+}(l)=-\frac{M_0\mathrm{Q}l^2}{2EJ}\mathrm{N}(x_0)\left(1-\frac{1}{1-2i\varepsilon\mathrm{Q}}\right), \qquad (36)$$

Similarly, at the frequency $(\kappa_0^{-}l)=x_0/\sqrt{1-\varepsilon}$ these displasements are

$$\hat{u}_x^{-}(l)=i\,\frac{M_0\mathrm{Q}l^2}{2EJ}\mathrm{N}(x_0)\left(1+\frac{1}{1+2i\varepsilon\mathrm{Q}}\right),$$
$$\hat{u}_y^{+}(l)=-\frac{M_0\mathrm{Q}l^2}{2EJ}\mathrm{N}(x_0)\left(1-\frac{1}{1+2i\varepsilon\mathrm{Q}}\right), \qquad (37)$$

where

$$\mathrm{N}(x_0)=\frac{1}{x_0^3\operatorname{sh}x_0\sin x_0}$$
$$\times\frac{\operatorname{sh}^2x_0-\sin^2x_0}{\operatorname{ch}x_0\sin x_0-\operatorname{sh}x_0\cos x_0}\,. \qquad (38)$$

Free and forced vibrations. Analysis of solution

Time-dependent displacements of arms ends are calculated in the form of real parts of amplitude functions multiplied by $\exp(i\omega_0^{\pm}t)$. So, for example, at the eigenfrequency ω_0^{+} corresponding to the root of resonance equation $(\kappa_0^{+}l)=x_0/\sqrt{1+\varepsilon}$, displacements of the ends are found in form

$$u_x^{+}(l,t)=\frac{M_0\mathrm{Q}l^2}{EJ}\mathrm{N}(x_0)\cos\left(\omega_0^{+}t+\frac{\pi}{2}\right),$$
$$u_y^{+}(l,t)=(\varepsilon\mathrm{Q})\frac{M_0\mathrm{Q}l^2}{EJ}\mathrm{N}(x_0)\cos\left(\omega_0^{+}t+\frac{\pi}{2}\right)$$
$$=(\varepsilon\mathrm{Q})u_x^{+}(l,t)\,. \qquad (39)$$

From these expressions is seen that ends of the arms move along stright lines at the angle $arctan\varepsilon\mathrm{Q}$ with respect to ox-axis. Similarly, at the resonant frequency ω_0^{-} corresponding to the root $(\kappa_0^{-}l)=x_0/\sqrt{1-\varepsilon}$ we have

$$u_x^{-}(l,t)=\frac{M_0\mathrm{Q}l^2}{EJ}\mathrm{N}(x_0)\cos\left(\omega_0^{-}t+\frac{\pi}{2}\right),$$
$$u_y^{+}(l,t)=-(\varepsilon\mathrm{Q})\frac{M_0\mathrm{Q}l^2}{EJ}\mathrm{N}(x_0)\cos\left(\omega_0^{-}t+\frac{\pi}{2}\right)$$
$$=-(\varepsilon\mathrm{Q})u_x^{-}(l,t)\,. \qquad (40)$$

In the latter case, the trajectory of motion of arms ends is a stright line at the angle $-em\arctan\varepsilon Q$ with respect to ox-axis.

It is easy to verify that in the problem of free vibrations of the fork under the uniform rotational motion around its symmetry axis, displacements of the arms at resonance frequencies are found in the form

$$u_x^+(l,t) = C\cos(\omega_0^+ t)\,, \quad u_y^-(l,t) = -C\sin(\omega_0^+ t)\,,$$
$$u_x^-(l,t) = C\cos(\omega_0^- t)\,, \quad u_y^-(l,t) = C\sin(\omega_0^- t)\,, \qquad (41)$$

where C is a real arbitrary constant. Now, as is seen from (41), ends of arms of the fork move along circles of radius C. In the first case ($\omega = \omega_0^+$) this motion is clockwise whereas the arm rotates counterclockwise. In the second case ($\omega = \omega_0^-$) directions of both motions coincide. Here is a principal difference between free and forced vibrations. It may be interesting to note that in the case of free vibrations we obtain the same regularities of rotational motion as studied by the author in the problem of plane harmonic waves propagation in rotating elastic medium [2].

Using measurements of angular velocities that are very small compared to circular frequency of vibrations as a guide, so that $\varepsilon = \Omega_0/\omega_0^{\pm} \sim 10^{-4} \div 10^{-5}$, we assume that forced vibrations are excited simultaneously at frequencies ω_0^+ and ω_0^-. Then, combined motion of arms ends accordingly to (39) and (40) takes the form

$$u_x(l,t) = u_x^+(l,t) + u_x^-(l,t)$$
$$= \frac{2M_0Ql^2}{EJ}\mathrm{N}(x_0)\cos\left(\omega_0 t + \frac{\pi}{2}\right)\cos(\varepsilon\omega_0 t)\,,$$
$$u_y(l,t) = u_y^+(l,t) + u_y^-(l,t)$$
$$= (\varepsilon\mathrm{Q})\frac{2M_0Ql^2}{EJ}\mathrm{N}(x_0)\sin\left(\omega_0 t + \frac{\pi}{2}\right)\sin(\varepsilon\omega_0 t)\,, \qquad (42)$$

So, we deduce that both components of displacements obey the law of modulated vibrations. It has high frequency of occilations ω_0 and low frequency of modulation $\omega_{\mathrm{mod}} = \Omega_0$. It is seen from (42) that the maximal amplitude of vibrations in the direction perpendicular to the fork plane of is (εQ) times less than that for in-plane vibrations.

The main characteristics of a wave gyroscope is the output potential difference at the detecting part. After similar trnsformations of derivatives of amplitude functions $\hat{u}_y^+$ and $\hat{u}_y^-$ this quantity can be written in the form

$$\frac{V}{h} = \frac{M_0\mathrm{Q}}{EJ}\mathrm{K}(x_0)\frac{h}{8}\frac{1-\overline{\nu}}{d_{31}}\frac{k_p^2}{1-k_p^2}$$
$$\times\cos\left(\omega_0 t + \frac{\pi}{2}\right)\cos(\varepsilon\omega_0 t)\,, \qquad (43)$$

where is designated

$$\mathrm{K}(x_0) = \frac{1}{x_0^2}\frac{\mathrm{sh}\,x_0 - \sin x_0}{\mathrm{sh}\,x_0\sin x_0}$$
$$\times\frac{\mathrm{ch}\,x_0\sin^2 x_0 - \mathrm{sh}^2 x_0\cos x_0}{\mathrm{ch}\,x_0\sin x_0 - \mathrm{sh}\,x_0\cos x_0}\,. \qquad (44)$$

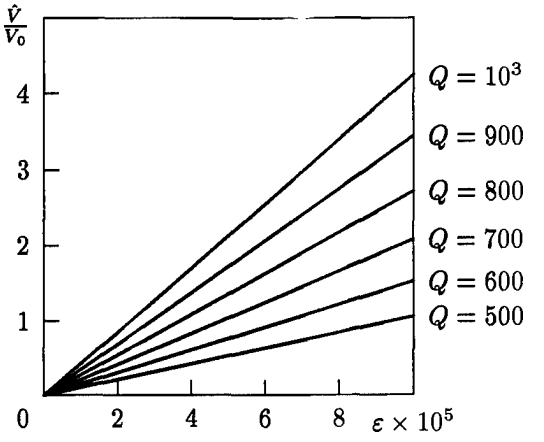


Fig. 2

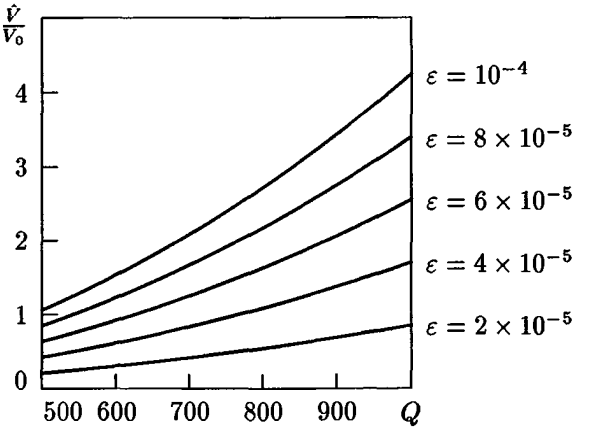


Fig. 3

Sensitivity of the gyroscope equals to the ratio between detecting and driving potential differencies. For this we have

$$\frac{\hat{V}}{V_0} = (\varepsilon\mathrm{Q})\frac{3\mathrm{Q}}{16}\frac{(1-\overline{\nu})^2}{1-\nu}\frac{k_p^2}{1-k_p^2}\frac{\mathrm{K}(x_0)}{\mathrm{sh}\,x_0\sin x_0}\,. \qquad (45)$$

Numerical evaluation of this ratio is performed in two ways. First time we assumed that the mechanical quality factor Q is constant, and obtained a linear

dependence of $\hat{V}/V_0$ on a nondimensional parameter $\varepsilon = \Omega_0/\omega_0$ for different values of Q. Another way is to assume that $\varepsilon = \Omega_0/\omega_0$ is constant, in which case this ratio behave itself as square of Q. Both of the dependencies are plotted in Fig. 2 and Fig. 3. Values of parameters in formula (44) are as follows: $x_0 = 1.875$, $\nu = 0.33$, $\overline{\nu} = 0.38$, $k_p^2 = 0.34$, From this data is seen that in suggested model the sensitivity of wave gyroscope is approximatelly 7.5×10^{-2} at Q = 500 and $\varepsilon = 2 \times 10^{-5}$.

Acknowledgements

The author would like to thank Prof. O. Yu. Zharii and A. F. Ulitko for continuous encouragement and profitable discussion. The assistance of Dr. A. Kutzenko in preparation of the manuscript is also very appreciated.

References

[1] J.D. Achenbach, *Wave Propagation in Elastic Solids*, North-Holland, Amsterdam (1973).

[2] I. A. Ulitko, "Dispersion of plane harmonic waves in uniformly rotating elastic space", *Dopovidi NAN Ukraine*, vol. 1, pp. 54–57, 1995. [in Russian]

[3] G.H. Bryan, "On the beats in the vibrations of a revolving cylinder or bell", *Proceedings of Cambridge Philosophical Society. Mathematical and Physical Sciences*, vol. 7, pp. 110–111, 1890.

[4] V.T. Grinchenko, A.F. Ulitko and N.A. Shul'ga, *Electroelasticity,* in *Mechanics of Coupled Fields in Elements of Structures*, vol. 5. Kiev, Naukova Dumka, 1989. [in Russian]

[5] S. Kudo, M. Konno, S. Sugawara and T. Yoshida, "Vibrational analysis of tuning fork gyroscope", *Japanese Journal of Applied Physics,* vol. 32, 32-1, pp. 2310–2313, 1993.

[6] N. Wakatsuki, M. Ono at al., "$LiTaO_3$ crystal fork vibratory gyroscope", *Proceedings of* 1994 IEEE *Ultrasonics Symposium,* Cannes, Ffance, November 1–4, 1994., vol. 1, pp. 581–584, 1995.

[7] D.A. Berlincourt, D.R Curran, and H.J. Jaffe, *Piezoelectric and piezomagnetic materials and their function in transducers* in *Physical Acoustics* edited by W.P. Mason, New York: Academic Press, vol. 1, pt. A., 1964.

[8] A.E.H. Love, *A Treatise on the Mathematical Theory of Elasticity,* Cambridge: University Press, 1927. vol. 35, pp. 2725–2726, 1964.

[9] S. Timoshenko, *Theory of elasticity,* New York: McGraw-Hill, 1937.

[10] S. Timoshenko, D.H. Young *Vibration Problems in Engeneering,* New York: Van Nostrad Co. Inc., 1955.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DYNAMIC BIMORPH MATRIX OF END-LOADED BIMORPHS

Jayanthy Goli[1,*], Jan G. Smits[2] and Arthur Ballato[3]
[1] Department of Biomedical Engineering, Boston University, Boston, MA 02215
[2] Department of Electrical Engineering, Boston University, Boston, MA 02215
[3] U.S. Army Research Laboratory, Fort Monmouth, NJ 07703

ABSTRACT

Piezoelectric homogeneous benders consist of two strips of piezoelectric elements with opposing polarities joined along their length. The matrix elements relating the driving parameter (voltage across the piezoelectric elements) with the response parameters have been calculated for the bimorph with tip mass. These bimorphs can be used to move an object, such as a mirror in optics, a pyramid in atomic force microscopy, or for a chemical sensor at the tip. The effect of such an object at the tip is to lower resonance frequency, which is one of the immediately recognized results. The extra mass affects the capacitance at all frequencies and also changes all the resonance frequencies. The results of this work will enable us to design bimorphs whose deflection and resonance frequency can be controlled.

INTRODUCTION

Piezoelectric bimorphs consist of two strips of piezoelectric material with opposing polarities joined longitudinally and sandwiched between two electrodes. When a voltage is applied across these electrodes, the resulting electric field across the piezoelectric material causes one strip to elongate and the other to contract. This results in a bending motion of the bimorph. This motion can be used to move microstructures or for chemical sensing applications.

*Current address: Battelle, Pacific Northwest Laboratories, Richland, WA 99352.

The static moment has been calculated by Smits et al. [1]. A full description of bending rotation, volume displacement and electrode charge caused by dynamic voltage has been determined [2]. An electric network equivalent for simple bimorphs is explained in [3].

The dynamic behavior of these bimorphs has to be understood well before they can be used as resonant pressure sensors or chemical sensors. The behavior of these bimorphs as a function of frequency and tip mass due to a driving voltage were investigated in this study.

EQUATION OF MOTION FOR THE BIMORPH

Let each of the piezoelectric strips be of length L, width w and thickness h. Two such pieces joined together along their lengths with opposing polarities yields a bender of thickness 2h sandwiched between electrodes of negligible thickness. The elastic constant of the material is designated as s^E, dielectric constant as ε^T and piezoelectric constant as d_{31}.

The length of the bimorph is along the x-coordinate, width along y and thickness along z. The deflection of the bender is along the z-axis and is represented as δ, slope of the tip as α, volume displacement as v and capacitance due to charge on the electrodes as C. The driving parameter V, the resulting parameters, and the bimorph are shown in Figure 1.

0-7803-2500-1/95 $4.00 © 1995 IEEE

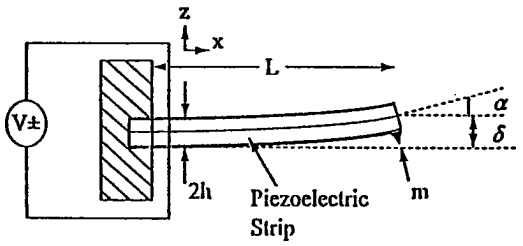


Figure 1: Homogeneous bimorph with tip mass, showing its driving and response parameters.

The equation for the curvature of the beam caused by a dynamic driving parameter is given as:

$$EIz'''' = -\rho A$$

where E is the Young's modulus, I is the moment of inertia, ρ is the specific density of the material and A is the area of cross-section of the bimorph.

The deflection of the beam is dependent on both x and t, i.e., z = z(x,t). The above equation can be solved by separation of variables, substituting $z = Z(x)\phi(t)$.

$$\frac{EI}{\rho A}\frac{Z''''}{Z} = -\frac{\ddot{\phi}}{\phi} = c$$

Substituting $c = \omega^2$ and $a^2 = EI/\rho A$ and solving for the above equation gives four possible solutions. Choosing a linear combination of these solutions yields the general solution given by:

$$z(x,t) = (c_1\cos\Omega x + c_2\sin\Omega x + c_3\cosh\Omega x + c_4\sinh\Omega x)$$

where $\Omega = \sqrt{\frac{\omega}{a}}$.

BIMORPH DRIVEN BY DYNAMIC VOLTAGE

When the bimorph is driven by an ac voltage, one of the piezoelectric strips elongates while the other contracts. This causes the bimorph to bend back and forth along the z-axis according to the frequency of the voltage. In this case, the moment is generated internally in the bimorph. By representing this internal moment caused by the voltage as an externally applied moment, the dynamic behavior of the bimorph due to an ac voltage is determined.

The expression derived for deflection, δ is:

$$\delta = \frac{wd_{31}hV\sin\Omega L\sinh\Omega L}{2\Omega^2 s_{11}^E\,(EI + EI\cos\Omega L\cosh\Omega L + A^1)}$$

where

$$A^1 = ma^2\Omega\,(-\cosh\Omega L\sin\Omega L + \cos\Omega L\sinh\Omega L)$$

The tip mass, m does not affect the anti-resonance frequencies. However, the resonance frequencies change when compared with those of a bimorph with no tip mass. The expression for the slope α of the bimorph with tip mass is:

$$\frac{d_{31}whV\,(B^1 + EI\,(\cosh\Omega L\sin\Omega L + \cos\Omega L\sinh\Omega L))}{2\Omega EIs_{11}^E\,(EI + EI\cos\Omega L\cosh\Omega L + A^1)}$$

where

$$B^1 = (-ma^2\Omega + ma^2\Omega\cos\Omega L\cosh\Omega L)$$

The tip mass affects both the resonance and anti-resonance frequencies compared to those with no tip mass. The expression for volumetric displacement of the bimorph, v is:

$$\upsilon = \frac{d_{31} h w^2 V (C^1 - EI \sin\Omega L + EI \sinh\Omega L)}{2\Omega EI s_{11}^E (EI + EI \cos\Omega L \cosh\Omega L + A^1)}$$

where

$$C^1 = ma^2\Omega (-\cos\Omega L + \cosh\Omega L - \sin\Omega L \sinh\Omega L)$$

The capacitance resulting from charge on the electrodes is:

$$C = P - Q\left(\frac{EID^1 + \sin\Omega L \cosh\Omega L + \cos\Omega L \sinh\Omega L}{EI + EI \cos\Omega L \cosh\Omega L + A^1}\right)$$

where

$$P = \frac{Lw}{2h}\left(\frac{d_{31}^2}{s_{11}^E} + \varepsilon_{33}^T\right)$$

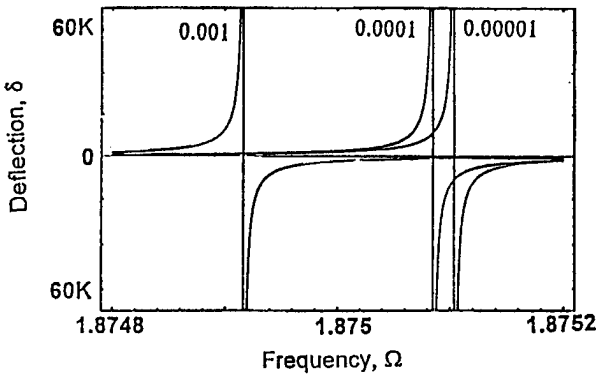
$$Q = \frac{d_{31}^2 w^2 h^2}{2 (s_{11}^E)^2 \Omega EI}$$

and $D^1 = -ma^2\Omega + ma^2\Omega \cos\Omega L \cosh\Omega L$

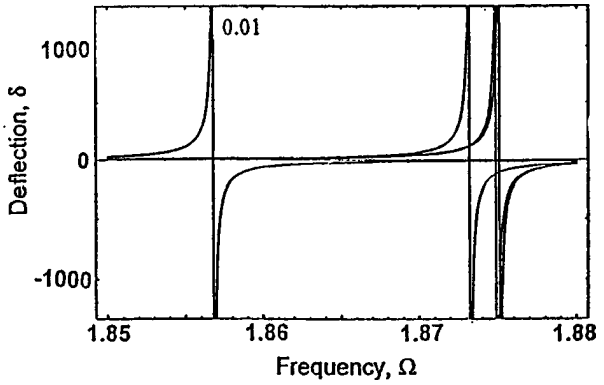
CONCLUSIONS

The dynamic behavior of end-loaded homogeneous piezoelectric bimorphs caused by ac voltage has been determined. The tip mass affects the denominator of all the response parameters in a similar fashion. The resonance frequency decreases with increasing tip mass as shown in Figure 2 (a), (b) and (c). The plot of first resonance frequency with normalized mass is shown in Figure 3. This dependence of bimorph resonance frequency on the tip mass allows it to be used for a chemical sensor, for application of small forces as in atomic force microscopy, and for moving small objects in microstructures.

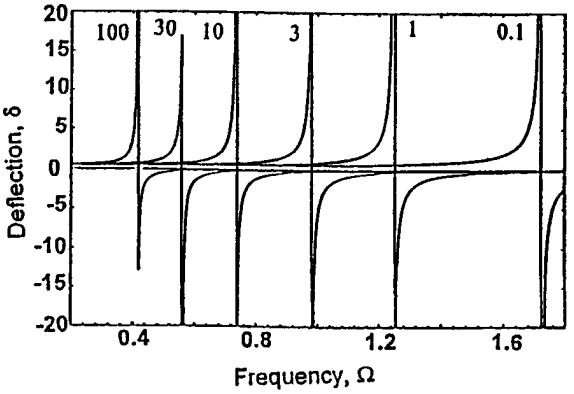
(a) δ vs Ω (reduced mass = 0.001 to 0.00001)



(b) δ vs Ω (reduced mass = 0.01)



(c) δ vs Ω (reduced mass = 0.1 to 100)



Figures 2 (a), (b), (c): First resonance frequency vs. deflection plots with varying normalized tip mass.

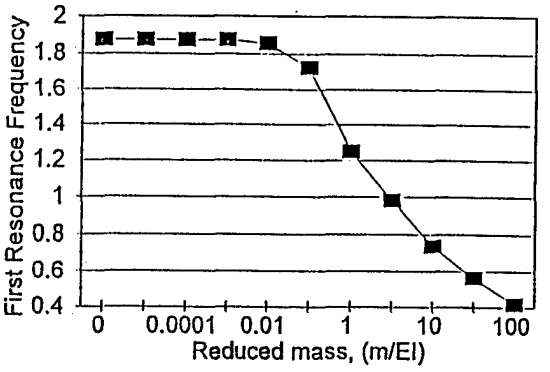


Figure 3: First resonance frequency vs. normalized tip mass plot.

REFERENCES

[1]. J.G. Smits, S.I. Dalke and T. K. Cooney, "The constituent equations of piezoelectric bimorph," *Sensors and Actuators A*, vol. 28, pp. 41-61, 1991.

[2]. J. G. Smits and A. Ballato, "Dynamic admittance matrix of piezoelectric cantilever bimorphs," *Journal of Microelectromechanical Systems*, vol. 3, No. 3, pp. 105-112, 1994.

[3]. J. G. Smits, "Equivalent circuit for end-loaded piezoelectric bimorph actuators," *Ferroelectrics*, vol. 60, pp. 141-148, 1984.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Torsional Vibrations of Quartz Crystal Beams

Hirofumi KAWASHIMA and Kenji SUNAGA

Seiko Instruments Inc.
Matsudo-shi, Japan 271

Abstract

An exact solution of a partial differential equation including elastic compliance constant s'_{56}, with respect to stress function ψ has been found for torsional modes of vibration of an arbitrary (singly, doubly, triply) rotated beam with a pair of parallel, free edges. The solution is obtained by relaxing the condition that the edge planes are perpendicular to the main faces of the beam. That is, the edges are off perpendicular by the angle $\Theta=\arctan(-s'_{56}/s'_{55})$. The exact solution can reduce the difference of the calculated and measured values for a thickness-to-width ratio where gives the first order temperature coefficient $\alpha=0$. Also, a comparatively large-inclination of the edge cuts is required to reduce the unwanted, complicated mode shapes to simple ones.

1. Introduction

Since a frequency constant of a resonator vibrating in torsion[1-3] lies intermediately to those for flexural and length extensional mode resonators, it is very available for realization of miniature resonators of a comparatively low-frequency, i.e., 400 to 600kHz. In order that one of the authors clarifies frequency temperature behavior of doubly rotated torsional quartz crystal resonators of tuning fork-type which can be approximately regarded as a clamped-free bar, and free-free bar-type, a theoretical and/or experimental examination was performed in detail. As a result, we have already obtained and reported[1,2] that a cut angle (ϕ, θ) where gives the first order temperature coefficient $\alpha=0$ and approximately one third of the second order temperature coefficient β for flexural mode quartz crystal resonators, exists for the torsional quartz crystal resonator by combination of the cut angle(ϕ, θ) and a thickness-to-width ratio, the so called TT-Cut classified into the TT(X_1)- and TT(X_2)-Cuts.

A vibration problem of torsion is equivalent to solving the problem of torsional rigidity C_t. In the previous report[1] C_t was derived from stress function ψ. However, the obtained stress function ψ is an approximate solution derived neglecting the elastic stiffness constant which couples two shear stresses.

In this paper, an exact solution of torsional rigidity C_t for a torsional mode resonator is found and applied to quartz crystal. In analysis procedure, first, a frequency equation is, as already reported[1-3], found to be given as a function of torsional rigidity C_t. Next, a general stress-strain relation is transformed into a simple one from the conditions of resonator shape and dimensions, and from the equation of equilibrium a partial differential equation is derived with respect to stress function ψ. In addition, the ψ which satisfies the boundary conditions is derived and torsional rigidity C_t is calculated. Finally, the obtained result is applied to a quartz crystal resonator and is compared with the measured ones of the quartz crystal resonator vibrating in torsion.

2. Analysis Procedure

2.1 Frequency equation

Figure 1 shows a resonator of a beam vibrating in torsion and its coordinate system. The resonator is expressed in the IEEE notation of (YXlt) ϕ/θ and has a dimension of length $2x_0$, width $2z_0$ and thickness $2y_0$. When length $2x_0$ is much larger than width $2z_0$ and thickness $2y_0$, lateral motion for a resonator vibrating in torsion can be neglected[4]. A resonant frequency f for a

0-7803-2500-1/95 $4.00 © 1995 IEEE

torsional resonator is, therefore, given as

$$f=\frac{m}{4(2x_0)}\sqrt{\frac{C_t}{\rho I_0}}, \tag{1}$$

where C_t : Torsional rigidity
ρ : Density of beam
I_0 : Polar moment of inertia
$2x_0$: Length of beam
m : Odd for clamped-free beam and even for free-free beam.

As can be easily understood from eq.(1), a vibration problem of torsion is equivalent to solving the problem of torsional rigidity C_t because I_0 is determined by the resonator shape of section.

2.2 Equations of equilibrium

Now, when displacements in the x,y and z axes direction are taken as u, v and w, the displacements u, v, w can be expressed by

$$\begin{aligned} u &= \tau\,\Phi(y,z) \\ v &= -\tau xz \\ w &= \tau xy, \end{aligned} \tag{2}$$

where torsion function Φ is a function of y and z, and τ is the twist. From the relation of eq.(2), the strain-components vanish and are

$$S_1 = S_2 = S_3 = S_4 = 0. \tag{3}$$

A general strain-stress relation has the form

$$\begin{aligned} S_1 &= s_{11}T_1 + s_{12}T_2 + s_{13}T_3 + s_{14}T_4 + s_{15}T_5 + s_{16}T_6 \\ S_2 &= s_{21}T_1 + s_{22}T_2 + s_{23}T_3 + s_{24}T_4 + s_{25}T_5 + s_{26}T_6 \\ S_3 &= s_{31}T_1 + s_{32}T_2 + s_{33}T_3 + s_{34}T_4 + s_{35}T_5 + s_{36}T_6 \\ S_4 &= s_{41}T_1 + s_{42}T_2 + s_{43}T_3 + s_{44}T_4 + s_{45}T_5 + s_{46}T_6 \\ S_5 &= s_{51}T_1 + s_{52}T_2 + s_{53}T_3 + s_{54}T_4 + s_{55}T_5 + s_{56}T_6 \\ S_6 &= s_{61}T_1 + s_{62}T_2 + s_{63}T_3 + s_{64}T_4 + s_{65}T_5 + s_{66}T_6 \end{aligned}, \tag{4}$$

where s_{qp} is elastic compliance constant. From eq.(3) the strain-stress relation reduces to

$$\begin{aligned} S_5 &= s'_{55}T_5 + s'_{56}T_6 \\ S_6 &= s'_{65}T_5 + s'_{66}T_6, \end{aligned} \tag{5}$$

where $s'_{qp} = \hat{s}_{qp} - \hat{s}_{4q}\hat{s}_{p4}/\hat{s}_{44}$ $(q,p = 5,6)$
$\hat{s}_{qp} = \tilde{s}_{qp} - \tilde{s}_{3q}\tilde{s}_{p3}/\tilde{s}_{33}$ $(q,p = 4,5,6)$
$\tilde{s}_{qp} = \bar{s}_{qp} - \bar{s}_{2q}\bar{s}_{p2}/\bar{s}_{22}$ $(q,p = 3,4,5,6)$
$\bar{s}_{qp} = s_{qp} - s_{1q}s_{p1}/s_{11}$ $(q,p = 2,3,4,5,6)$.

In addition, the equations of equilibrium yield

$$\begin{aligned} \frac{\partial T_1}{\partial x} + \frac{\partial T_6}{\partial y} + \frac{\partial T_5}{\partial z} &= 0 \\ \frac{\partial T_6}{\partial x} + \frac{\partial T_2}{\partial y} + \frac{\partial T_4}{\partial z} &= 0 \\ \frac{\partial T_5}{\partial x} + \frac{\partial T_4}{\partial y} + \frac{\partial T_3}{\partial z} &= 0. \end{aligned} \tag{6}$$

Since stress-components T_1, T_5 and T_6 are given as a function of y and z, we find that

$$\frac{\partial T_1}{\partial x} = \frac{\partial T_5}{\partial x} = \frac{\partial T_6}{\partial x} = 0. \tag{7}$$

When length $2x_0$ for a torsional resonator is much larger than width $2z_0$ and thickness $2y_0$, stress-components T_2, T_3 and T_4 become

$$T_2 = T_3 = T_4 = 0. \tag{8}$$

From eqs.(7) and (8), the second and third equations of eq.(6) are always satisfied, so that the equation of equilibrium is given by

$$\frac{\partial T_6}{\partial y} + \frac{\partial T_5}{\partial z} = 0. \tag{9}$$

Equation(9) is satisfied if

$$T_5 = -\frac{\partial \psi}{\partial y} \quad \text{and} \quad T_6 = \frac{\partial \psi}{\partial z}, \tag{10}$$

ψ is stress function. Substitution of eqs.(2) and (10) into eq.(5) then yields

$$s'_{55}\frac{\partial^2 \psi}{\partial y^2} + s'_{66}\frac{\partial^2 \psi}{\partial z^2} - 2s'_{56}\frac{\partial^2 \psi}{\partial y \partial z} = -2\tau, \tag{11}$$
$$(\because s'_{56} = s'_{65})$$

which must hold over the area of the cross section, while the boundary conditions must be satisfied with

$$T_5 = 0 \quad on \quad z = \pm z_0$$

and

$$T_6 = 0 \quad on \quad y = \pm y_0. \tag{12}$$

Hence, the stress function ψ is derived from eq.(11) and the boundary conditions of eq.(12).

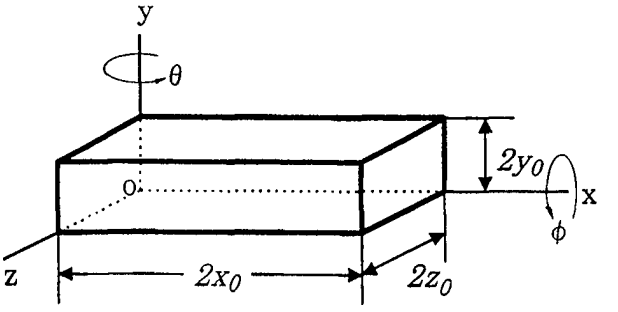


Fig.1 Torsional resonator of a beam and its coordinate system.

2.3 Stress function ψ

Now, we take the stress function ψ as

$$\psi = a_n\left\{1 - b_n \cosh\xi\left(z + \zeta y\right)\right\}\cos\eta y, \tag{13}$$

where a_n, b_n, ξ, ζ and η are constant. Substitution of eq.(13) into eq.(11) then gives

$$a_n\Big\{b_n\big(s'_{55}\eta^2 - s'_{55}\xi^2\zeta^2 - s'_{66}\xi^2 + 2s'_{56}\xi^2\zeta\big)\cosh\xi\big(z+\zeta y\big)\cos\eta y + 2b_n\big(s'_{55}\eta\xi\zeta - s'_{56}\eta\xi\big)\sinh\xi\big(z+\zeta y\big)\sin\eta y - \eta^2 s'_{55}\cos\eta y\Big\} = -2\tau . \quad (14)$$

Accordingly, to hold eq.(14) at an arbitrary value of $\cosh\xi\big(z+\zeta y\big)\cos\eta y$ and $\sinh\xi\big(z+\zeta y\big)\sin\eta y$, we should find that

$$s'_{55}\eta^2 - s'_{55}\xi^2\zeta^2 - s'_{66}\xi^2 + 2s'_{56}\xi^2\zeta = 0$$

and

$$s'_{55}\eta\xi\zeta - s'_{56}\eta\xi = 0 . \quad (15)$$

From eq.(15) ζ and ξ, respectively, become

$$\zeta = \frac{s'_{56}}{s'_{55}} \quad (16)$$

and

$$\xi = \sqrt{\frac{s'_{55}}{s'_{66}}}\frac{1}{\sqrt{1-\dfrac{s'^2_{56}}{s'_{55}s'_{66}}}}\eta , \quad (17)$$

so that equation(14) is simplified into

$$a_n\eta^2 s'_{55}\cos\eta y = 2\tau . \quad (18)$$

Now when $-y_0 < y < +y_0$, the 2τ of the right term in eq.(18) can be expanded in a Fourier's series[5] as follows:

$$2\tau = 2\tau\sum_{n=1,3,5}^{m}\frac{4}{n\pi}(-1)^{(n-1)/2}\cos\eta y . \quad (19)$$

Hence, a_n of eq.(18) can be expressed in the form

$$a_n = \frac{8\tau}{\eta^2\pi}\frac{1}{s'_{55}}\sum_{n=1,3,5}^{m}\frac{1}{n}(-1)^{(n-1)/2} . \quad (20)$$

Next, the constants η and b_n are derived from the boundary conditions. The stress-component T_6 of eq.(10) becomes

$$T_6 = \frac{\partial\psi}{\partial z} = a_n b_n\xi\sinh\xi\big(z+\zeta y\big)\cos\eta y , \quad (21)$$

the boundary condition of $T_6 = 0$ on $y = \pm y_0$ then gives

$$\eta = \frac{n\pi}{2y_0} \quad (n = 1,3,5\cdots) . \quad (22)$$

In addition, since we can not find a solution which satisfies the boundary condition of $T_6 = 0$ on $z = \pm z_0$, the edges are, as illustrated in Fig.2, off perpendicular by the angle Θ. When a resonator vibrating in torsion is rotated about the x-axis, a transformation of an old coordinate system(x,y,z) into a new one (x', y', z') yields

$$\begin{pmatrix}x'\\y'\\z'\end{pmatrix} = \begin{pmatrix}1 & 0 & 0\\0 & \cos\Theta & \sin\Theta\\0 & -\sin\Theta & \cos\Theta\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} . \quad (23)$$

Accordingly, the transformed stress-components T'_5 and T'_6 is given by

$$T_5' = T_5\cos\Theta - T_6\sin\Theta$$

and

$$T_6' = T_5\sin\Theta + T_6\cos\Theta . \quad (24)$$

Here we note that

$$y' = y\cos\Theta$$

and

$$z' = \big(z+\zeta y\big)\cos\Theta , \quad (25)$$

the stress-components T_5 and T_6 are, therefore, transformed into

$$T_5 = -\frac{\partial\psi}{\partial y} = -\cos\Theta\frac{\partial\psi}{\partial y'} - \zeta\cos\Theta\frac{\partial\psi}{\partial z'}$$

and

$$T_6 = \frac{\partial\psi}{\partial z} = \cos\Theta\frac{\partial\psi}{\partial z'} . \quad (26)$$

From the relation of eq.(26) T'_5 reduces to

$$T_5' = -\cos^2\Theta\frac{\partial\psi}{\partial y'} - \big(\zeta\cos\Theta + \sin\Theta\big)\cos\Theta\frac{\partial\psi}{\partial z'} . \quad (27)$$

In eq.(27), we set

$$\zeta\cos\Theta + \sin\Theta = 0 , \quad (28)$$

so that we obtain

$$\Theta = \arctan\big(-s'_{56}/s'_{55}\big) . \quad (29)$$

Accordingly, T'_5 is simplified into

$$T_5' = -\cos^2\Theta\frac{\partial\psi}{\partial y'} , \quad (30)$$

and finally, T'_5 reduces to

$$T'_5 = a_n\eta\cos\Theta\sin\eta y\left(1 - b_n\cosh\frac{\xi}{\cos\Theta}z'\right) . \quad (31)$$

From the boundary condition of $T'_5 = 0$ on $z' = \pm z_0\cos\Theta$, we should find that

$$b_n = \frac{1}{\cosh\xi z_0} . \quad (32)$$

From these results, stress function ψ is given by

$$\psi = \frac{8\tau(2y_0)^2}{\pi^3}\frac{1}{s'_{55}}\sum_{n=1,3,5}^{m}\frac{1}{n^3}(-1)^{(n-1)/2} \times\left[1 - \frac{\cosh\xi\big(z+\zeta y\big)}{\cosh\xi z_0}\right]\cos\eta y . \quad (33)$$

Namely, ψ of eq.(33) is an exact solution which includes the elastic compliance constant s'_{56}.

2.4 Torsional rigidity C_t

Since torsional rigidity C_t is given by a ratio of torsional moment M_t and the twist τ (M_t/τ), the M_t which is defined by the integration of ψ, i.e., $2\int_{-z_0}^{z_0}\int_{-y_0}^{y_0}\psi\,dy\,dz$ is calculated, so that M_t is

$$M_t=\frac{1}{3}\tau\frac{1}{s'_{55}}(2y_0)^3(2z_0)\left[1-\frac{192}{\pi^5}\left(\frac{y_0}{z_0}\right)\right.$$
$$\times\sqrt{\frac{s'_{66}}{s'_{55}}}\sqrt{1-\frac{s'^2_{56}}{s'_{55}s'_{66}}}\left(1-\frac{s'^2_{56}}{s'_{55}s'_{66}}\right)$$
$$\times\sum_{n=1,3,5}^{m}\frac{1}{n^5}\cosh\left(\frac{n\pi}{2}\frac{s'_{56}}{\sqrt{s'_{55}s'_{66}}}\frac{1}{\sqrt{1-\frac{s'^2_{56}}{s'_{55}s'_{66}}}}\right)$$
$$\left.\times\tanh\left(\frac{n\pi}{2}\sqrt{\frac{s'_{55}}{s'_{66}}}\frac{1}{\sqrt{1-\frac{s'^2_{56}}{s'_{55}s'_{66}}}}\frac{z_0}{y_0}\right)\right]. \qquad (34)$$

Here, we note that the constant m must be so determined that it fulfills the convergent condition of M_t. Namely, it may need the following relation

$$\sqrt{s'_{55}s'_{66}-s'^2_{56}}>ms'_{56}\,, \qquad (35)$$

From the relation of eq.(34) torsional rigidity C_t is given by

$$C_t=\frac{1}{3}\frac{1}{s'_{55}}(2y_0)^3(2z_0)\left[1-\frac{192}{\pi^5}\left(\frac{y_0}{z_0}\right)\sqrt{\frac{s'_{66}}{s'_{55}}}\sqrt{1-\frac{s'^2_{56}}{s'_{55}s'_{66}}}\left(1-\frac{s'^2_{56}}{s'_{55}s'_{66}}\right)\right.$$
$$\left.\times\sum_{n=1,3,5}^{m}\frac{1}{n^5}\cosh\left(\frac{m\pi}{2}\frac{s'_{56}}{\sqrt{s'_{55}s'_{66}}}\frac{1}{\sqrt{1-\frac{s'^2_{56}}{s'_{55}s'_{66}}}}\right)\tanh\left(\frac{m\pi}{2}\sqrt{\frac{s'_{55}}{s'_{66}}}\frac{1}{\sqrt{1-\frac{s'^2_{56}}{s'_{55}s'_{66}}}}\frac{z_0}{y_0}\right)\right]. \qquad (36)$$

In addition, when $s'_{56}=0$, C_t reduces to

$$C_t=\frac{1}{3}\frac{1}{s'_{55}}(2y_0)^3(2z_0)\left[1-\frac{192}{\pi^5}\left(\frac{y_0}{z_0}\right)\sqrt{\frac{s'_{66}}{s'_{55}}}\right.$$
$$\left.\times\sum_{n=1,3,5}^{\infty}\frac{1}{n^5}\tanh\left(\frac{m\pi}{2}\sqrt{\frac{s'_{55}}{s'_{66}}}\frac{z_0}{y_0}\right)\right], \qquad (37)$$

this result agrees with that[1,2,5,6] derived from the first condition of $s'_{56}=0$.

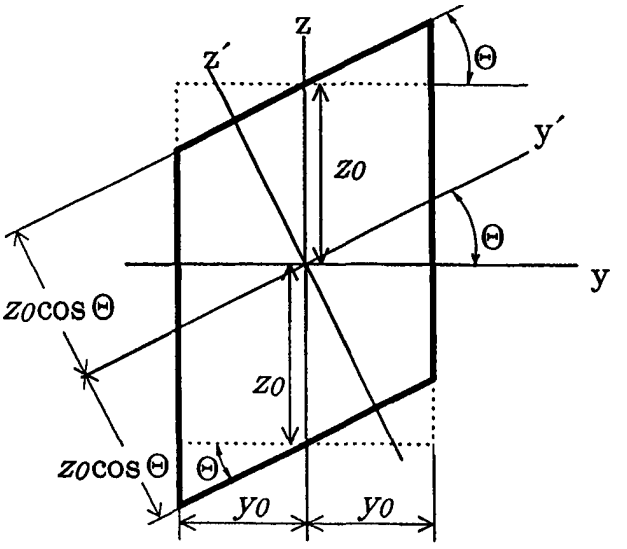


Fig.2 Sectional view and rotated axes of quartz crystal resonator.

3. Resonator Shape and Excitation Electrodes

Figure 3 shows resonator shape and dimensions of tuning fork-type vibrating in torsion which are used in the present experiments(a), and the sectional view of electrodes construction(b). The edges of arms of tuning fork are off perpendicular by the angle Θ. The resonators are formed from a rotated Y-plate with a cut angle of ϕ and θ about the x- and y'-axes. In addition, the excitation electrodes are deposited on obverse and reverse of the arms of tuning fork and so that the adjoining electrodes have different polarity.

4. Results and Discussion

The results obtained in the present analysis are compared with the measured values of torsional quartz crystal resonators of tuning fork-type which can be approximately regarded as a

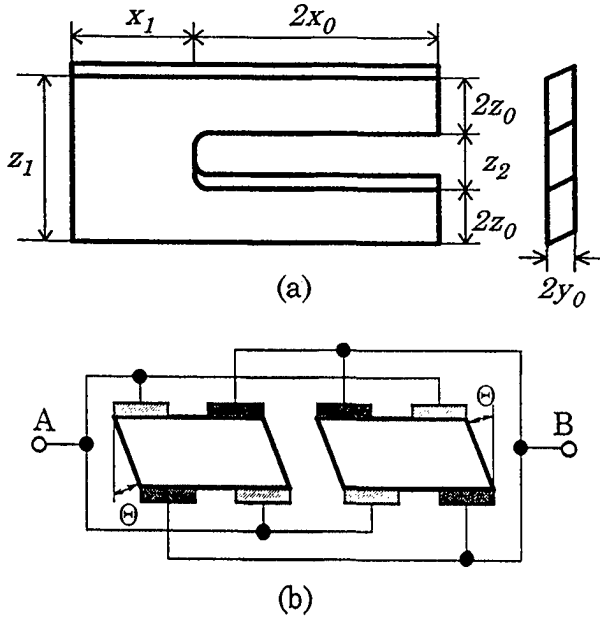


Fig.3 Shape and dimensions of torsional quartz crystal resonator of tuning fork-type and its sectional view of electrodes construction.

clamped-free bar. First, a relationship of a frequency constant versus a thickness-to-width ratio R_{yz} is shown. Next, a relationship of the first- and second- order temperature coefficients α, β versus a thickness-to-width ratio R_{yz} is shown. In the calculation the constants of the reference(7) are employed.

4.1 Frequency constant

Figure 4 shows the relationship of a frequency constant($f \cdot 2x_0$) to a thickness-to-width ratio R_{yz} of a quartz crystal resonator vibrating in torsion with a cut angle of $\phi = 66°$ and $0 = 0°$, and an edge angle of $\Theta = 24°$. The frequency constant($f \cdot 2x_0$) is approximately proportional to the aspect ratio R_{yz} when it is small. As is shown in Fig.4, the frequency constant($f \cdot 2x_0$) increases gradually according to increase of the aspect ratio R_{yz}. However, further increase of R_{yz} gives the maximum of the frequency constant in the vicinity of R_{yz}=0.9, and then, an inclination to decrease gradually. For example, the frequency constant($f \cdot 2x_0$) is 91.7kHz·cm in the calculation and 89.9kHz·cm in the experiments when a thickness-to-width ratio R_{yz} is 0.917 where the first order temperature coefficient α becomes fairly small at $\phi = 66°$, $0 = 0°$ and $\Theta = 24°$. Since these values lie intermediately to those for flexural and length extensional mode resonators, a miniature torsional

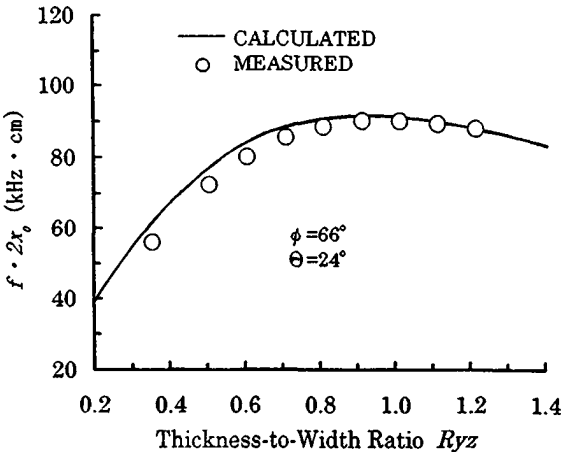


Fig.4 Relationship of frequency constant ($f \cdot 2x_0$) and thickness-to-width ratio Ryz.

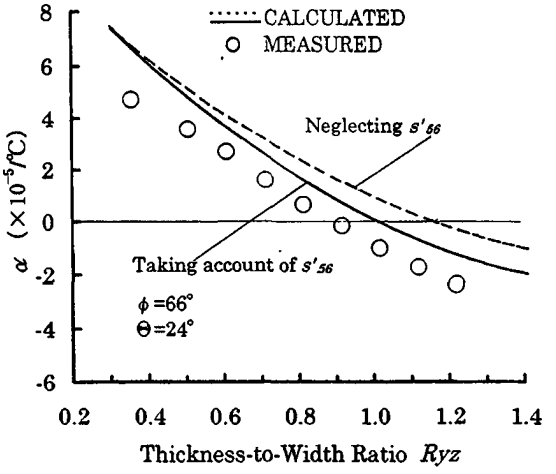


Fig.5 Relationship of the first order temperature coefficient α and thickness-to-width ratio Ryz.

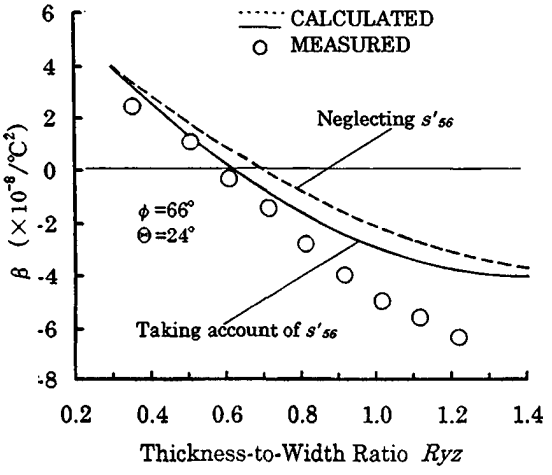


Fig.6 Relationship of the second order temperature coefficient β and thickness-to-width ratio Ryz.

quartz crystal resonator of 400 to 600kHz can be realized when the resonator size is taken into account.

4.2 Frequency temperature coefficients

Figure 5 shows the relationship of the first order temperature coefficient α and a thickness-to-width ratio R_{yz} of a torsional quartz crystal resonator with a cut angle of $\phi=66°$, $\theta=0°$ and $\Theta=24°$. When the elastic compliance constant s'_{56} is taken into account, a thickness-to-width ratio R_{yz} where gives $\alpha=0$, is 1.01 in the calculation and 0.906 in the experiments, and the difference of about 10 percent occurs between them, while, when $s'_{56}=0$, the aspect ratio R_{yz} where gives $\alpha=0$ is 1.16 in the calculation, and the difference of the calculated and measured values is about 22 percent. Thus, the neglect of s'_{56} is found to give a larger difference between the calculated and measured values. In addition, when s'_{56} is taken into account, the theoretical relationship of R_{yz} and α is in good agreement with the inclination of the experimental one, and the calculated values for the small-shifted R_{yz} by 0.1 agree well with the measured ones. As one of the reasons why the difference takes place between the calculated and measured values, it is conceivable that the piezoelectric terms were neglected in the theoretical analysis. An analysis which includes the terms will be performed for a coming subject.

Figure 6 shows the relationship of the second order temperature coefficient β and a thickness-to-width ratio R_{yz} for the resonator of the same condition as in Fig.5. β decreases as R_{yz} increases. However, the difference of the calculated and measured values for β to the increase of R_{yz} becomes large gradually from the vicinity of R_{yz} =0.8 because the variation quantity of the calculated values becomes smaller than that of the measured ones. In addition, the influence of s'_{56} on β is the same inclination as that on α, and the calculated values which took account of s'_{56}, as is shown in Fig.6, approach to the measured ones.

5. Conclusions

In this paper, an exact solution of torsional rigidity C_t for a resonator vibrating in torsion has been found. This theoretical result was applied to torsional quartz crystal resonators of tuning fork-type and the calculated values are compared with the measured ones.

First, the general stress-strain relation was transformed into a simple one from the conditions of resonator shape and dimensions, and from the equation of equilibrium the partial differential equation was derived with respect to the stress-function ψ. In addition, the ψ which satisfies the boundary conditions was derived and then, the torsional rigidity C_t was calculated.

Next, the relationship of a frequency constant($f \cdot 2x_0$) versus a thickness-to-width ratio R_{yz} was clarified from the theoretically obtained equation, so that the calculated values were found to agree well with the measured ones, e.g., at the aspect ratio R_{yz} where an absolute value of the first order temperature coefficient α becomes fairly small.

Finally, the relationship of the first- and second-order temperature coefficients α, β versus a thickness-to-width ratio R_{yz} was clarified. As a result, it was shown that the calculated values where s'_{56} was taken into account, approach further to measured ones. Especially, for the relationship of α versus R_{yz} the inclination of the calculated values was found to be in good agreement with that of the measured ones.

For the coming subject the influence of the piezoelectric terms on α and β will be clarified.

References

1) H.Kawashima and M.Nakazato: "A New Cut Torsional Quartz Crystal Resonator of Tuning Fork-type" , Trans. IEICE(A), Vol.J76-A, No.1, pp.1-8, 1993 (in Japanese).

2) H.Kawashima and M.Nakazato : "TT- Cut Torsional Quartz Crystal Resonators of Free-Free Bar-Type" , Jpn.J.Appl. Phys.Vol.33 (1994) Pt.1,No.5B,pp.2989-2993.

3) J.Hermann : "DT-Cut Torsional Resonator" ,Proc. 31st Annual Symp. on Freq. Control, pp.55-61(1977).

4) H.Kawashima and K.Sunaga : "A Vibration Analysis of Torsional Quartz Crystal Resonators Taking Account of Lateral Motion" , Jpn.J.Appl.Phys. Vol.34 (1995) Pt.1 No.5B, pp.3316-3319.

5) A.E.H.Love : "A Treatise on the Mathematical Theory of Elasticity" , Fourth edition, Cambridge University Press, Cambridge, England, 1952.

6) H.Kawashima,M.Nakazato and Y.Shimizu :"Quartz Temperature Sensors of Torsional Vibration Formed from Doubly Rotated plates" ,Jpn.J.Appl.Phys.Vol.33,(1994) Pt.1, No.5B, pp.3005-3009.

7) M.Aruga : "On the elastic constants of quartz and their temperature characteristics" , Bull.Tokyo Inst. Tech., Vol.A-2, pp.88-182, (1956) (in Japanese).

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DEPENDENCE OF FREQUENCY-TEMPERATURE CHARACTERISTICS OF QUARTZ RESONATORS ON FORM, DIMENSIONS AND LOCAL VARIATIONS OF THICKNESS OF PIEZOELECTRIC CRYSTAL

I.I.Postnikov

Special Research Bureau of Moscow Power Engineering Institute,
Krasnokazarmennaya, 14, Moscow, Russia, 111250

Abstract

An algorithm and computing program for solving a boundary-value problem are developed to determine temperature dependencies of eigenvalues and eigenfunctions of quartz resonators with rotated Y-cut piezoelectric crystals with axially symmetry and varying thickness. The possibility of determining of these dependencies is demonstrated for crystals with varying thickness and different cut directions for both harmonic and nonharmonic vibrations.

Introduction

The problem for determining of the dependence of temperature on frequency in crystal-controlled and selectivity acoustoelectronic devices is the object of numerous theoretical and experimental studies aimed at searching for ways of compensating circuitry and by limiting temperature variations, or else aimed at suppressing temperature dependence on the frequency of a piezoelectric crystal itself by choosing its crystallographic orientation about crystallographic axis.

Studies of variously oriented piezoelectric crystals have revealed more that twenty possible quartz cut directions, which provide thickness shear type vibrations. However only about one third of these type vibrations were used in practical applications (Table 11.1 in [1]). All these cuts refer to the family of rotated Y-cut, where particle displacements in a thickness shear wave are directed exactly along the electrical axis [2].

It is found experimentally that frequency-temperature characteristics (FTC) of all quartz resonators may have either a quadric or cube parabola form or a straight line and may be represented in a temperature range of -200 to +200° C with accuracy good enough for practical applications as a third degree polynomial [1,3,4]:

$$\delta_f = (f - f_o)/ f_o = \sum_{n=1}^{m} T_f^{(n)} (t - t_o)^n \qquad (1)$$

where $T_f^{(n)} = 1/(n! \cdot f_o) \cdot (\partial^n f/\partial t^n) /t=t_o$ is the temperature coefficient of a frequency of the nth order; t is the temperature of the piezoelectric crystal; t_o is initial temperature; f and f_o are resonance frequencies of the piezoelectric crystal at temperatures t and t_o, respectively.

For widely used temperature intervals, it is reasonable to restrict consideration to case m=3. Then, with regard to (1), we can write:

$$f(t) = f_o[1+T_f^{(1)}(t-t_o)+T_f^{(2)}(t-t_o)^2+ + T_f^{(3)}(t-t_o)^3] \qquad (2)$$

For resonators of rotated Y-cuts with thickness shear vibrations, the resonance frequency for infinite plate can be determined by the expression [4]:

$$f(t) = [N/h_o(t)] \sqrt{C_{66}^*(t)/\rho_\kappa(t)}, \qquad (3)$$

$$C_{66}^*(t)=C_{66}(t)\cdot\cos^2\beta+C_{44}(t)\cdot\sin^2\beta+ +C_{14}(t)\cdot\sin 2\beta \qquad (4)$$

where N is the frequency coefficient, h_o is the thickness of the piezoelectric crystal, C_{ij} are elasticity module, ρ_κ is density of the material , and β is the angle of rotation of the piezoelectric crystal about the x axis.

0-7803-2500-1/95 $4.00 © 1995 IEEE

Temperature dependencies of elasticity module C_{ij}, of the density ρ and the thickness of the piezoelectric crystal h_0 (the frequency determining measure), which are involved in (3) and (4), can also be represented by third-degree polynomials:

$$C_{ij}(t) = C_{ij}(t_0)[1+\sum_{n=1}^{3} T_c^{(n)} (t - t_0)^n]; \quad (5)$$

$$h(t) = h(t_0)[1+\sum_{n=1}^{3} T_h^{(n)} (t - t_0)^n]; \quad (6)$$

$$\rho_K(t) = \rho(t_0)[1+\sum_{n=1}^{3} T_\rho^{(n)} (t - t_0)^n]; \quad (7)$$

where $T_c^{(n)}$, $T_h^{(n)}$, and $T_\rho^{(n)}$ are nth-order temperature coefficients of the elasticity modules, thickness, and piezoelectric crystal density, respectively.

Effective elastic constant C_{66} depends on the elastic constants C_{ij} of the crystal and on angle β between the z axis and the quasi-neutral plane of the quartz crystal parallel to the cut plane of the zero temperature coefficient of frequency. Because the temperature coefficients of module C_{ij} in (4) have opposite signs ($T_{c66}^{(1)} = 178 \cdot 10^{-6}$ 1/°C; $T_{c44}^{(1)} = -177 \cdot 10^{-6}$ 1/°C; $T_{c14}^{(1)} = 101 \cdot 10^{-6}$ 1/°C [4]), it is possible to find an angle of the crystal orientation β such (Fig.1) that the resulting temperature coefficient will be $T_f^{(1)} = 0$.

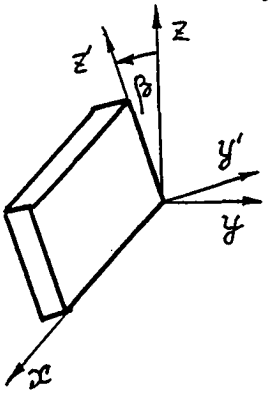


Fig.1. Rotated Y-cuts

By differentiating dependencies (2),(5)-(7), three times with respect to temperature we can easily find a relation between temperature coefficients of frequency of all three orders and temperature coefficients of ρ, C_{66}^*, and h [2,4], and, consequently, we can determine the form of FTC.

There are a large number of theoretical and experimental studies devoted to analysis and synthesis of FTC and the temperature coefficient of frequency of quartz resonators [4-6, 7-9, 10]. These studies show possibilities of compensating for and attaining a small or zero temperature coefficient of frequency by appropriate selection of piezoelectric crystal orientation. However, all studies on FTC of quartz resonators have so far been based on the above relations for infinite plates and provide no quantitative estimates of the effect on resonance frequency of the full complex of factors, such as piezoelectric crystal geometry, crystal orientation, and temperature. A lack of corresponding studies is most upsetting in the case of multimode resonators, where the above-listed factors influence all working vibrations.

Meanwhile, certain methods for eigenvalue and eigenfunction calculations in axially symmetric piezoelectric crystals of varying thickness [11-13], where AT-cut piezoelectric crystals are considered while ignoring temperature effects on resonator characteristics, can be modified according to the new requirements. Such modification is demonstrated below by developing an algorithm to determine eigenvalues and eigenfunctions of resonators with a rotated, Y-cut, piezoelectric crystal of circular shape and varying thickness.

In order to determine their FTC and temperature coefficients of frequency for piezoelectric crystals, characterized by axial symmetry and varying thickness by constructing equations for thickness shear vibrations [13] it is necessary to take into account temperature dependencies of ρ and C_{ij} values and the dimensions of a given piezoelectric crystal:

$$\begin{aligned} &2[A_1F''_m + A_3F'_m+(A_9-mA_7)F_m] \\ &+ A_2F''_{m+2} + [A_4 - (m+2)A_7]F'_{m+2} \\ &+[A_{10} - (m+2)A_6 - (m+2)^2A_8]\ F_{m+2} \\ &+A_2F''_{m-2} - [A_4 - (m-2)A_5]\ F'_{m-2} \\ &+[A_{10} + (m-2)A_6 - (m-2)^2A_8]\ F_{m-2} \\ &+k\{A_2F''_{2-m} + [A_4 + (m-2)A_5]F'_{2-m} \\ &+[A_{10} + (m-2)A_6 - (m-2)^2A_8]\ F_{2-m}\} = 0. \end{aligned}$$

The primes here indicate differentiation with respect to r (designations of geometric parameters of piezoelectric crystal are shown in Fig.2), and varying coefficients A_j are determined from the following relations:

$$A_1 = g_5 + g_6;\ A_2 = g_1 - g_6;$$
$$A_3 = A_1/r+\delta g_1/\delta r+2g_6\delta D_6/\delta r;$$
$$A_4 = -A_2/r+\delta g_1/\delta r+2g_6\delta D_6/\delta r;$$
$$A_5 = -2A_2/r^2;$$
$$A_6 = 2A_2/r^2-(\delta g_1/\delta r+2g_6\delta D_6/\delta r)/r;$$
$$A_7 = A_1/r^2;\ A_8 = -A_2/r^2;$$
$$A_9 = g_6((1/r)\delta D_6/\delta r + \delta^2 D_6/\delta r^2) + h^3\rho_\kappa\omega^2/6 - 2D_6; \quad (9)$$
$$A_{10} = g_6(\delta^2 D_6/\delta r^2 - (1/r)(\delta D_6/\delta r));$$
$$g_6 = D^2_6/ h\rho_\kappa\omega_o^2;\ g_1 = D_1 - D_5;\ g_5 = D_1 + D_5;$$
$$D_1 = h^3[C_{11}^* - (C_{12}^*)^2/C_{22}^*]/12;$$
$$D_5 = h^3C_{55}^*/12;$$
$$D_6 = \pi^2 hC_{66}^*/12;\ k = -1,0,1,$$

where ω and ω_o are angular frequencies of a piezoelectric crystal and an infinite plate of thickness h_o, respectively; and C_{ij}^* are elasticity module in piezoelectric plate coordinate system x,y',z'.

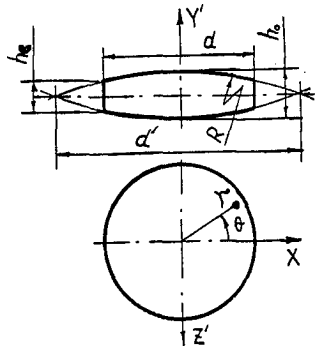


Fig.2. Geometric dimensions of lens-type piezoelectric crystal

For quartz as a trigonal structure, there exist only two coefficients of linear thermal expansion: $\alpha_x = \alpha_y$, and α_z with $\alpha_z < \alpha_x$. The value of the coefficient of linear thermal expansion in a crystal depends on these two coefficients and on direction with respect to the crystal axes; it is equal to [6]:

$$\alpha_y = \alpha_z - (\alpha_z - \alpha_x)\cos^2\beta, \quad (10)$$

where β is an angle between the present direction and z axes:

$$\alpha'_y = \alpha_z - (\alpha_z - \alpha_x)\cos^2\beta,$$
$$\alpha'_z = \alpha_z - (\alpha_z - \alpha_x)\sin^2\beta.$$

Coefficients of thermal expansion of quartz crystals along crystallographic axes in the temperature range -30°C to +60°C are nearly constant and equal to [4]:

$$\alpha_x = \alpha_y = 14.9\text{x}10^{-6};\ \alpha_z = 7.81\text{x}10^{-6}. \quad (11)$$

With more precision, α_j can be found with the help of expression (6) using the corresponding coefficients from [4]:

$$\alpha_x^{(1)} = 13.71\text{x}10^{-6}\text{deg}^{-1};$$
$$\alpha_z^{(1)} = 7.48\text{x}10^{-6}\text{deg}^{-1};$$
$$\alpha_x^{(2)} = 6.5\text{x}10^{-9}\text{deg}^{-1}; \quad (12)$$
$$\alpha_z^{(2)} = 2.9\text{x}10^{-9}\text{deg}^{-1};$$
$$\alpha_x^{(3)} = -1.9\text{x}10^{-12}\text{deg}^{-1};$$
$$\alpha_z^{(3)} = -1.5\text{x}10^{-12}\text{deg}^{-1}.$$

After determining $\alpha_h = \alpha_y$ from (10) using (11) or (12), we can easily find the dependence of frequency-determining measure $h(t)$ on temperature:

$$h(t) = h(t_o)(1+\alpha_h).$$

The maximum change in diameter of a piezoelectric crystal with temperature occurs in the x-axis direction

$$\Delta d = d\alpha_x(t-t_o),$$

which means $\Delta d/d = 0.000014(t-t_o)$.
This change is very small, and besides, this size is not frequency-determining; hence, we can consider diameter to be a temperature-independent constant value.

To determine $\rho(t)$ from (7), we use values $T_\rho^{(n)}$ [4]

$$T_\rho^{(1)}\text{x}10^6 = -34.92;\ T_\rho^{(2)}\text{x}10^9 = -16.9;$$
$$T_\rho^{(3)}\text{x}10^{12} = 5.3;$$

and value $\rho = 2.65 g/cm^3$ at $t_o = 20°C$.

Dependencies of the elasticity coefficients involved in expressions (9) on temperature and cut angle have the form [14]:

$$C_{11}^*(t,\beta) = C_{11}(t);$$
$$C_{22}^*(t,\beta) = c^4C_{22}(t)+s^4C_{33}(t) 2c^2s^2[C_{13}(t)+2C_{44}(t)]-4c^3sC_{14}(t);$$
$$C_{55}^*(t,\beta) = c^2C_{44}(t)+s^2C_{66}(t)+2csC_{14}(t);$$
$$C_{66}^*(t,\beta) = c^2C_{66}(t)+s^2C_{44}(t)+2csC_{14}(t);$$
$$C_{12}^*(t,\beta) = c^2C_{12}(t)+s^2C_{13}(t)+2csC_{14}(t);$$

where $c = cos(\beta)$ and $s = sin(\beta)$.

Each coefficient C_{ij} in (13) is determined according to (5) by corresponding temperature coefficients. Values of coefficients C_{ij} and T_{cij} at 25°C [10] with corrected $T_{c12}^{(n)}$ $(n=1,2,3)$ [15] are presented in the table. It is necessary to note that frequency ω_o, which is included in formulas (9) for coefficient calculation, must be relevant to a temperature of 25°C.

Temperature coefficients

i,j	$T_{cij}^{(1)}x10^6$ °C^{-1}	$T_{cij}^{(2)}x10^9$ °C^{-1}	$T_{cij}^{(3)}x10^{12}$ °C^{-1}	C_{ij} Nm^{-2}
11	-48.5	-107	-70	86.75
33	-160	-275	-250	107.2
12	-2637	-2678	-1110	5.95
13	-550	-1150	-750	11.91
44	-177	-216	-216	57.8
66	178	118	21	40.4
14	101	-48	-590	-17.8

After substituting (5)-(7) and (13) into equation of motion (8) and introducing certain modifications into previously developed computational programs [11,13], we can determine eigenvalues and eigenfunctions of piezoelectric crystals axial symmetry and varying thickness with allowances for simultaneous variations in the cut angle of the crystal plate, in its geometry, and in its temperature. This makes possible numerical studies of FTC for an entire spectrum of thickness shear vibrations of rotated Y-cut quartz piezoelectric crystals with axial symmetry and varying thickness. Below, some results are presented, which were obtained by studying the effect of local variations in piezoelectric crystal thickness and cut angle on a resonator FTC with the help of an upgraded computer program.

Figure 3 shows a family of curves - FTC of AT-cut quartz resonators for different angles of piezoelectric crystal orientation; these are calculated according to the developed algorithm using the above relations and the coefficient given in the table. FTC for an infinite plate agree well with similar dependencies calculated by other methods [3]. Figures 4,5 show calculated FTC for AT-cut piezoelectric crystals with a given diameter ($d/h_o = 20$) and a curvature varying from zero (curve 1, a disk-shaped piezoelectric crystal with $h_b/h_o = 1$) to maximum (curve 5, a lens-shaped piezoelectric crystal with $h_b/h_o = 0$).

FTC for a lens-shaped piezoelectric crystal with $R/h_o = 500$ and $\Delta\beta = 0$ is nearly coincident with FTC for an infinite plate with $\Delta\beta = 8'$ (dashed line in Fig.4), which agrees with [4].

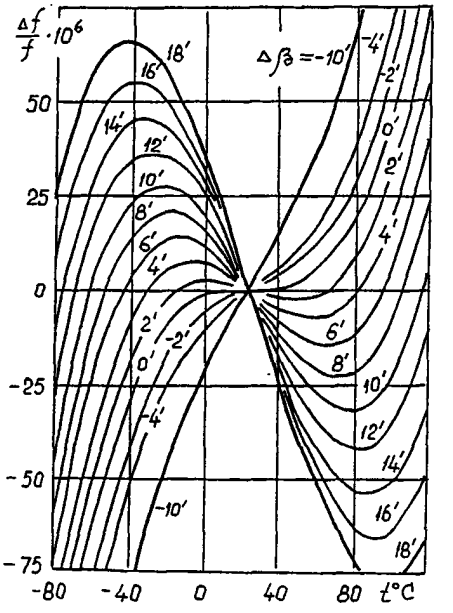


Fig.3. Frequency-temperature characteristics of a piezoelectric crystal with various cut angles $\beta = \beta_o + \Delta\beta$, $\beta_o = 35.0862°$

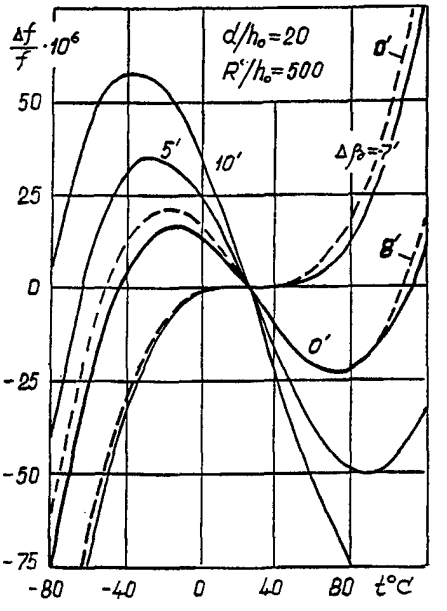


Fig.4. Frequency-temperature characteristics of a biconvex lens-type piezoelectric crystal of AT-cut ($\beta_o = 35.0862°$). Frequency-temperature characteristics for a disk-shaped piezoelectric crystal with the same diameter (dashed line)

In order to obtain a zero temperature coefficient of frequency for the lens-shaped piezoelectric crystal under consideration, we have to change its cut angle by $\Delta\beta=-7'$ (dashed line in Fig.4); at the same time, for a lens-shaped piezoelectric crystal with $R/h_o=200$, the temperature coefficient of frequency will already be zero at $\Delta\beta=13'...1'$. A lens-shaped piezoelectric crystal with $R/h_o = 100$ and $h_b/h_o = 0$ will have a zero temperature coefficient of frequency at the bend point of the characteristic curve when the cut angle changes by $\Delta\beta=-16'$ (Fig.5).

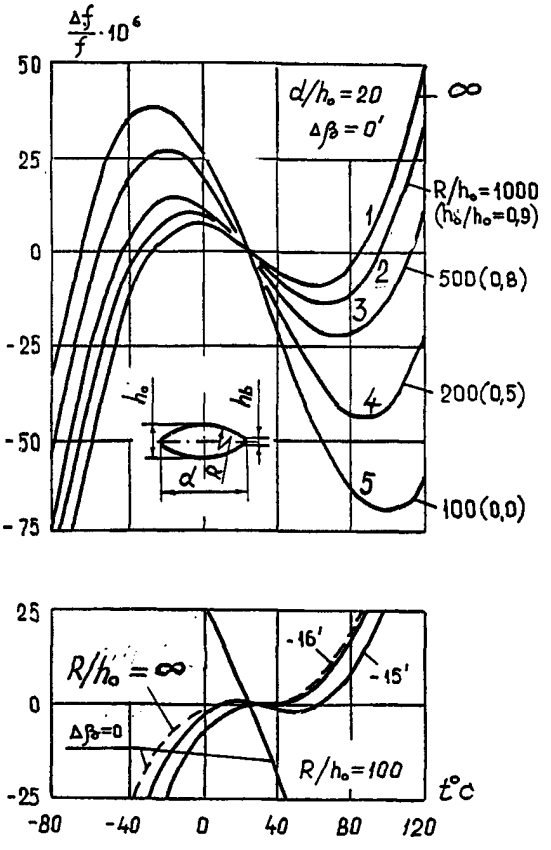


Fig.5. Frequency-temperature characteristics of piezoelectric crystal with different contoured plates and $\beta = 35.0862°$

When analyzing the FTC of the nonharmonic vibrations shown in Fig.6, we must note that the dependencies of their resonance frequencies on temperature are similar to the temperature dependence of the fundamental vibrations, but the magnitude of this effect for nonharmonic vibrations is several times grater.

If we change the cut angle of the plate by $\Delta\beta=-12'$ (β=34.88°), we obtain for nonharmonic vibration 113 a FTC coincident with the FTC of fundamental vibrations at β=35.08° (Fig.6). The results presented agree well with experimental results of FTC measurements for similar vibrations in AT-cut quartz resonators [16].

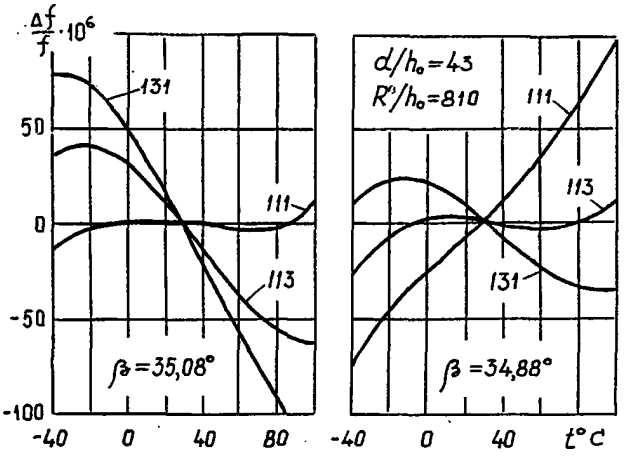


Fig.6. Frequency-temperature characteristics of the fundamental and two nearest nonharmonic vibrations of lens-type piezoelectric crystal

The curves in Fig.7 demonstrate the influence of beveling a disk-shaped piezoelectric crystal on its FTC of resonance vibrations. With the help of these curves, it is possible to make quantitative estimates for correcting the crystal plate cut angle in order to obtain a required temperature dependence.

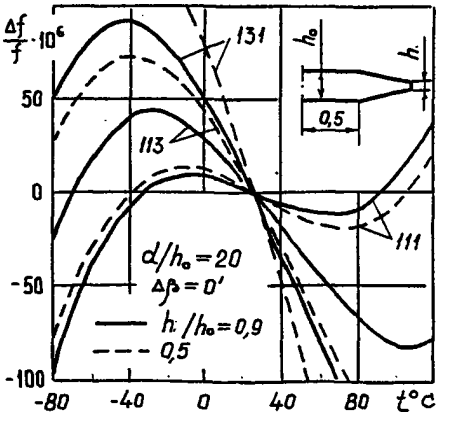


Fig.7. The influence of beveling a disk-shaped piezoelectric crystal on its FTC of resonance vibrations

The developed FTC calculation technique is quite helpful when estimating the effect on these characteristics of local changes in piezoelectric crystal and electrode geometry. A comparison of calculated FTC for a lens-shaped piezoelectric crystal with a groove or with an electrode bulge in the middle or at the edges shows, along with the curves in Fig.8, that local changes in the geometry of a piezoelectric crystal and its electrode are equivalent to a change in cut angle by 1'; rotation of FTC can be done both clockwise and counterclockwise.

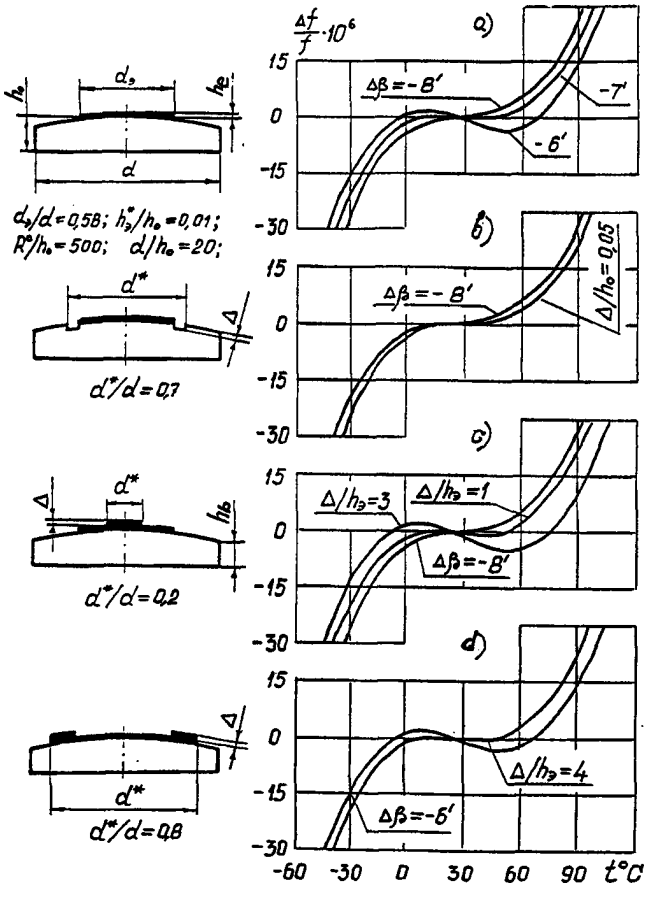


Fig.8. The influence of local changes in the geometry of a piezoelectric crystal and its electrode on FTC of quartz resonator

Thus, the developed calculation technique for FTC of thickness shear vibrations, including nonharmonic vibrations, for resonators with axially symmetric piezoelectric crystals of varying thickness allows estimation of the combined effect of the following factors on a resonator FTC; cut angle of the crystal plate, piezoelectric crystal geometry, and local variations of the thickness of a piezoelectric crystal and its electrode. This last factor is especially important for multimode resonators and high-frequency quartz resonators, which have small piezoelectric crystal thickness and, hence, are highly sensitive to local variations in the geometry of a piezoelectric crystal or electrode.

References

1. Androsova V.G., Bronnikova E.G., Vasil'ev A.M. et.al., Piezoelectric Resonators: A Handbook, Moscow: Radio i Svyaz',1992.

2. Malov V.V., Piezoelectric Resonator Sensors. Moscow: Energiya, 1978.

3. Zelenka J., Piezoelectric Resonators with Bulk and Surface Waves: Materials, Technology, Design, and Application, Moscow: Mir, 1990.

4. Smagin A.G. and Yaroslavskii M.I., Piezoelectricity of Quartz and Quartz Resonators, Moscow: Energia,1970.

5. Smagin A.G., Precision Quartz Resonators, Moscow: Izd. Gos, Kom. Standartov, 1964.

6. Smagin A.G., Piezoelectric Resonators and Their Applications), Moscow: Izd. Gos, Kom. Standartov, 1967.

7. Pekar' I.K., Method of Temperature Measurement and Prospects for Quartz Resonator Application in Thermometry, Obzory Elektron. Tekn., 1970, no. 7(23).

8. Il'ichev V.A., Temperature Dependence of Dynamic Parameters of Quartz Resonators, Elektron. Tekn., Ser.10,1972,no.3, pp.45-53.

9. Pozdnyakov P.G., Calculation of Piezoelectric Crystals with Asymmetric Frequency-Temperature Characteristics, Elektron. Ten., Ser.9, 1968, no.4, p.68.

10. Bechmann R., Ballato A.D., and Lukaszek T.J., Higher Order Temperature Coefficients of the Elastic Stiffness and Compliancies of Alpha-Quartz, Pros. IRE, 1962, vol.50, no.8, pp.1812-1822.

11. Postnikov I.I., Computer Study of High-Frequency Quartz Resonator Characteristics, Radiotekh., 1989, no.12, pp.85-86.

12. Postnikov I.I., Frequency Stability and Signal Formation in Radio and Optical Ranges: Collected Articles, Minsk: Rad. Tech.Inst., 1992.

13. Postnikov I.I., Eigenvalues and Eigenfunctions of Lens-Shaped Piezoelectric Crystals, Acoust. Phys., 1994, vol.40, no.3, pp.381-385.

14. Cady W., Piezoelectricity, New York: McGraw-Hill, 1946.

15. Chernykh G.G., On the Temperature Coefficients of Alpha-Quartz, Electron. Tekn., Ser.10, 193, no. 4, pp.91-93.

16. Morozov E.I. Frequency-Temperature Characteristics of Nonharmonic Vibrations in Quartz Resonators with AT-Cut Convex-Plane Piezoelectric Crystals, Electron. Tekh., Ser.9, 1970, no.2, pp.23-25.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

INTENSITY OF RESONANCE VIBRATION OF AT-CUT QUARTZ RESONATORS WITH CIRCULAR ELECTRODES

I.I. Postnikov, T.V.Christophorova

Special Research Bureau of
Moscow Power Engineering Institute,
Krasnokazarmennaya, 14, Moscow, Russia, 111250

Abstract

There has been obtained analytical expressions for estimation of dynamic inductance and capacitance as well as the intensity of resonance vibrations of AT-cut lens-type piezoelectric element resonators with the complicated axial symmetric configurations of electrodes enabling to determine the frequency characteristics of quartz resonators and the above-listed parameters for both harmonic and nonharmonic vibrations. Here I will be presented the results of theoretical researches of the dependence of the given parameters of quartz resonators on the forms and dimensions of the crystal.

Inroduction

When designing of both monomode and multimode piezoelectric resonators it is necessary to estimate the intensity of resonance vibrations as well as to know the admissible range of values of their dynamic inductance and capacitance for the working vibrations which can be overlapped when changing the geometric dimensions of the piezoelectric crystal and the electrode. All studies on calculation of a dynamic inductance (capacitance) of quartz resonators with harmonic vibrations of lens-type electrodes which as a rule have circular electrodes [1-4] do not allow to estimate these dynamic parameters of quartz resonators for nonharmonic vibrations and for configuration of electrodes different from circular one. Therefore designing of multimode resonators and scopes for improving of features and parameters of monomode resonators are impeded.

Analytical expressions obtained [5] for determination of values of the eigenfunctions of the lens-type electrodes will allow to obtain analytical expressions for these calculations as it will be shown below and partly will fill the a.m. gap in calculations for AT-cut lens-type quartz resonators.

An equivalent dynamic capacitance value (C) of a quartz crystal resonator on resonance frequencies can be defined as follows [6,7]:

$$C_{nmp}=(\int_{Se} D_{nmp}\, dS_e)^2/\rho_\kappa\, \omega^2_{nmp} \int_V U^2_{nmp}\, dV, \quad (1)$$

Then, with regards to Thompson's equation dynamic inductance value (L) can be determined by the expression:

$$L_{nmp} = 1/\omega_{nmp}{}^2 C_{nmp}, \quad (2)$$

where D is a projection of electric induction vector on the normal to the surface of piezoelectric element; U is a component of a vector displacement along electric axes; ρ_κ is a quartz density; ω is an angular frequency; *nmp* are mode indexes (here below the mode indexes will be omitted, but they should be implied).

In the equation (1) the integral in the numerator is taken over the surface of electrode (S_e), in the denominator the integral is taken over the piezoelectric crystal volume (see Fig.1). The electric induction at thickness-shear vibration (is directly dependent on) (is a straight function of) deformation (S_6):

$$D = e_e S_6,$$

where $e_e = k_e e_{26}$ is an effective electric constant; e_{26} is a piezoelectric coefficient; k_e is a correction coefficient.

To determine the deformation S_6, we use a value of displacement (U) at the surface of piezoelectric crystal [8]:

$$U_x(x,y,z) = y\Psi_x(x,z)$$

0-7803-2500-1/95 $4.00 © 1995 IEEE

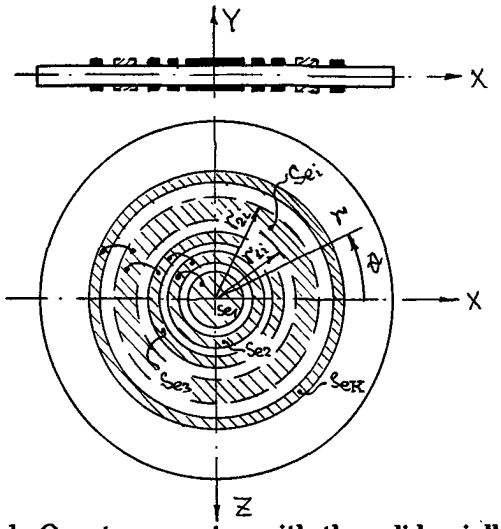


Fig.1. Quartz resonator with the solid axially symmetric electrodes

Then, the deformation S_6 is equal:

$$S_6 = \delta U_x(x,y,z)/\delta y = y\Psi_x(x,z).$$

The integral value in the numerator of the Eq.1 with the account of the displacement function Ψ_x from [9]

$$\Psi_x(r,\vartheta)=P_o(r)+ \sum_{m=1}^{M}[P_m(r)\cos(m\vartheta)+Q_m(r)\sin(m\vartheta)]$$

$$M = 1.2.3,...$$

will then take the form:

$$D = \int_{Se} DdS_e = e_e \int_{Se} \Psi_x(x,z)S_e = e_e \sum_{i=1}^{K} \int_{Sei} \{P_o(r)+ \sum_{m=1}^{K}[P_m(r)\cos(m\vartheta)+Q_m(r)\sin(m\vartheta)]\}dS_e, \quad (3)$$

where k is the number of elementary lands areas into which the area of a composite electrode is split.

Resonators with axial symmetric electrodes (circle, ring, circle plus ring, ring plus a number of rings, rings set and etc., see Fig.1) harmonic (*nll*) and nonharmonic (*nmp*) vibrations with the odd indexes *m&p* which are the vibrations of the first group [5,9] are the most active. So, for axial symmetric electrodes from Eq.(3), we obtain:

$$\int_{Se} DdS_e = 2\pi\, e_e\, R^2_p h^2_o F_s,$$

$$F_s = \sum_{i=1}^{K} \int_{r11}^{r21} \bar{P}_o(r)\bar{r}d\bar{r}, \quad (4)$$

$$R_p = d'/2h_o; \quad \bar{r} = 2r/d'; \quad H = h/h_o.$$

$$\sum_{i=1}^{K} \int_{o}^{2\pi} \int_{r11}^{r21} [P_m(r)\cos(m\vartheta)+ +Q_m(r)\sin(m\vartheta)]rdrd\vartheta = 0, \quad (5)$$

as each integral over the angle ϑ from functions presented in (5) is equal to zero.

The integral in the denominator (Eq.1) is taken over the whole volume of the piezoelectric crystal and for the lens-type quartz resonators (Fig.2) is the following:

$$\int_v U^2 dV = \int_{-h/2}^{h/2} \int_0^1 \int_0^{2\pi} y^2\{\bar{P}_o(\bar{r})+ \sum_{m=1}^{M}[\bar{P}_m(\bar{r})\cos(m\vartheta)+ +\bar{Q}_m(\bar{r})\sin(m\vartheta)]\}^2\bar{r}d\bar{r}dyd\vartheta =$$

$$= (R^2_p h^5_o \pi/12)\cdot F_v \quad (6)$$

$$F_v = \int_0^1 [2\bar{P}^2_o(\bar{r})+ \sum_{m=1}^{M}\bar{P}^2_{2m}(\bar{r})]H^3\bar{r}d\bar{r}.$$

where the function F_v is given only for the first group vibrations.

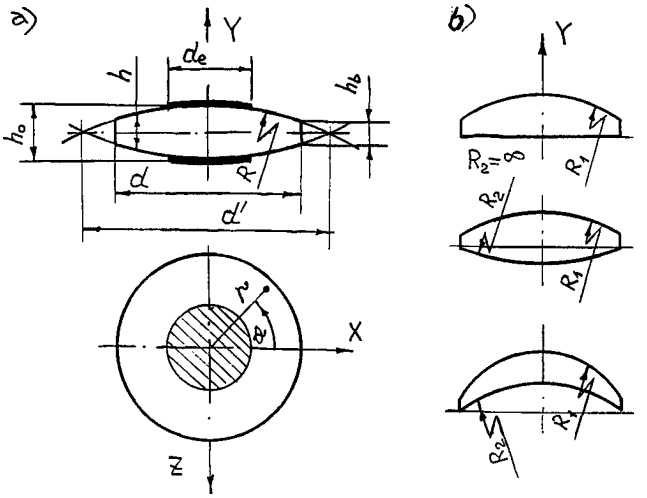


Fig.2. Quartz resonator with lens-type piezoelectric crystal: a) biconvex piezoelectric crystal with circular electrode; b) piezoelectric crystal with varying curvature surface

The value of maximum possible diameter of the lens-type piezoelectric element (d') presented in (4) at the given values of R and h_0 is easily found from the right angle triangle ABC (see Fig.3a):

$$(O'B)^2 = AO' \cdot O'C,$$

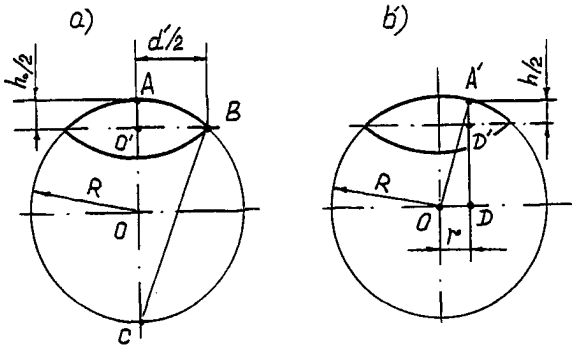


Fig.3. Dimensions' proportions for lens-type piezoelectric crystal

For a biconvex (b/c) and a plane (p/c) convex lens-type piezoelectric element we will obtain the following values:

$$d'_{b/c} = \sqrt{4R_{b/c}h_0 - h_0^2};$$
$$d'_{p/c} = \sqrt{8R_{p/c}h_0 - 4h_0^2};$$

where equating $d'_{b/c} = d'_{p/c}$ one can write:

$$2R_{p/c} / R_{b/c} = 1+3/4(h_0/R_{b/c})$$

which at $(h_0/R_{b/c}) \ll 1$ is always fulfilled, thus one can get the equivalence condition of plane convex and biconvex lens-type electrode:

$$R_{b/c} = R_{p/c}$$

further we will consider parameters of quartz resonators with reference to double convex lens-type piezoelectric crystal, which can also be applied to the quartz resonators with the plane convex lens-type piezoelectric crystal. In case of lens-type piezoelectric crystal with varying radiuses of surface curvature (Fig.2b) the equivalent radius of curvature of biconvex lens-type piezoelectric crystal will be equal in general to:

$$R = 2\,|R_1 \cdot R_2/(R_1+R_2)\,| \qquad (7)$$

The expression (7) will cover a "lentil" type piezoelectric crystal (Fig.2b), when one of the radiuses (R_1 or R_2) is negative, so Eq.7 is taken as the modulo. After substituting (4) and (6) into the Eq.(1) and than into the Eq.(2) we write:

$$L_{nmp} = k_o F_v^{nmp}/(F_s^{nmp})^2;$$
$$C_{nmp} = (F_s^{nmp})^2/k_o F_v^{nmp}\omega^2_{nmp}; \qquad (8)$$
$$k_o = \rho_к h_0/48e_e^2R_p^2\pi,$$

Coefficient k_o in (7) depends on geometrical dimensions and material of piezoelectric crystal and is constant for all modes, at the same time function F_v depends on displacement distribution and for a specified vibration nmp it is a constant value. Function F_s, in its turn, depends not only on the type of vibration but also on geometrical dimensions of the electrode and can have for nonharmonic vibrations even zero values. This is the function which defines both qualitative and quantitative dependence of equivalent motional inductance and capacitance of quartz resonators from forms and dimensions of electrodes and the intensity of vibration.

For axially symmetric piezoelectric crystal the values of functions in (8) can be easily calculated at the specified values of functions p_i by a computer. The program text for computer calculation of equivalent motional inductance and capacitance of resonance vibrations of contoured axially symmetric piezoelectric crystal with variable electrode configurations and the influence of mass on the dynamic parameters (L,C) are written in FORTRAN and presented in the book [9]. Fig.4 depicts experimental [1,10] and calculated values of inductance dependencies of the fundamental vibration and attenuation of the nearest nonharmonic vibration of AT-cut lens-type piezoelectric crystal A_{imp} ($A_{imp} = 20\lg(L_{imp}/L_{iii})$ on electrode diameter which prove the equivalence of experimental and estimated results. Numerous theoretic researches (11-14) made with the help of algorithms and programs for the estimation of the dynamic L and C and capacitance proportion confirm their efficiency and make it possible to replace experimental researches by theoretical experiments.

Detailed researches of the influence of the variations of diameter and mass of the circular electrode on the spectral parameters of the

quartz resonator with axial symmetric electrodes according to the presented above algorithm will be shown in [7, 9]. Nevertheless as shown in [9] features and parameters of the AT-cut lens-type quartz resonator are defined to the major extent on the curvature of piezoelectric crystal surface and in unimportant extent depend on the electrode mass. That is why for their evaluation there should by defined the lien between L, C and capacitance proportion with the dimensions of the piezoelectric crystal and the electrode. which can be easily done taking into account the results of formula (5) where for the resonance vibrations of the first group there have been obtained analytic results of the dependence of P_1 on the geometrical dimensions of lens-type piezoelectric crystal. Taking into account that F_v does not depend on the form of the electrode and is a constant value for the given mode with fixed dimensions of the piezoelectric crystal it looks reasonable to consider the relative values of L and C for estimating of the influence of the electrode forms and dimensions on the dynamic L and C:

$$\bar{L} = L/L^* ; \qquad \bar{C} = C/C^* ; \qquad (9)$$

where $\bar{L}$, $\bar{C}$ are normalized values of L and C; L^*, C^* - L and C values when the electrode covers the piezoelectric crystal completely (d_e=d).

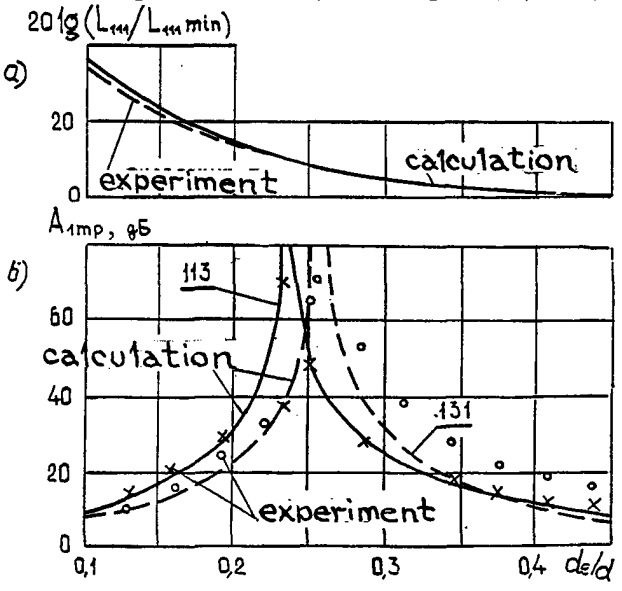


Fig.4.Dependance of the dynamic inductance of the fundamental vibration (a) and attenuation of nonharmonic vibrations (b) on diameter of electrode of lens-type resonator: R_s=370, R_d=15(R_d=d/2h_o, L_{iiimin} - inductance at d_e=d).

Taking into account (8) Eq.(9) will look as follows:

$$\bar{L} = F^2_{sc}/F^2_{sv}, \quad \bar{C} = F^2_{sv}/F^2_{sc}, \qquad (10)$$

where

$$F_{sv}=\sum_{I=1}^{k}\int_{r11}^{r21}\rho_o(r)rdr, \quad F_{sc}=\int_{o}^{1}\rho_o(r)rdr. \qquad (11)$$

In this case when considering the influence of the forms and dimensions of the axially symmetric electrodes on dynamic L and C it will be enough to know only the dependence of the function P_o on the radius of the electrode which makes the task less complicated. Generalized dependencies of the function P_o on the radius of the sphere of piezoelectric crystal and the radius of electrode for the resonance vibrations (till the nonharmonic *n51*) presented in this paper [5] will look as follows:

$$P_o(r)=(\chi_o\lambda+\xi_o)a_1\cdot(1-a_2\rho^2+a_3\rho^4)\exp(-\alpha\rho^2), \qquad (12)$$

The estimated radius of piezoelectric crystal (ρ) for any resonance vibration is connected with the normalized radius (r) and the radius of the sphere of the biconvex (R_s) in the following way:

$$\rho_{nmp} = \sqrt{nr R_s^{\mu 1mp}}, \qquad (13)$$

$\mu_{1mp} = \nu_1^{1mp} + \nu_2^{1mp}\lambda$, $R_s = R/h_o$, $\lambda = \lg R_s$.

Coefficients presented in formulas (12)-(13) for some modes are presented in table 1.

Coefficient	Coefficients for (13)-(14) for the corresponding modes [5]					
	111	113	131	115	133	151
ν_1	0.21	0.159	0.139	0.11055	0.0925	0.0761
ν_2	0.006	0.014	0.017	0.02105	0.0238	0.0268
a_1	1.0	0.8	0.9	1.0	1.0	1.0
a_2	-	5.75	5.62	17.13	18.10	18.10
a_3	-	-	-	37.29	41.82	42.82
α	1.76	2.75	2.85	4.17	4.50	4.60
χ_o	0	0	0	0.1210	-0.0106	-0.023
ξ_o	1	1	1	0.4326	0.7762	0.8697

Eq.(12) will be presented in the following way:

$$\bar{P_o}(\bar{r}) = (b_1 - b_2\bar{r}^2 + b_2\bar{r}^4\, b_2\bar{r}^4)\cdot(\exp-\gamma\bar{r}^2), \qquad (14)$$

where $b_1 = a_1(\chi_o \alpha + \xi_o)$,

$b_2 = a_2 b_1 n R_s^{2\mu}$;

$b_3 = a_3 b_1 n^2 R_s^{4\mu}$;

$\gamma = \alpha n R_s^{2\mu}$.

Inserting P_o form into Eq.(11) and obtaining the values of the integrals we'll have the following values for F_{sc} and F_{vc} from (10):

$$F_{sc}=1/2[(b_1-b_2-b_3)/\gamma-(b_2+2b_3)/\gamma^2- -2b_3/\gamma^3]\cdot\exp(-\gamma)- - b_1/\gamma + b_2/\gamma^2 - 2b_3/\gamma^3,$$

$$F_{sv} = 1/2 \sum_{i=1}^{\kappa} \sum_{j=1}^{2}(-1)^{}\{1/\gamma(b_1-b_2\bar{r}^2_{ji} - b_3\bar{r}^4_{ji})- - 1/\gamma^2 (b_2 + 2b_3\bar{r}^2_{ji}) - 2b_3/\gamma^3\}\exp(-\gamma\bar{r}^2_{ji}).$$

In case of a solid circular electrode presented at Fig.2a under nonharmonic vibrations of the quartz resonator:

$$\bar{C} = [\exp(-\gamma\bar{r}_e^2)-1]^2/\exp(-\gamma)-1]^2;$$

$$\bar{L} = [\exp(-\gamma)-1]^2/\exp(-\gamma\,\bar{r}_e^2)-1]^2;$$

Lets show the efficiency of the presented above analytic expressions when considering the dynamic parameters of resonators with axially symmetric electrodes. Fig. 5 presents the dependencies of the changes of inductance of the main vibration and attenuation of several vibrations of the first group on the diameter of electrode for the biconvex lens-type piezoelectric element: R_s=370; R_d=15, showing the character of changes of the resonance vibrations' intensity with different electrode diameters. With the diameter close to the diameter of the piezoelectric element ($d_e/d \to 1$) the intensity of nonharmonic vibrations which is in verse proportion to the value of dynamic inductance will be the maximum (Fig.5).

As shown in Fig.5 with d_0/d=0,27 not only nonharmonic vibrations 113, 131 (the cross point on the lines L_{113} and L_{131}) will be attenuated but attenuation will effect other unharmonics as well. This is the condition of obtaining the mono-mode resonators as unharmonics 113, 131 are the next to the main vibration [9] and they are the first to be attenuated. This electrode diameter is the "optimal". Flactuation of the diameter in any way from the optimal will increase the intencity of one of the nonharmonic vibration. (Fig.5).

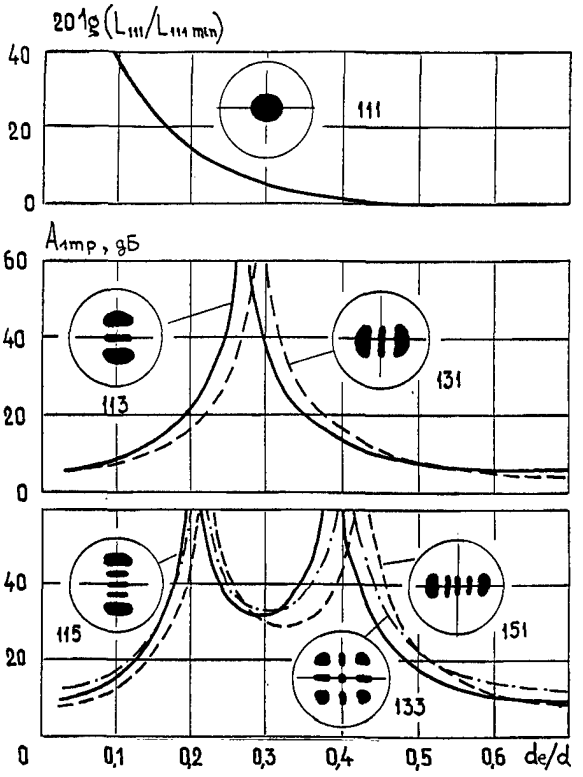


Fig.5.Dependencies of the attenuation of vibrations with nodes of displacement in the center of piezoelectric element on the diameter of electrode

In its turn simultaneous maximum attenuation of the unwanted vibrations 113, 131 has a limit as well. That is why in practice for improving the monomode of teh quartz resonator there will be use for example [15-18] various forms of active electrodes, passive electrode systems forming additional resonators , etc. Thus in the construction of the piezo element presented in [17] in order to increase attenuation of the unwanted vibrations the piezo element is covered with the system of additional electrodes which as distinct from the electrodes of the main resonaor are shirt circuited. But this construction of electrodes is effective for attenuation of vibrations with the maximum displacement of the elements of the PE surface being in the zone of low amlitudes of displacement of the main vibration. Otherwise with attenuation of the unwanted vibration there will take place a

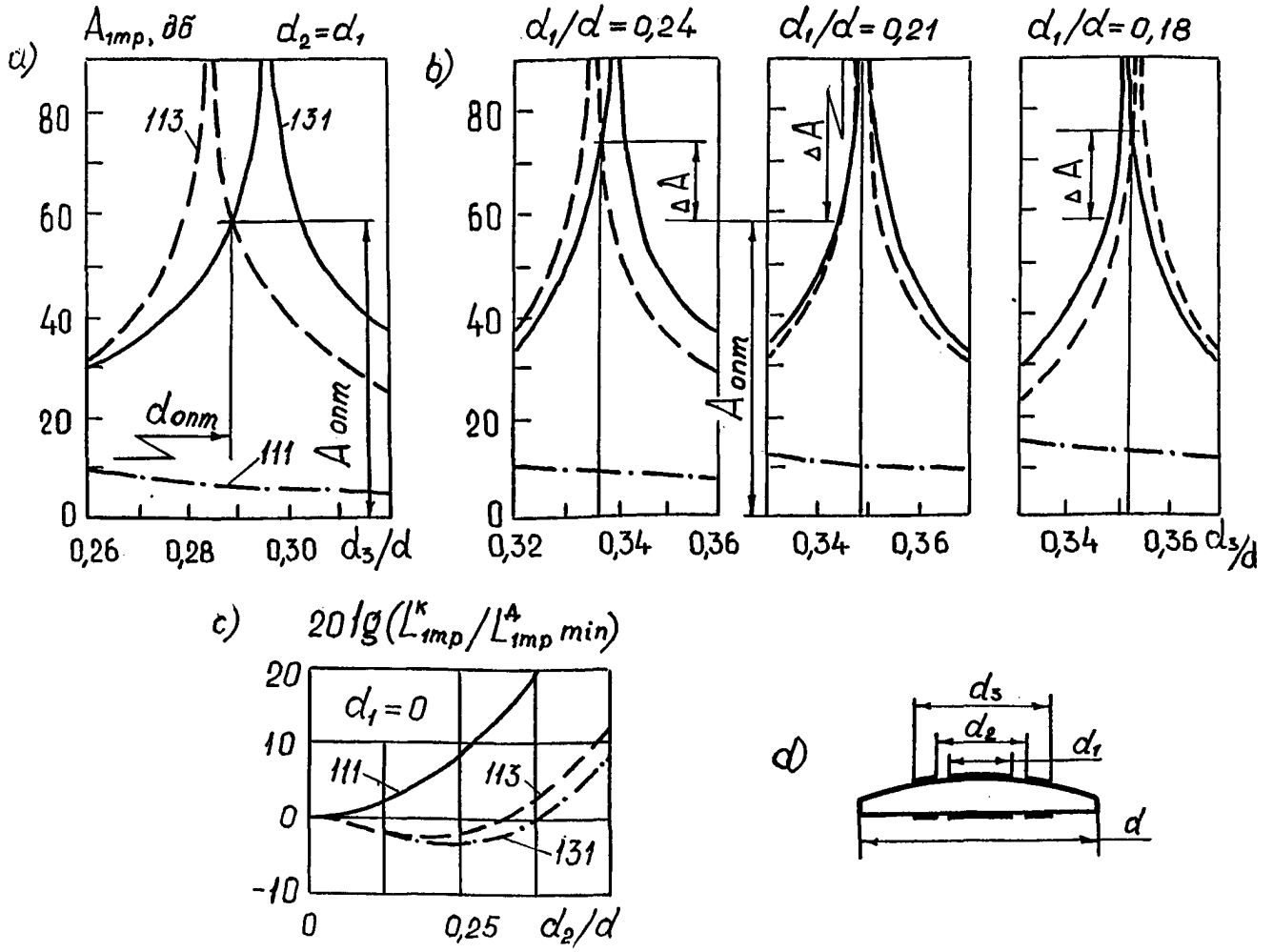


Fig.6. Dependencies of the attenuation of nonharmonic vibrations (a,b) and dynamic inductance (c) on geometric dimensions of complicated axially symmetric electrode.

sufficient attenuation of the fundamental resonance. Thus for the disk shaped piezoelectric element with the optimal dimentions of the circular electrode (by energy consumption) the amplitude of displacement of the fundamental vibration in the zones of the extreme values of displacements for unharmonics 113,131 will be 0,38 and 0,39 adequate form the maximum value [9].

We'll show that using the axially symmetric electrode in the form of electrically connected circle and ring there may be achieved a noticeable attenuation of the unwanted resonances while preserving the sufficient intensity of the fundamental vibrations. Results of the calculations A_{imp} for various compositions of the electrode dimensions are presented in Fig.6. Due to the fact that the ring shaped surface of the piezoelectric element between the diameters d_1 and d_2 (Fig.6c) does not participate in the emergence of the piezo-electic impulse on the active electrode it is possible to achieve simultaneous zero impulse for both vibrations. Thus for fixed values $d_2/d=0,3$ when changing values d_1/d_2 presented at Fig.6b dependencies A_{113}, A_{131} displace against each other and coincidence of their extremes (for example $d_1/d_2=0.21$ *and* $d_3/d=0.348$) there will appear the maximum simultaneous attenuation of the given nonharmonic vibrations which for dozens dB will exceed the one which may appear in case of the solid circular electrode (Fig.6a) and at the same time with the given electrode form the fundamental vibration will attenuate noticeably (Fig.6b- dash-dotted line).

Serious interest for the designers and consumers will be awaken by circular electrodes

not only due to the simplicity of analytic expressions in this case are rather simple but as well by the simplicity of production due to absence of the necessity of orienting the piezoelectric plate when pasting the electrodes .

For resonators with such electrodes under nonharmonic vibrations 113, 131 one may obtain lower inductancies than for resonators with solid circular electrodes (Fig.6c, L^{d}_{imp},L^{k}_{imp} - inductancies with solid circular and ring shaped electrodes relatively: $R_s = 100$, $R_d = 4$). Fig.6c also shows that with given absolute values of inductancies of the considered modes it is possible to define easily the value of the internal radius of the ring shaped electrode under which all three vibrations will have similar inductance.

Thus analytic formulas presented hereabove for defining of the attenuation of the resonance vibrations and dependencies of inductancies and capacitancies of a lens-type piezo-electric quartz resonators on dimensions and forms of electrodes' make it possible to analyze the a.m. parameters of resonators under harmonic and nonharmonic resonance vibrations which is important for developing of both monomode and multimode resonators.

References

1. Mindlin R.D. "High frequency vibrations of crystal plates", Quarterly of Applied Mathematics, Vol.XIX, No.1.- pp. 51-61, 1961

2. Tiersten H.F., Mindlin R.D. "Forced vibrations of piezoelectric crystal plates", Quart. Appl. Math., V.XX, No.2, pp.107-119, 1962.

3. Mindlin R.D.,Medic M.A. "Extensional vibrations of elastic plates", J.Appl. Mech. Trans. ASME, V.26, P.561,1959

4. Holland R., EerNisse E.P. "Design of resonant piezoelectric devices", The M.I.T.Press, Cambridge, Massachussetts, London, 1979.

5. Mindlin R.D., Spencer W.J. "Anharmonic Twist Overtones of Thickness - Shear and Flexural Vibrations of Rectangular AT-Cut Quartz Plates", J.Acoust.Soc. Am., V.42, No.6, pp. 1268-1277, 1967.

6. Bleustein J.L.,Tiersten H.F. "Forced thickness-shear vibrations of discon-, tinuously plated piezoelectric plates", J.Acoust.Soc.Amer., V.43, No.3, pp. 1311-1318, 1968.

7. Lee P.C., Spencer W.J. "Shear flexural-twist vibrations in rectangular AT-cut quartz plates with partial electrodes", J.Acoust.Soc.of Amer., V.45, No.3, pP.637-645, 1968.

8. Mason W.P., Thurston R.N. "A compact electromechanical band-pass filter for frequencies below 20 kilocycles",IRE Trans. Ultrason. Engineering. V.UE-7., p.59, 1960.

9. Nakamura K., Skimizu H. "Analysis of two-dimensional energy trapping in piezoelectric plates with rectangular electrodes", Proc. IEEE Ultrasonic Symposium.,P.606,1976.

10. Sykes R. Bell. Syst. Tech.J., V.23, pp.52-96, 1944.

11. Mindlin R.D. "High frequency vibrations of plated crystal plates", Progress in Appl. Mechanics, McMillan, New York, pp. 73-84, 1961.

12. Byrne R.J., Lloyd P., Spencer W.J. "Thickness- shear vibration in rectangular AT-cut quartz plates with partial electrodes", J.Acoust. Soc. Amer., V.43, P.231, 1968

13. Mindlin R.D. "High frequency vibrations of piezoelectric crystal plates", Report, Columbia University, Departament of Civil Engineering and Engineering Mechanics.-New York.-1972.

14. Zelenka J. "The influence of the electrodes on the resonance frequency of AT-cut quartz plates", Int.J.Solids Structure, V. 11.- pp. 871-876, 1975.

15. Kantor V.M."Monolithic piezoelectric filters". Publ. House "Swjaz", 1977 (in Russian).

16. Nakamura K., Skimizu H. "Analysis of two-dimensional energy trapping in piezoelectric plates with rectangular electrodes", Proc.IEEE Ultrasonic Symposium., P.606, 1976.

17. Bechmann R. Patent USA N 49933,1941.

18. Mortley W.S.Proc.IEE., V.B104,P.239, 1956.

19. Mindlin R.D., Lee P.C.Y. "Thickness-Shear and Flexural Vibrations of Partially Plated crystal Plates", Int.J.Solids and Structures, V.2, No.1, P.125-139, 1966.

20. Shockley W., Curran D.R., Koneval P.T. "Trapped enargy modes in quartz filter crystals", J.Acoust.Soc.Amer., V.41, N.4, pt.2, pp.981-983, 1967.

21. Mindlin R.D. "Optimum size and shapes of electrodes for quartz resonators", J. Acoust. Soc. Amer., V.43, pp.1329-1331, 1968.

22. Wilson C.J. "Vibration modes of AT-Cut Convex Quartz Resonators", J. Phys. D: Appl. Phys., V.7, No.18, pp.2449-2456, 1974.

23. Tiersten H.F., Smythe R.C. "An analysis of contoured crystal resonators operating in overtones of coupled thickness shear and thickness twist", J.Acoust. Soc. Am., V.65, No 6, pp.1455-1460, 1979.

24. Physical Acoustics. Principles and Method. Edited by Warren P.Mason, Volume Y, Academic Press, New York and London, 1968.

25. Mindlin R.D. "Thickness-Shear and Flexural Vibrations of Crystal Plates", J.Appl.Phys., V.22, No.3, pp.316-323, 1951.

26. Mindlin R.D. "High Frequency Vibrations of Piezoelectric Crystal Plates", Int. J. Solids and Structures., V.8, No.7, pp.895-906, 1972.

27. Samojlov V.S., Lubimov L.A., Postnikov 4I.I. "Three-dimensional theory of contoured resonators in approximation of energy trapping", Electronic Engineering, Ser.5, Radiodetails and Radiocomponents, Issue 4, pp.3-7, 1974 (in Russian).

28. Postnikov I.I., Vesselov G.I., Lubimov L.A., Samojlov V.S. "Resonators with lens-type piezoelectric elements" in the book "Collected articles on scientific works dealing with microelectronics problems"(UHF and measuring techniques), Issue XXY, MIIET 1976, pp.207-216 (in Russian).

29. Vesselov G.I., Postnikov I.I., Samojlov V.S. "On the design of piezoelectric elements of high-frequency quartz crystal units", Radioengineering and electronics, 1981, V.XXXI, No.12, pp.2531-2539 (in Russian).

30. Postnikov I.I. "Coupling of thickness-shear vibrations with flexural ones in high-frequency piezoelectric resonators, Radioengineering, No.10, 1991, p.26 (in Russian).

31. Smagin A.G., Jaroslavsky M.I. "Piezoelectricity of quartz and quartz crystal units", Publ.House "Energija", 1970 (in Russian).

32. Postnikov I.I. "Eigenfrequencies of lens-type piezoelectric elements", Radiophysics, V.29, No.6, 1987, pp.740-747 (in Russian).

33. Postnikov I.I. "Eigenvibration modes of high-frequency piezoelectric resonators", V.XXXII, No.10, 1987, pp.2052-2060 (in Russian).

34. Postnikov I.I. "Eigenfunction of anharmonic vibrations n15, n33, n51 of lens-type piezoelectric elements", Radiophysics, V.33, No.10, 1990, pp.1176-1180 (in Russian).

35. Smirnov V.I."High mathematics course", V.2, Publ.House "Nauka" (Scince) 1974 (in Russian).

36. Bachvalov N.S. "Numerical methods", Publ.House "Scince", 1973 (in Russian).

37. Kalitkin N.N. "Numerical methods", Publ.House "Scince", 1977 (in Russian).

38. Hall G. and Watt J.M. "Modern Numerical Methods for Ordinary Differential Equations", Clareudon Press.Oxford, 1976.

39. Vesselov G.I., Postnikov I.I., Samojlov V.S. "Steady-state algorithm of numerical modelling at electronic computer of eigenfrequencies of HF quartz crystal units", Radioelectronics, V.XXYI, No.6, pp.27-32, 1983 (in Russian).

40. Postnikov I.I. "Computer study of HF quartz crystal unit characteristics", Radioengineering, No.12, pp.485-86, 1989 (in Russian).

41. Vesselov G.I., Postnikov I.I., Samojlov V.S. "Modeling of lens-type piezoelectric elements for HF resonators", "Problem of microelectronics", MIET, pp.133-139, 1979 (in Russian).

42. Postnikov I.I. "Frequency distances up to anharmonic vibrations in lens-type piezoelectric elements", Radioengineering and Electronics, V.XXXIX, No.3, pp.446-448, 1994 (in Russian).

43. Morgalovsky V.P., Postnikov I.I. Simultaneous maximum attenuation of nearest unwanted responses of quartz crystal units", Techniques of Communication Means, Series OT, Issue 1, pp.139-144, 1990 (in Russian).

44. Postnikov I.I. "Generalized dependence of capacitance relation of HF quartz crystal units", Electrocommunication, No.2, pp. 54-56, 1985 (in Russian).

45. Postnikov I.I. "Calculation of capacitance relation of HF resonators with lens-type piezoelectric elements of AT-cut", Radioengineering, No.8, pp.85-87, 1985 (in Russian).

46. Postnikov I.I. "Dependence of capacitance relation on electrode radius anharmonic vibrations of quartz crystal units with lens-type piezoelectric elements", Radioengineering, No.11, pp.58-60, 1986 (in Russian).

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

High Frequency Fundamental Resonators and Filters fabricated by Batch Process using Chemical Etching

O. Ishii, T. Morita, T. Saito and Y. Nakazawa

Toyo Communication Equipment Co.,Ltd.
Kanagawa-pref., Japan

Abstract

The small sized 1st intermediate frequency (IF) filters at center frequency range of 70MHz to 150MHz, passband widths of ± 5 to ± 100kHz with sharp selectivity are required in mobile communication systems such as mobile and portable cellular phone. Our solution for its requirement was to employ fundamental mode monolithic crystal filter (MCF) assembled in surface mountable package.

We describe here in detail, the design approach including batch process etching technology. Through photolithography, 56 patterns are chemically etched on one wafer (25mm × 20mm). Then, a similar etching process automatically adjusts the wafer thickness in accordance with frequency. For the MCF, the frequency of split electrodes and the degree of coupling between them are automatically adjusted by an accurately positioned mask evaporation process controlled by a computer. Further, we describe suppression of the spurious response, technology for realizing wide bandwidth and high stopband attenuation characteristics.

By our above developed technology, we achieved 90MHz miniaturized IF filter with 1 / 15 volume reduction of conventional 3rd overtone mode MCF, still possessing the same characteristic of conventional one.

Also, we achieved 130MHz of middle band with suppression of spurious response in wide stopband frequency range, and 71MHz linear phase wide ± 88kHz band with the group delay distortion 0.9 μ s over f_0 ± 80kHz.

Introduction

New development in mobile communication systems requires higher operating frequencies, and smaller size and lighter weight of equipment. Therefore, resonators and MCFs are required to be miniature models with higher operating frequencies. In resonators and MCFs using the fundamental mode of thickness shear AT-cut crystals with excellent temperature characteristics, obtaining high operating frequencies involves to thin the wafer thickness. With the mechanical restrictions on fabricating, 50MHz has been the upper limit for the fundamental operating frequency at thickness of about 33 μm. Even if it were possible to lap and polish to a thinner wafer, it would be difficult to implement mass production and ensure high reliability. Because it would have insufficient mechanical strength and, as a result, such problems as damage in assembly process, by vibration and mechanical shock. Especially in case of MCF, when we constitutes cascade connection, since the capacitance ratio is large, it is necessary to use inductors. In addition, since the characteristic impedance becomes more than several k Ω , tuning transformers are required at the input and output terminals. It is disadvantageous in terms of miniaturization. Furthermore, there is a problem that spurious response are large and difficult to be suppressed.

In recent years, several studies have suggested that the problems mentioned above may be eliminated by etching technology which employs photolithography and that higher frequencies may be achieved by milling only the central portion of the

0-7803-2500-1/95 $4.00 © 1995 IEEE

wafer to a minimum thickness[1]–[4].

In this paper, we reported that we have accordingly employed a batch processing method with chemical etching that has not been previously attempted, in order to achieve high frequency. We have produced a high frequency fundamental (HFF) resonator and HFF–MCF. We have developed a surface mount device (SMD) type 90MHz narrow band filter, which is compact, and suitable IF filter for portable telephones. In addition, we described that a 130MHz mid band filter and a 71MHz wide band filter were also achieved as IF filters for digital portable telephones.

HFF–MCF structure

In case of thickness shear AT–cut crystal, the thickness is about 18.5 μm for a resonator frequency of 90MHz with the fundamental mode, and 12.8 μm for 130MHz. These are both extremely thin. It is physically not possible to process and maintain a crystal wafer so thin.

Figure 1 shows the structure of a HFF–MCF with which higher frequencies are achieved by inverted mesa only one surface of an AT–cut crystal wafer with a thickness of 80 μm. It is relatively easy to handle a crystal of this thickness, and the high frequency is achieved by etching the central area.

Aluminum with low density was used for the electrodes and grounding electrode was formed by vacuum evaporation on the entire surface of the crystal wafer on the side that had been etched. On the flat surface on the other side, a split electrode was formed through photolithography. The mount was constituted as one end in the inverted mesa and the split electrode is connected to the holder terminal through wire bonding. This made surface mounting possible, which had been extremely difficult in conventional MCFs.

Manufacturing process

Figure 2 is a block diagram showing a simplified process sequence for the HFF–MCF fabrication. The AT–cut crystal wafer to be used for fabrication is a polished wafer (25mm × 20mm × 80 $^{t}\mu$m), cut out from a synthetic quartz which has etch channel density 3 or less of per sq.cm.

The crystal wafer is formed through surface

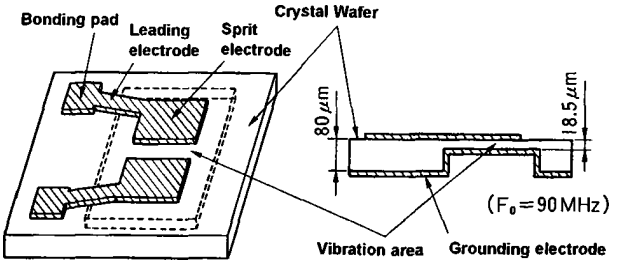


Fig.1 Structure of HFF–MCF.

treatment with a hydrofluoric acid solution in order to completely remove machining distortions on the surface and vacuum evaporation of gold with chrome as a base for enhancing the strength of the bond with the crystal. A crystal etching pattern is formed through photolithography, which makes it possible to form 56 elements at a time on a wafer.

A hydrofluoric acid solution which is heated to approximately 80 ℃ is used during the crystal etching process to etch only the vibration area to the thickness which corresponds to the desired frequency.

All the frequencies of the individual elements in the vibration area on the crystal wafer thus etched are measured, and adjustment is performed using a final etching equipment, which etches only the vibration area through computer control so that all the elements will achieve the target frequency.

After aluminum vacuum evaporation, the electrode pattern for MCF is formed through photolithography on the flat side, which is not etched. The entire surface of the etched side is covered by a grounding electrode. Because of this, an electrode pattern is achieved with higher dimensional accuracy than mask evaporation methods, which have been conventionally used for electrode formation. In addition, deviation of the electrode patterns of the front and rear is eliminated, contributing to accuracy in filter characteristics. Frequency adjustment is performed for the individual elements in which MCF electrode patterns are formed on the crystal wafer show in Figure 3.

This adjustment includes the adjustment of the frequency interval between the symmetric frequency and the antisymmetric frequency, which determines the passband width of the filter characteristics,

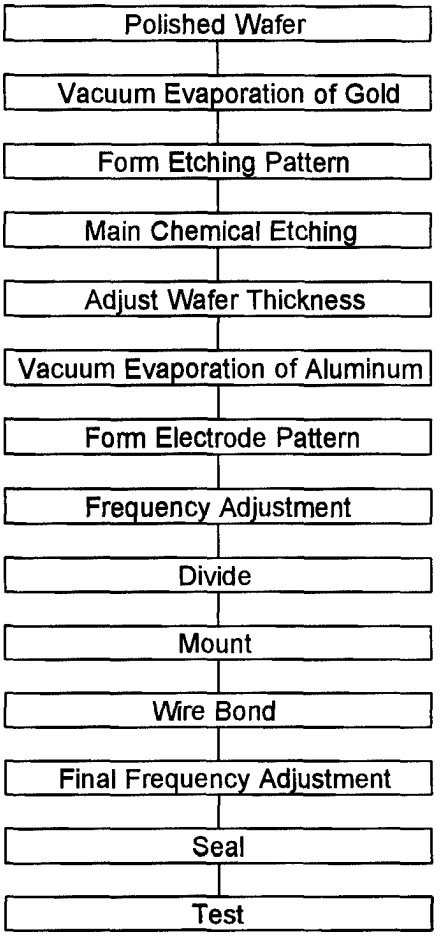


Fig.2 Simplified process sequence.

and the adjustment to achieve the target frequency. It is performed through silver vacuum evaporation on the side whose entire surface is a grounding electrode, which has been etched using the equipment shown in Figure 4. Note that since this equipment performs image processing through automatic computer control of positioning of the electrode pattern and the fine tuning mask, the adjustment can be performed with a high degree of positional accuracy. Moreover, two crystal wafers can be placed within the vacuum chamber for adjustment and the time required for the adjustment is approximately 20 sec per element.

The 56 HFF–MCF elements on the crystal wafer for which the frequency adjustment has been performed are then cut out in segments of approximately 3mm × 3mm and these are mounted in a ceramic package with conductive paste. Then, each of the split electrodes is connected to an input

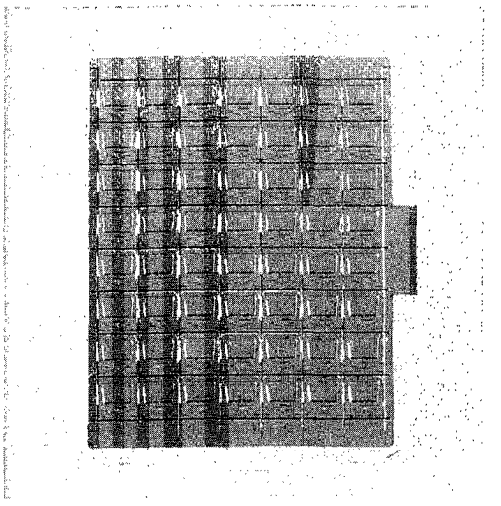


Fig.3 HFF–MCF wafer formed electrode patterns.

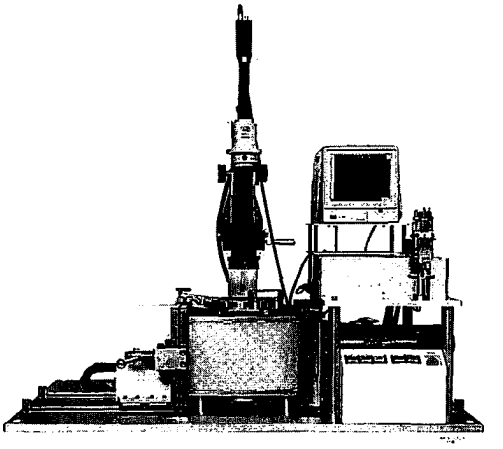


Fig.4 Automatic frequency adjusting equipment.

and output terminal of the ceramic package through wire bonding with aluminum.

The final frequency adjustment is performed from the split electrode side through silver vacuum evaporation. After the adjustment is complete, it is sealed within a nitrogen gas atmosphere, employing the seam welding method, which offers a high degree of reliability since the temperature at the time of sealing does not affect the elements.

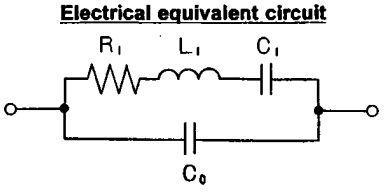
HFF resonator

A MCF is constituted by providing a plurality of electrodes (two electrodes in the case of 2-pole) which are in close proximity to each other on a crystal wafer so as to take advantage of acoustic coupling among the resonators. It means that the filter characteristics are determined by the characteristics of the resonator at each electrode. In order to design HFF-MCF, we first made a prototype sample HFF resonator provided with one electrode and synthetic quartz with its Q value at 2×10^6 was used and the measurement was performed with the external dimension of the resonator at approximately 3mm × 3mm, the vibration area dimensions at approximately 2mm × 2mm, the electrode area dimensions at approximately 0.3mm × 0.3mm and the electrode film thickness at 1,000 ~ 3,000 Å with the resonance frequency ranging from 80MHz to 310MHz. The constants of the equivalent circuits of the prototypes with resonance frequencies at 90MHz, 220MHz and 310MHz are shown in Table 1.

The capacitance ratio γ (C_0 / C_1), which restricts the passband width of the filter characteristics in the equivalent circuit constant is 380 ~ 500, is 1.5 ~ 2 times larger than that of a fundamental mode resonator in the regular 10 ~ 30MHz (approximately 250). Approximately 0.2% can be expected as the maximum relative bandwidth.

Table 1 Measured circuit equivalent values of HFF-resonators.

f_s (MHz)	R_1 (Ω)	L_1 (μH)	C_1 (fF)	C_0 (pF)	Q	γ
90.0	9.7	832	3.51	1.80	50,300	484
220.0	10.5	192	2.81	1.38	24,900	491
310.0	9.4	57	4.64	1.76	11,800	380



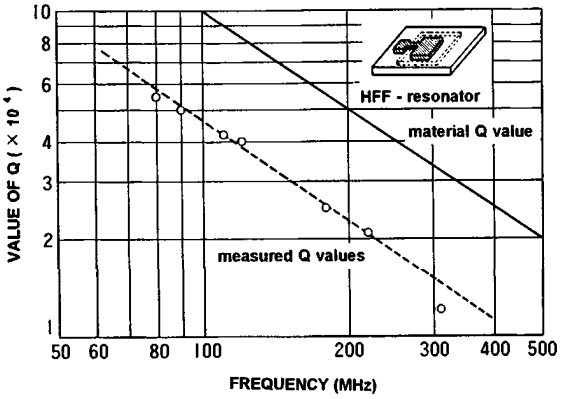


Fig.5 Frequency vs. measured Q values of HFF-resonators.

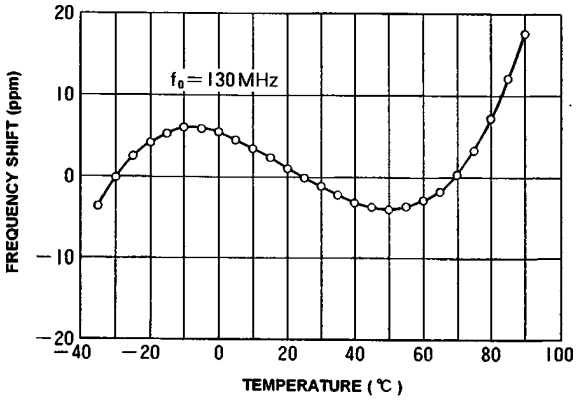


Fig.6 Frequency temperature characteristics of HFF-resonator.

The relationships between the resonance frequencies and the measured Q values of the resonators are shown in Figure 5. Also, the material Q value of the synthetic quartz shown in Figure 5. The Q value of the resonator is approximately 40% ~ 50% of the material Q value. This value is sufficiently high to achieve a small insertion loss when filters are connected in cascade to improve stopband attenuation characteristics.

Figure 6 shows the frequency temperature characteristics of the HFF resonator in the resonance frequency of 130MHz. The Figure 6 shows that the cubic curve characteristics particular to the AT-cut resonator are achieved. Thus, we were able to verify that the fabricating method employing etching technology does not present any problems.

Features of HFF–MCF

HFF–MCFs use fundamental mode, so they have a number of advantages,

(1) The capacitance ratio is small, a wide passband width can be achieved compared with overtone mode MCFs.

(2) Inductors are not required for interstage connection in cascade, a smaller crystal unit is achieved with no inductors.

(3) Low impedance facilitates termination and high attenuation is achieved with one section.

(4) Suppression of spurious responses is facilitated.

Figure 7 shows the relationships between the operating frequency and the relative bandwidth in various filters. The application range of the HFF–MCF has a center frequency of 50 ~ 150MHz and a passband width of ± 5 ~± 100kHz. The HFF–MCF covers the areas where the frequency is high with conventional MCFs, and in which miniaturization is difficult with conventional SAW filters, and achieves a stable, compact filter.

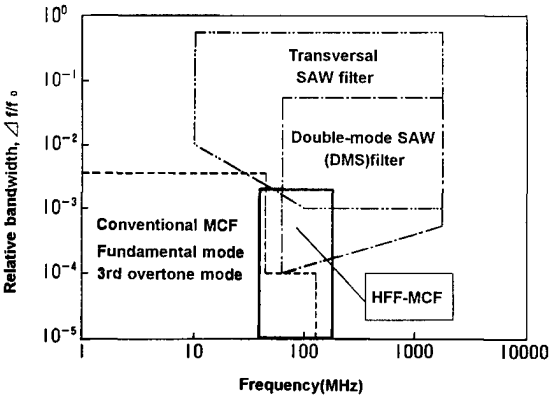


Fig.7 Realization range by means of HFF–MCF.

Application for portable telephone IF filters

Lower prices and further miniaturization of compact portable telephones, which are currently typified by "Micro TAC" and "Mova" will continue to be pursued in the future. The following examples demonstrate the application of this HFF–MCF, which is suited for mass production and advantageous for miniaturization, as an IF filter for portable telephones in various systems from a narrow band through wide band.

90MHz narrow band IF filter

A 90MHz 4–pole (two section configuration) MCF that operates on a 3rd overtone mode is used for IF filters for Japan's analog portable telephones that are currently in use domestically. In order to achieve miniaturization, we reviewed a 2–pole (one section configuration) HFF–MCF. The passband characteristics and the group delay characteristics of a 90MHz 2–pole HFF–MCF are shown in Figure 8 and its stopband attenuation characteristics are shown in Figure 9.

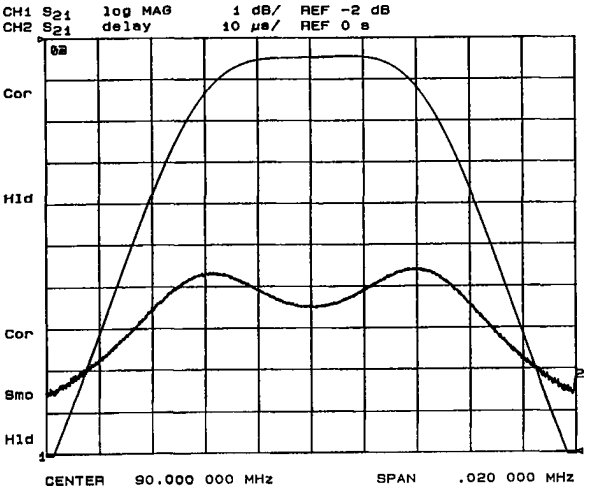


Fig.8 Passband characteristics and group delay characteristics of 90MHz 2–pole HFF–MCF.

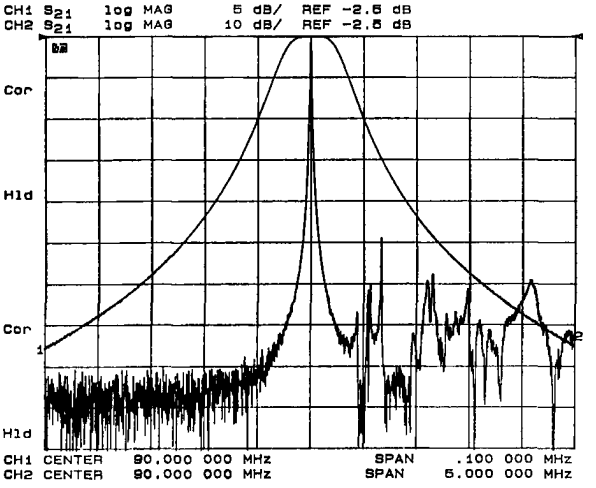


Fig.9 Stopband attenuation characteristics of 90MHz 2–pole HFF–MCF.

At the center frequency of 90MHz, the MCF offers a 2.5dB insertion loss, a 3dB passband width of ± 6kHz and a group delay distortion of 8 μ s at $f_0 \pm$ 4kHz, a stopband attenuation of 13dB for $f_0 \pm$ 12.5kHz, 35dB for $f_0 \pm$ 50kHz, and more than 80dB for $f_0 -$ 910kHz.

Here, the terminal impedance was as low as 150 Ω // 9.5pF for input and output. In addition, by using the 2-pole HFF-MCF that was mounted in a SMD type ceramic package, the dimensions of which were 7mm × 5mm × 1.5 ^{t} mm, the Q value reached approximately 40,000. Furthermore, in comparison with a SMD type MCF (18.5mm × 10.5mm × 4.0 ^{t} mm) with an internal matching circuit using a conventional 3rd overtone mode, an extremely thin profile was realized for miniaturization, with a volume ratio of approximately 1 / 15. Figure 10 shows an external view of the unit in contrast with the conventional product[5].

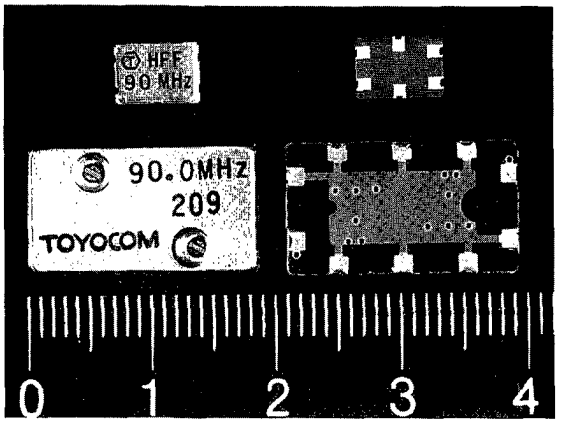
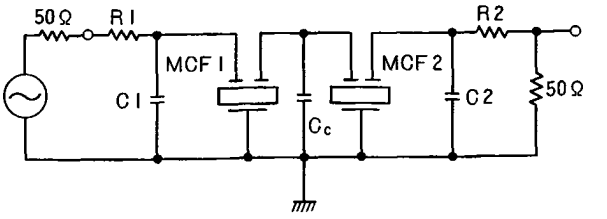


Fig.10 Photograph of 90MHz 2-pole HFF-MCF and the conventional product.

130MHz middle band IF filter

A 130MHz filter with a bandwidth of ± 10.5kHz is required for an IF filter for digital portable telephones to be used in the sub-microwave band. We produced of such a filter in a 4-pole configuration in which two 2-pole HFF-MCF elements are connected in cascade. Its filter configuration and the measuring circuit are shown in Figure 11. Its passband characteristics and group delay characteristics are shown in Figure 12 and its stopband attenuation characteristics are shown in Figure 13.



R1=R2=180Ω, C1=C2=3.0pF, C_c=13.0pF

Fig.11 Circuit configuration and measured circuit of 130MHz 4-pole HFF-MCF.

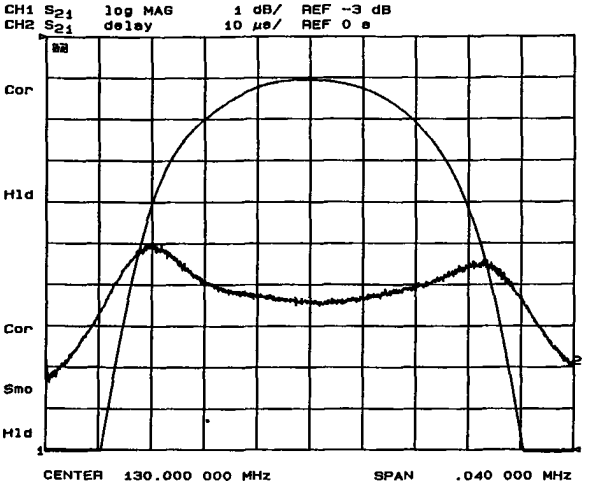


Fig.12 Passband characteristics and group delay characteristics of 130MHz 4-pole HFF-MCF.

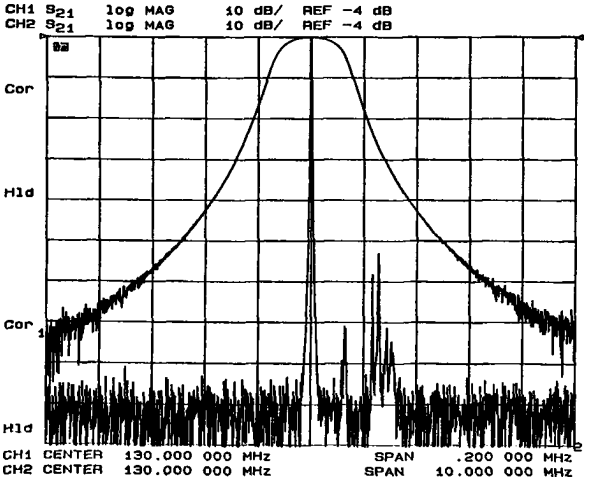


Fig.13 Stopband attenuation characteristics of 130MHz 4-pole HFF-MCF.

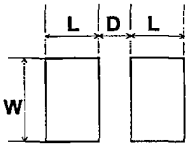
The results show that the characteristics at the center frequency of 130MHz, the product attains a 4.1dB insertion loss, a 3dB passband width of ± 11.9kHz and a group delay distortion of 10 μ s at $f_0 \pm$ 10.5kHz. Note that since the terminal impedance is 230 Ω // 3pF and the interstage capacitance (Cc) is 13pF, the impedance is low and capacitive. Therefore, it does not require an inductors and this facilitates miniaturization and weight lightening.

Moreover, attenuation values of more than 85dB is assured in the lower side stopband, while in the higher side, due to the generation of spurious responses, attenuation is about 55dB. Since an analog type portable phone only requires a certain level of attenuation in the lower side, this does not present any problems. However, it is expected that in the future, attenuation values of more than 80dB will be required in the both side for digital type portable phones.

So, we examined methods for suppressing spurious responses generated in the higher side stopband. The method which is normally used to suppress spurious responses in MCF involves determining the dimensions of the electrode and the volume of plateback based upon the theory of energy trapping. However, it is very difficult to suppress more than 80dB. We solved the problem with a 4-pole configuration in which two elements are connected in cascade and by suppressing spurious responses generated at the individual elements by shifting the frequency positions of these spurious responses in relation to each other. We attempted an improvement in the attenuation by designing a pattern in which the ratio (W / L) in the electrode dimensions is varied. The design parameters are shown in Table 2.

Table 2 Electrode pattern size of 130MHz 2-pole HFF-MCF.

	Pattern A			Pattern B		
Dimension [mm]	W	L	D	W	L	D
	0.35	0.22	0.13	0.40	0.20	0.15
Ratio (W / L)	1.6			2.0		
Area(W × L) [mm^2]	0.08			0.08		
Film thickness [Å]	1,000			1,000		



MCF electrode pattern

The electrode patterns A and B shown in Table 2 have different ratio (W / L) and different distance (D) dividing the split electrode. In order to match the terminal impedance and the passband width between the input and the output, the electrode area (W × L) and the electrode film thickness were made the same in patterns A and B.

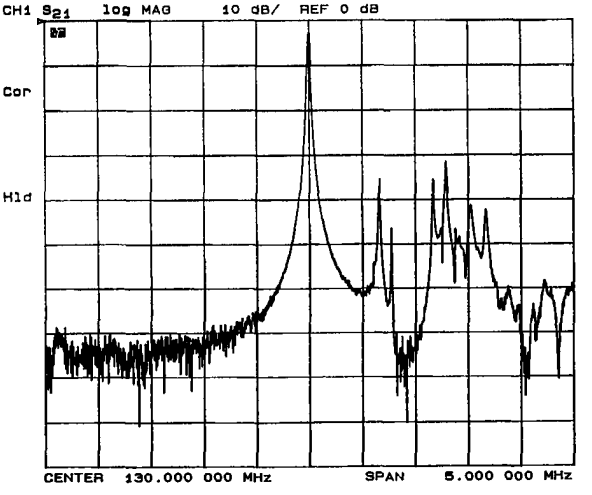


Fig.14 Spurious response characteristics of 130MHz 2-pole HFF-MCF (electrode pattern A).

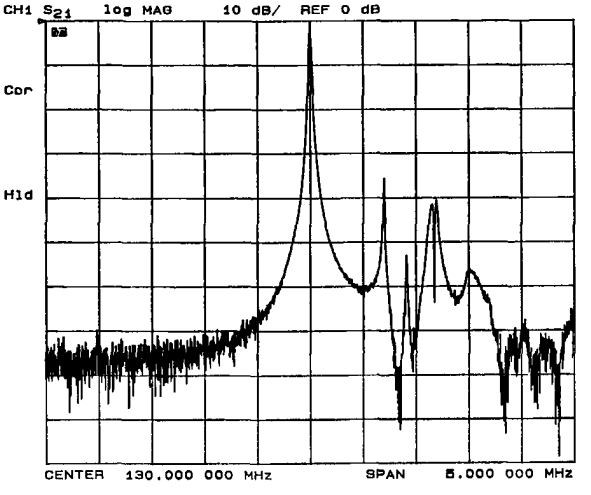


Fig.15 Spurious response characteristics of 130MHz 2-pole HFF-MCF (electrode pattern B).

Figures 14 and 15 show the spurious response characteristics of the 2-pole filter with the electrode patterns A and B respectively in the prototypes. The results clearly reveal that the spurious responses generating positions are offset.

Therefore, we made a 4-pole configuration with two elements connected in cascade, using the 2-pole filters in which the spurious response characteristics are different from each other. Figure 16 shows its stopband attenuation characteristics. In the attenuation, the spurious response is suppressed at both high and low frequencies, while permitting attenuation characteristics greater than 75dB for more than $f_0 \pm 200$kHz. In addition, characteristics for the passband and the group delay were achieved that are completely identical to those described earlier.

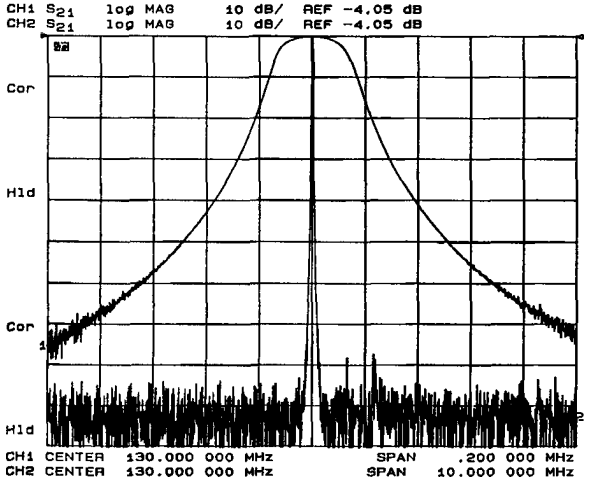


Fig.16 Stopband attenuation characteristics of 130MHz 4-pole HFF-MCF (after suppression of spurious).

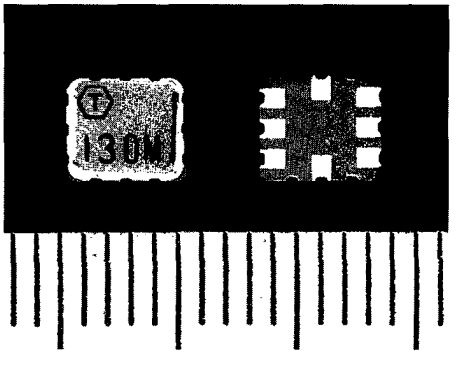


Fig.17 Photograph of 130MHz 2-pole HFF-MCF.

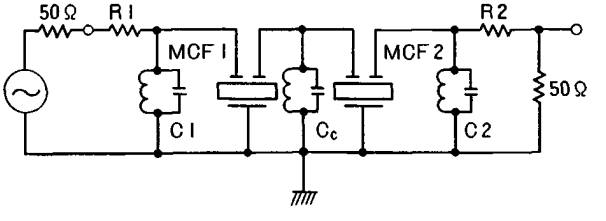
The Q value of the resonator of the prototype 2-pole filter also achieved approximately 30,000. Each HFF-MCF was mounted on SMD ceramic package with dimensions of 5.2mm × 4.5mm × 1.1 ^{t} mm, smaller than that used for the filter in the 90MHz. An external view of the unit is shown in Figure 17.

Featuring excellent attenuation characteristics, the filter works as the 1st IF filter in the RF section of a digital portable telephone, helping to cut down on the number of orders of ceramic filter for the 2nd IF filter[6]. This helps decrease the height of ceramic filters, contributing greatly to achieving thin designs in portable telephones.

71MHz wide band IF filter

HFF-MCFs have a smaller capacitance ratio than filters using overtone mode, so they enable use of a comparatively large relative bandwidth. In the past, it was difficult to work with overtone mode filters when fabricating wide band filters for the Group Special Mobile (GSM). However, we now have designed such a filter as a phase linear HFF-MCF with a 71MHz center frequency and a ± 80kHz bandwidth. To achieve a wide bandwidth while working within the designed requirements for MCFs, we had to make the distance split electrode extremely small, just a few μ m in design. Conventional vacuum evaporation processes can not fabricate such a filter. However, we fabricated electrodes accurately by using photolithography.

The filter configuration and the measuring circuit are shown in Figure 18. Its passband characteristics and group delay characteristics are shown in Figure 19 and its stopband attenuation characteristics are shown in Figure 20.



R1=R2=4.2kΩ, C1=C2=−1.4pF, Cc=−1.9pF

Fig.18 Circuit configuration and measured circuit of 71MHz 4-pole HFF-MCF.

In the 4-pole configuration in which two elements are connected in cascade, characteristics with the center frequency at 71MHz, a insertion loss of 0.2dB, a 3dB passband width of ± 88kHz and a group delay distortion of 0.9 μ s at $f_0 \pm$ 80kHz are achieved. Furthermore, stopband attenuation values of 20dB at $f_0 \pm$ 200kHz, 40dB at $f_0 \pm$ 400kHz and more than 60dB at $f_0 \pm$ 800kHz are achieved.

In this product, the terminal impedance is 4.2k Ω // −1.4pF, the interstage capacitance (Cc) is −1.9pF. Because it is for the wide band, the filter requires an inductor.

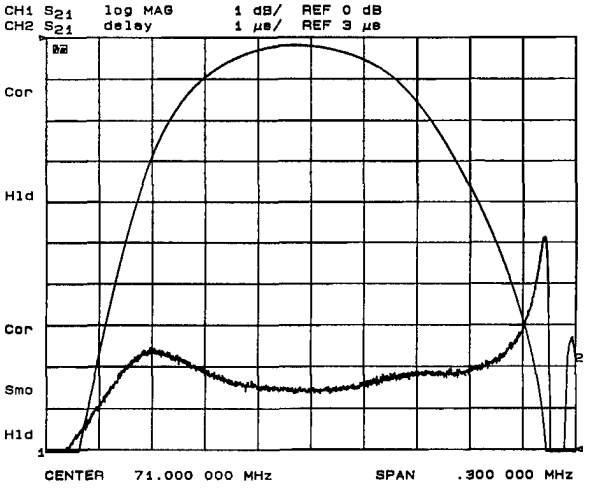


Fig.19 Passband characteristics and group delay characteristics of 71MHz 4-pole HFF-MCF.

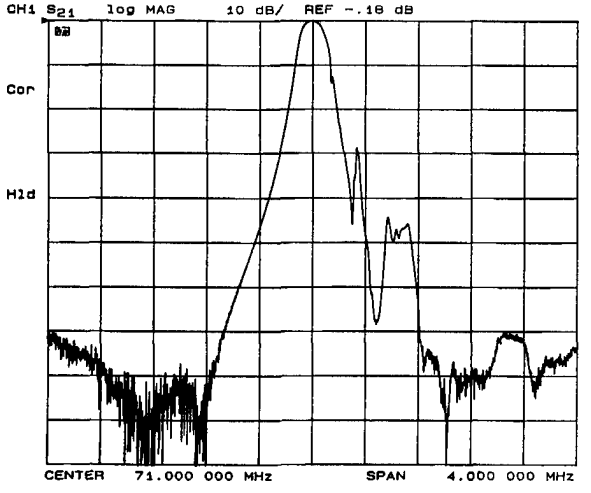


Fig.20 Stopband attenuation characteristics of 71MHz 4-pole HFF-MCF.

Each 2-pole HFF-MCF mounts on the 7mm × 5mm × 1.5 ^{t} mm SMD ceramic package, in the same manner as the 90MHz HFF-MCFs. In the future, we will work to solve the problem of the large impedance resulting from the wide bandwidth.

Conclusions

High frequency resonators and MCFs operating in the fundamental mode were fabricated in the prototype through etching technology using a batch processing system which is effective to mass production. HFF-MCFs were compared with MCFs using conventional overtone mode. A compact filter with low loss, inhibited spurious responses and low impedance was achieved. In an application of this filter as an IF filter for portable telephones, we fabricated of SMD type filters in the 90MHz narrow band, the 130MHz middle band and the 71MHz wide band, the characteristics of all these filters were satisfied with the standard specification.

We still plan to achieve higher frequency, better performance and further miniaturization.

References

[1] J. S. Wang, S. K. Watson and K. F. Lau, "Reactive Ion Beam Etching for VHF Crystal Resonators", Proc. 38th Freq. Cont. Symp., pp.101-104, 1984.

[2] R. C. Smythe, M. D. Howard and J. R. Hunt, "VHF Monolithic Crystal Filters Fabricated by Chemical Milling", Proc. 39th Freq. Cont. Symp., pp. 481-485, 1985.

[3] J. Brauge, M. Fragneau and JP. Aubry, "Monolithic Filters using Ion Etched Fundamental Mode Resonators", Proc. 39 th Freq. Cont. Symp., pp.504-513, 1985.

[4] F. M. Stern and J. Dowsett, "The Fabrication of High Frequency Fundamental Crystals by Plasma Etching", Proc. 43rd Freq. Cont. Symp., pp.634-639, 1989.

[5] T. Saito, O. Ishii, T. Morita and Y. Nakazawa, "VHF Range MCF using Fundamental Mode", 22nd EM Symposium of Japan, pp.51-56, 1993.

[6] O. Ishii, T. Morita, T. Saito and Y. Nakazawa, "Fundamental Mode IF MCF for Handheld Phones", TOYO TECHNICAL REPORT, No.52, pp.12-20, 1995.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SELF LIMITING ETCHING OF PIEZOELECTRIC CRYSTALS

K.M. Lakin, G.R. Kline, and K.T. McCarron

TFR Technologies, Inc.
701 SE Salmon Ave.
Redmond, OR 97756

ABSTRACT

Piezoelectric plates have been thinned by various techniques including chemical and plasma etching, This paper reports on a novel method of plasma thinning piezoelectric plates in order to make high frequency resonators in inverted mesa or other configurations. The method involves using plasma excitation at the frequency where resonance is desired. Theoretical and experimental results are presented.

I. INTRODUCTION

Significant advances have been made in plasma processing for etching and deposition in microelectronic device fabrication. To a limited extent plasma processing has been applied to crystal resonators as an alternative to chemical etching but without satisfactory control of the resonant frequencies [1-6]. A plasma processing method is described here that should produce large area UHF crystal plates with precise control of frequency and be suitable for automatic trimming of resonators and filters fabricated on wafers.

Consider the plasma etching system drawn schematically in Fig. 1. Here a crystal wafer is placed between capacitor plates similar to a typical parallel plate plasma or reactive ion etching (RIE) system. It is desired to control the resonant frequency of the crystal by etching the plate in a gas plasma to reduce its thickness and raise the resonant frequency to some predetermined value. Assume for example that the frequency of the RF generator driving the plasma is set to a desired resonant frequency that is above the resonant frequency in the crystal, i.e. the crystal plate is too thick for the desired frequency. In the presence of the gas plasma the surface of the crystal is eroded by the action of the plasma and the crystal becomes thinner, thus raising its resonant frequency towards the generator setting.

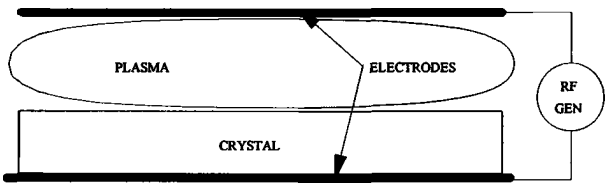


Fig. 1. Basic plasma etching system with plasma and crystal in series with RF power source.

The etch rate of the crystal is dependent upon the rate of ion impingement and energy imparted to the surface and hence the RF current and power. The RF current is controlled by the impedance of the parallel plate capacitor which includes the crystal and the plasma. The plasma discharge is approximated by a constant voltage breakdown and hence a small differential impedance. Current through the crystal is therefore largely determined by the impedance of the crystal. This condition is most apparent if the external drive impedance to the plasma is low (acting more like a voltage source). Under plasma etching the crystal plate thins and can be arrange so that the plate resonance changes relative to the generator frequency.

THEORY

The plasma etch rate has been modeled for this system with some reasonable approximations. Assume that the etch or deposition rate is proportional to the magnitude of the RF power and that the external circuit impedance is low. Under these conditions the voltage across the crystal is determined by the low impedance external source and constant voltage drop nature of a plasma discharge.

The differential equation for thickness is given by;

$$\frac{da}{dt} = r\frac{|V|^2}{|Z|} \qquad (1)$$

0-7803-2500-1/95 $4.00 © 1995 IEEE

where V is the voltage, Z circuit impedance, a the thickness, r a proportionality constant and t is time. In (1) Z is evaluated at the RF generator frequency f_g.

Equation (1) is more easily solved for time as a function of thickness because Z is a function of thickness. The differential expression for time becomes;

$$t = \frac{1}{r|V|^2}|Z(\phi(\alpha))|d\alpha \qquad (2)$$

where,

$$\alpha = \frac{V}{2f_p}\ ,\ \phi(\alpha) = \frac{\pi f_g}{2 f_p}$$

and where V is the velocity. Using the frequency normalization for thickness α suggested above, (2) may be integrated to give t as a function of the parallel resonant frequency f_g ;

$$t = \frac{1}{R}\int_{\alpha_o}^{\alpha} \left|1 - K^2 \frac{\tan\phi}{\phi}\right| \frac{d\alpha}{\alpha^3} \qquad (3)$$

where $\phi = \frac{\pi}{2}\frac{1}{\alpha}$, $\alpha = \frac{f_p}{f_g}$

and where α_o is a normalized parallel resonant frequency corresponding to the initial plate thickness, α is the normalized frequency corresponding to the thickness at time t and the convenient normalization frequency is the plasma generator frequency, f_g. The coefficient R is an effective deposition or etching rate. The impedance of the crystal is assumed to be of the form plotted in Fig. 2 as used analytically in the integrand of (3).

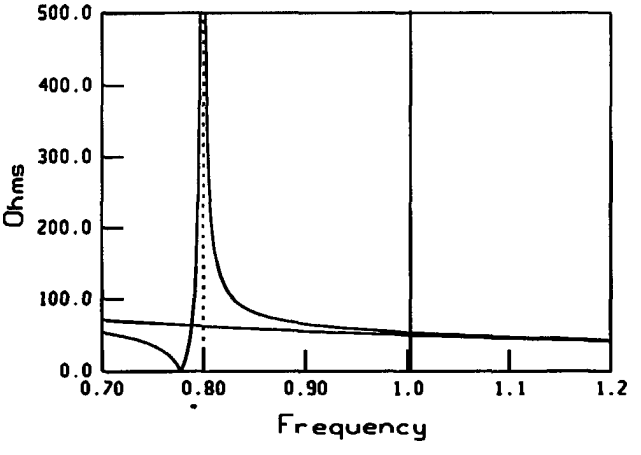


Fig. 2. Absolute value of impedance of crystal resonator versus normalized frequency. Plasma generator at 1 and parallel resonance at 0.8.

The general results of this modeling are shown in Fig. 3. Etching starts from a thickness equivalent to a parallel resonant frequency 0.8 of the generator frequency. The etching proceeds at a uniform rate, indicated by the initial slope of the curve, causing the resonant frequency of the plate to increase toward the generator frequency. The change of resonant frequency with time then levels off when the current decreases as parallel resonance is approached. Etching proceeds slowly during parallel resonance, due to decreased plasma current, causing a slow drift towards series resonance wherein etching proceeds very rapidly. Finally, etching continues again at a rate determined by the simple dielectric capacitance of a plate that is getting thinner.

The deposition process is almost the mirror image of the etching process with the plasma generator set to a frequency below parallel resonance of the plate. Deposition occurs at a normal rate until series resonance, where the current increases, and then levels off during parallel resonance where the current is minimum.

Because the etching process automatically approaches the parallel resonant frequency of the crystal, a high degree of

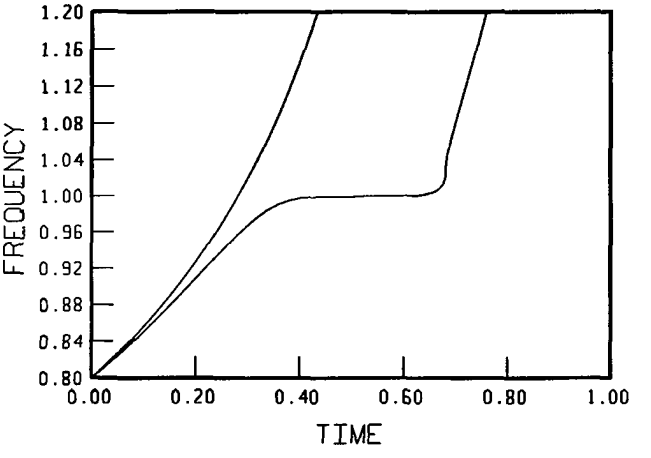


Fig. 3. Frequency change with time during etching, left curve for non-piezoelectric plate.

planarization of a crystal blank or wafer should occur. Suppose there was a thickness variation across the wafer. Setting the generator to a frequency at the parallel resonant frequency of the thinner region would cause greater etching in thicker regions until parallel resonance is obtained. Etching would occur on a localized basis until the entire wafer reached the same parallel resonant condition. This assumes that the parallel resonant etch rate of the initially thinner region is sufficiently slow to prevent drift into series resonance.

The effects of thickness non-uniformity are illustrated in Fig. 4 where initial wafer non-uniformities of 5% and 10% were modeled. By programming the RF generator to shift upward in frequency towards the end of the first parallel resonance condition the wafer would be moved out of the first region's parallel resonance and away from the danger of drifting into series resonance. The effect of a step in RF generator frequency is equivalent to starting the curves from zero time but now with the relative thickness variations much reduced. This procedure provides for an

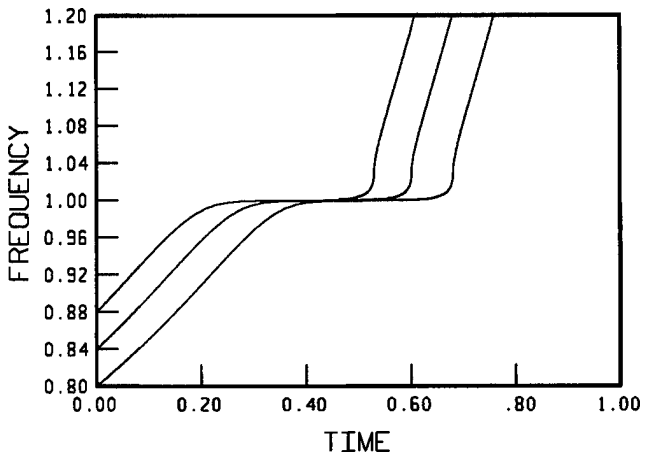


Fig. 4 Etching characteristics with three different starting thicknesses on the wafer normalized to their parallel resonant frequencies.

effectively wider time band at parallel resonance at the new generator frequency setting. In this manner a production system could be programmed to planarize a wafer in steps and get a high degree of thickness precision and resonator frequency uniformity.

The characteristics of the plasma are such that as current is reduced to near the point of extinction the plasma departs from the constant voltage relationship, the current decreases, and the plasma may extinguish effecting and automatic etch stop at the parallel resonant frequency.

EXPERIMENTAL RESULTS

A very simple demonstration of the use of a crystal to control the current through a plasma can be conducted in a one dimensional sense by connecting a mounted quartz crystal in series with a small neon bulb and a signal generator. The brightness of the plasma glow is an indicator of RF current flow and hence resonance conditions.

In order to demonstrate actual etching a small vacuum chamber was built of stainless steel and fitted with plasma etching fixtures. The steel chamber provided adequate shielding of the RF so that spurious radiation did not occur and cause interference.

The chamber pressure was monitored by a capacitance manometer and regulated manually by use of the roughing valve. The system was designed to use two different gases which are mixed and flow into the chamber under the control of mass flow meters. Chamber pumping was done with a rough pump having FOMBLIN oil. A small 1.27 cm diameter cathode and accompanying ground shield were installed for the purpose of etching small diameter quartz crystals.

The matching network at low frequencies consisted of the conventional circuit shown in Fig. 5 but tunable over a much wider frequency range than commercial plasma matching networks.

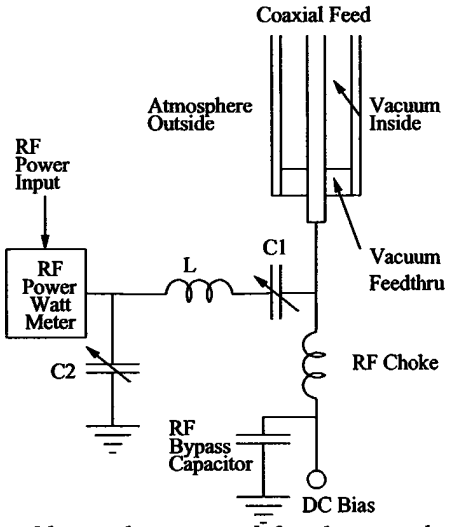


Fig. 5 Tunable matching network for plasma etcher.

Initially a 5 cm x 2.5 cm x 0.22 mm AT quartz blank was centered on the cathode for etching. The parallel resonant frequency of the wafer was measured to be 7500.78 KHz by observing plasma extinguishment at parallel resonance. By rocking the generator frequency across the resonant frequencies, parallel and series resonance could be observed. At series resonance the plasma glow became more intense and moved closer the surface of the quartz while at parallel resonance the dark space increased significantly, lifting the plasma off the surface of the quartz blank. With higher power the plasma did not extinguish at parallel resonance.

The plasma generator was placed a few KHz above the parallel resonance of the crystal and the crystal was allowed to etch to the point where the parallel resonant frequency was equal to the generator frequency. At parallel resonance the plasma extinguished demonstrating the automatic etching stop. The process was repeated by resetting the generator to a higher frequency and allowing the crystal to continue etching. It was necessary to set the generator power (approximately 2 watts) to a level that would allow the plasma to extinguish at parallel resonance. The overall power density was equivalent to 125 watts for a 100 mm diameter wafer which is in the range for normal RIE of silicon wafers.

With higher power the plasma does not extinguish at parallel resonance. In this case the generator frequency can be swept over the resonances and the resonance conditions observed by noting the character of the plasma glow as described above.

Gas (Freon, CF4) pressure was maintained in the 20 to 100 mTorr range to assure sufficient energy for reactive ion etching..

Plasma etching at UHF was initiated using a miniature version of the original plasma etcher. This was necessary in order to keep the cathode size within reason and to allow for efficient matching of the generator to the plasma load. The alternate chamber consisted of nominal 1.75 inch vacuum system components including a standard cross and viewing ports. The matching network, Fig. 6 consisted of a quarter wave coaxial resonator fed off center for impedance matching. One end of the resonator constitutes the cathode in the vacuum chamber and the end outside the vacuum is DC open circuited to allow for a DC bias to be established within the plasma. The position of the input tap along the resonator allows impedance matching to a 50 source.

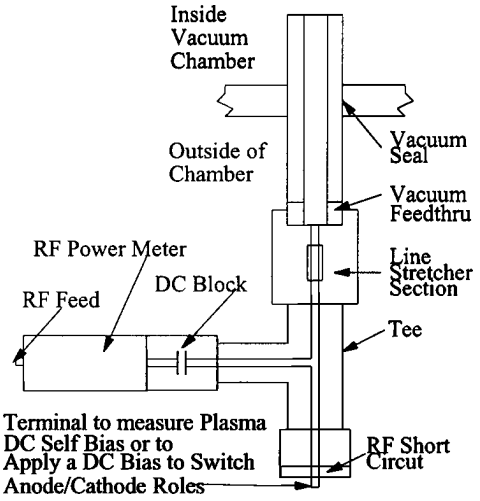


Fig. 6. Microwave matching network for plasma etching.

In practice the surface of the crystal must remain flat during normal etching so that a parallel resonance can be obtained at the proper thickness. Accordingly, some means must be implemented to assure enough flatness to realize a parallel resonance.

During normal etching, away from any resonance in the plate, the etching profile is determined by the electric field and displacement current distribution in the crystal plate. It was found necessary to pattern an electrode on the quartz side in contact with the cathode in order to assure a uniform potential across the quartz Fig. 7. Without the electrode it was found that the machining marks on the cathode post were replicated on the top of the quartz wafer. Although undesirable, the results showed the extent to which the etching was being controlled by series impedance of the circuit since air gaps under the quartz created by groves in the cathode post constitute low capacitance high impedance paths for RF current.

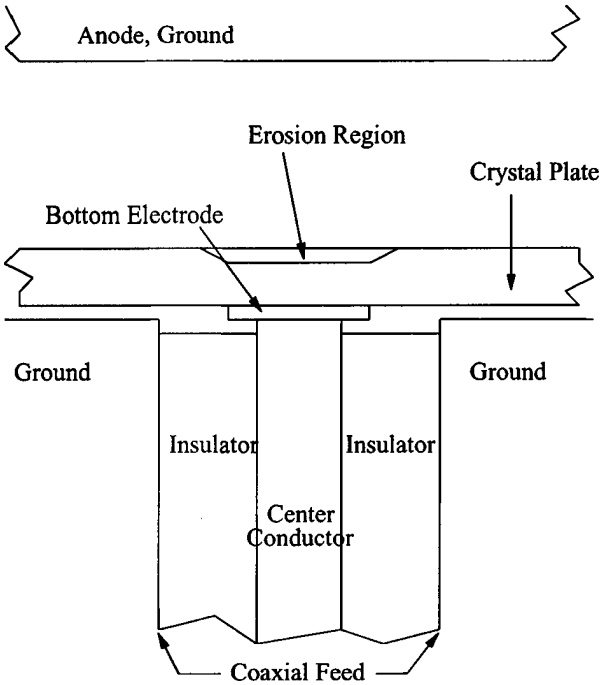


Fig. 7 Inverted mesa etching with electrode formed on the bottom of the piezoelectric plate.

An anode plate placed above the quartz wafer and cathode region provided a more uniform current flow and flatter surface. It was configured so that the cathode to anode distance was adjustable. The anode produces a more perpendicular electric field distribution above the crystal rather than the fringing fields that extend to the edge of the crystal. Shown in fig. 8 is a profile of a 25 micrometer deep pocket etched in the surface of a quartz blank.

As an alternative to automatic stopping by extinguishing the plasma at parallel resonance, the resonance conditions can be detected optically by observing the position of the plasma glow relative to the top surface of the piezoelectric plate, Fig. 9

SUMMARY AND CONCLUSIONS

A novel technique for plasma etching piezoelectric crystals has been demonstrated that uses the impedance of the thinning resonator to control the etch rate. Self limiting

etching was successfully demonstrated on quartz. This was done, however, under very narrow operating conditions where the generator frequency was set so that relatively little etching was required for the crystal to reach resonance.

Plasma etching at UHF was demonstrated using a unique cathode and matching network that consisted of a coaxial resonator that was part of the high vacuum feed through components and cathode.

Further work needs to be done to develop the process for production. Issues such as crystal heating during plasma etching and mode effects need to be addressed.

V. REFERENCES

1. "Chemically Polished High Frequency Resonators", W.P. Hanson, Proc. 37 th Ann. Freq. Contr. Symp., (AFCS), 1983, pp. 261-264.

2. "Chemically Milled VHF and UHF AT-Cut Resonators", J.R. Hunt and R.C. Smythe, Proc. 39 th Ann. Freq. Contr. Symp., (AFCS), 1985, pp. 292-300.

3. "A New Design for High Frequency Bulk Resonators", A. Lepek and U. Maishar, Proc. 43 rd Annual Frequency Control Symposium, Denver, CO, pp. 544-547, May 31-June 2, 1989.

4. "Acoustic-Bulk-Wave Resonators and Filters Operating in the Fundamental Mode At Frequencies Greater Than 100 MHz", M. Berte, Electronic Letters, Vol. 13, No. 9, pp. 248-250, Apr. 28, 1977

5. "The Fabrication of High Frequency Fundamental Crystals By Plasma Etching", F.M. Stern, J. Dowsett, D.J. Carter, and R.J. Williamson, Proc. 43 rd Ann. Freq. Contr. Symp., (AFCS), 1989, pp. 634-639.

6. "Reactive Ion Beam Etching for VHF Crystal Resonators, Proc. 34 th Ann. Freq. Contr. Symp., (AFCS), 1984, pp. 101-104.

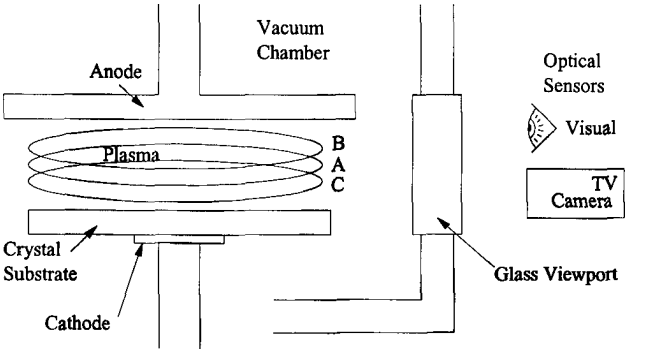


Fig. 9. Optical detection of resonance. The position of the plasma moves from A to B at parallel resonance, to C and series resonance, and back to A for normal etching.

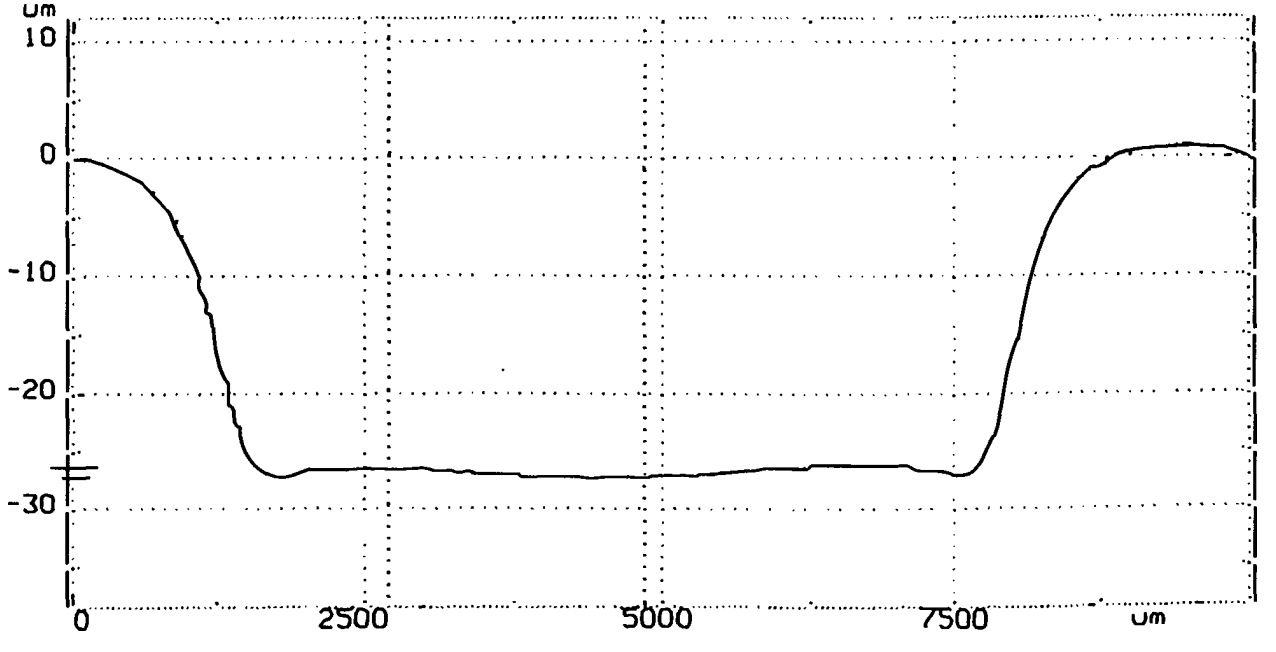


Fig. 8. Etch profile at the surface of a quartz plate used to form an inverted mesa.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

EXPERIMENTS TO MINIMIZE FREQUENCY SHIFTS AND YIELD LOSSES FOR SURFACE MOUNT FILTERS

IYAD ALHAYEK AND JOE VOISSEM

Component Products Group, Motorola
1305 E. Algonquin Rd.
Schaumburg, IL. 60193

ABSTRACT

Piezoelectric quartz surface mount filters typically are composed of an electroplated quartz element, a ceramic carrier, a ceramic or metal lid, and an adhesive that attaches the quartz element to the ceramic carrier. The ceramic carrier and lid provide a hermetic seal for the quartz element after the device is provided with an inert environment. The hermetic seal is accomplished by either a solder seal process or a seam weld process. Electrodes on the quartz element couple the electrical signals into and out of the piezoelectric quartz material. A conductive adhesive attaches the quartz element to the carrier and the electrical signal is transmitted to the outside world via electroplated conductors.

Mechanical shock isolation and thermal compliance are two requirements of the blank mount system to insure that mechanical or thermal shocks do not result in unwanted frequency shifts or shifts in the device parameters. A third requirement is that the adhesive and the sealing methods do not expose the quartz element to contaminants and gases, which could result in unacceptable device shifts. The rate of adhesive curing and the dimensions and locations of the adhesives can also result in unwanted stresses on the quartz element, which then can cause unwanted frequency shifts or unwanted vibration modes. Recovery from these frequency shifts or a continuation of these frequency shifts can then take place during and after the reflow to the customer PC board. To provide frequency stability for the customer, the built up stresses often can be minimized by thermally exercising the adhesive/quartz/carrier system. Such thermal exercising is provided by a method called high temperature pre-conditioning.

It is believed that the high temperature pre-condition step minimizes future outgassing of the adhesive, and that it also improves the cross-linking of the adhesive and its adhesion to the carrier and quartz elements.

This paper will deal with designed experiments that were used to determine an acceptable thermal conditioning sequence. To be acceptable, thermal conditioning must be sufficiently high to provide cross-linking of the material to itself and to the bond sites. Also it must be sufficiently low so as not to shock or degrade the adhesive material. Because the size and location of the adhesive dots are related to the amount of thermal stresses, this paper will also describe experiments that were related to robotic adhesive settings and robotic quartz element placement. These were important because they are directly related to frequency stability in the customer application.

The experiments were performed for a solder seal package and a seam weld package. The package design for each seal process is unique; therefore, each process requires its own set of pre-conditions and its own amounts and locations of adhesives.

SOLDER SEAL HIGH TEMP EXPERIMENT

Previous to the studies of this paper, process information suggested that a lowering of the high temperature pre-condition would improve the shifts through seal, thereby improving the process yields. A Process optimization team developed a two-way ANOVA designed experiment with three levels each of high temperature pre-condition peak time and peak temperature to determine if a lower temperature would be robust. In order to reduce the number of cells in the study, a single replicate design was used. A single-replicate, two factors study can be analyzed if the interaction term is shown to be insignificant. John Tukey developed a method to break the interaction term into a non-additivity term and an error term. [1] If the non-additivity term is shown to be insignificant, then the conclusion is made that interaction is insignificant and the remaining conclusions related to the main effects are valid. If the interaction term is shown to be significant, then the statistical conclusions related to main effects are not valid.

0-7803-2500-1/95 $4.00 © 1995 IEEE

The design of the experiment was as follows:

HIGH TEMP EXPERIMENT FOR SOLDER SEAL					
			PEAK TIMES		
			t. Ref. - 5	t. Ref.	t. Ref. + 5
P E A K	T E M P S	T. Ref. - 20			
		T. Ref.			
		T. Ref. + 20			

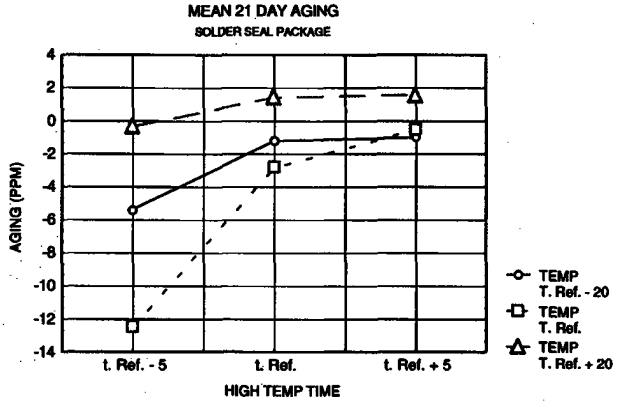
The filters were placed into aging and ANOVA analysis was performed on the aging results.

Following is the ANOVA table for the mean aging:

MEAN 21 DAY AGING ANOVA TABLE					
SOURCE OF VARIATION	SUM OF SQUARES	DEG. OF FREEDOM	MEAN SQUARES	F CALC	ALPHA RISK
PEAK TIME	57.3278	2	28.6639	F = 28./1.8 =15.54	0.05
PEAK TEMP	64.9024	2	32.4512	F = 32./1.8 = 17.60	0.01
NON-ADDITIVITY	24.5830	1	24.5830	F = 24/1.8 = 13.33	0.05
ERROR	5.533	3	1.8440		
TOTAL	152.3462	8			

The peak time and peak temperature were significant, but the non-additivity term (and therefore, the interaction) was also significant. Because the interaction term was significant, the conclusions made relating to the significance of the peak time and peak temperature were not valid.

Though valid statistical conclusions could not be made, the following peak time and peak temperature effect plots were made to provide suggestions for future experiments:



The above plot agrees with previous experience that a low peak time could lead to large frequency shifts during and after the seal operation. The lowest peak time of t. Ref. - 5 appeared to be inadequate to assure acceptable aging.

SEAM WELD HIGH TEMP EXPERIMENT

Because the previous high temp precondition experiment could not be statistically tested, new samples were built, this time with replication. Because it was desirable to convert from solder seal to seam weld, this new experiment and all new experiments utilized a seam weld package. The high temp temperatures were lowered because experience gained elsewhere suggested that the seam weld package would behave better with a lower peak temperature. In the new experiment, the peak time range being studied was increased from (t. Ref. - 5 to t. Ref. + 5) to (t. Ref. - 10 to t. Ref. + 10).

Following is the design matrix for the second experiment:

HIGH-TEMP EXPERIMENT FOR SEAM WELD DESIGN					
			PEAK TIMES		
			t. Ref. - 10	t. Ref.	t. Ref. + 10
P E A K	T E M P S	T. Ref. - 20			
		T. Ref.			
		T. Ref. + 20			

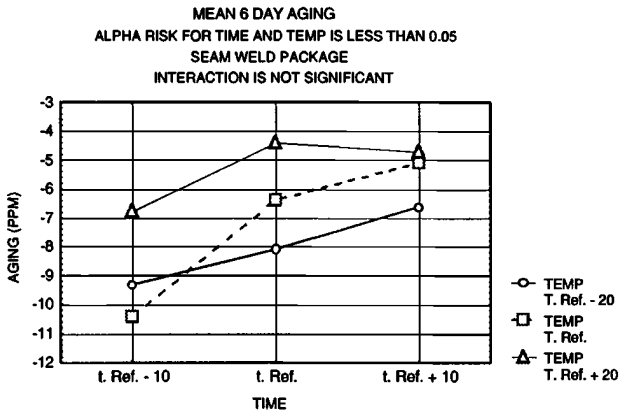
The filters were placed in the PQE environments and ANOVA analyses were performed on the results.

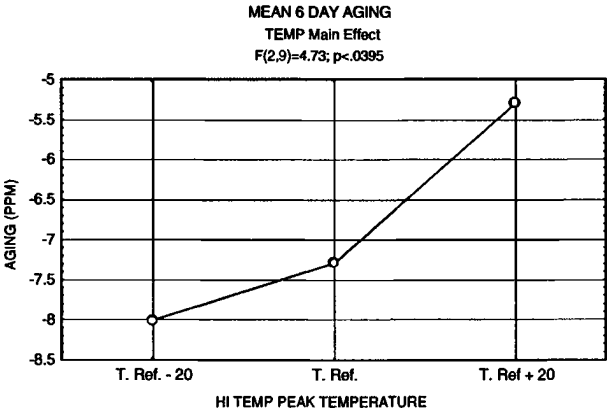
The ANOVA table from the analysis for mean 21 day aging was as follows:

MEAN 21 DAY AGING ANOVA TABLE					
SOURCE OF VARIATION	SUM OF SQUARES	DEG. OF FREEDOM	MEAN SQUARES	F CALC	ALPHA RISK
PEAK TIME	36.988	2	18.49740	F = 18.5/2.5 = 7.4	0.013
PEAK TEMP	23.848	2	11.92402	F = 11.9./2.5 = 4.8	0.040
TIME X TEMP	7.984	4	1.99589	F = 1.99/2.5 = 0.8	0.560
ERROR	22.702	9	2.52243		
TOTAL	91.522	17			

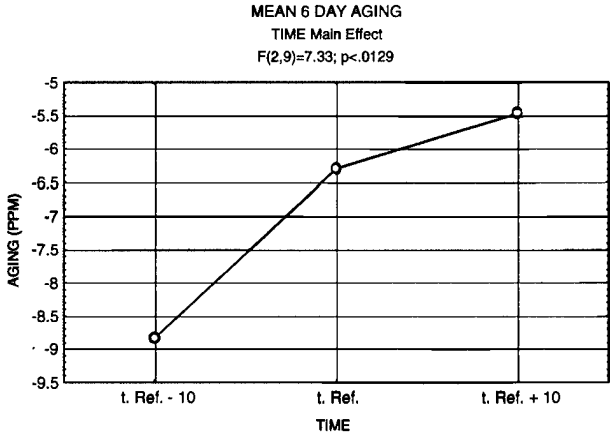
The above ANOVA table indicates that the peak time and peak temperature were significant to an alpha risk of less than 0.05, but that the interaction was not significant.

Following is the effects plot for the 6 day aging:





The above plot could be presented and analyzed as in the two plots and Duncan's post hoc tables that follow:



DUNCAN'S POST HOC TEST FOR PEAK TIME				
6 DAY AGING		-8.84 PPM	-6.29 PPM	-5.46 PPM
		t. Ref. - 10	t. Ref.	t. Ref. + 10
-8.84 PPM	t. Ref. - 10		0.02	0.03
-6.29 PPM	t. Ref.	0.02		0.39
-5.46 PPM	t. Ref. + 10	0.03	0.39	

The above plot and table indicate that t. Ref. - 10 was significantly worse than t. Ref. + 10 and t. Ref. It also indicates that t. Ref. was not significantly different from t. Ref. + 10.

The effects plot and Duncan's table for the peak temperature effect on the aging was as follows:

DUNCAN'S POST HOC TEST FOR PEAK TEMP				
MEAN 6 DAY AGING		-8.01 PPM	-7.29 PPM	-5.29 PPM
		T. Ref. - 20	T. Ref.	T. Ref. + 20
-8.01 PPM	T. Ref. - 20		0.450	0.075
-7.29 PPM	T. Ref.	0.450		0.057
-5.29 PPM	T. Ref. + 20	0.075	0.057	

The above table indicates that no temperature was significantly different from any other temperature to an alpha risk of 0.05, but T. Ref. - 20 and T. Ref. were significantly worse than T. Ref. + 20 to alpha risks of 0.057 and 0.075, respectively.

Combining conclusions from the above Duncan's tables results in the conclusion that a peak time between t. Ref. and t. Ref. + 10 and a peak temperature of T. Ref. + 20 will result in the best frequency shift through aging.

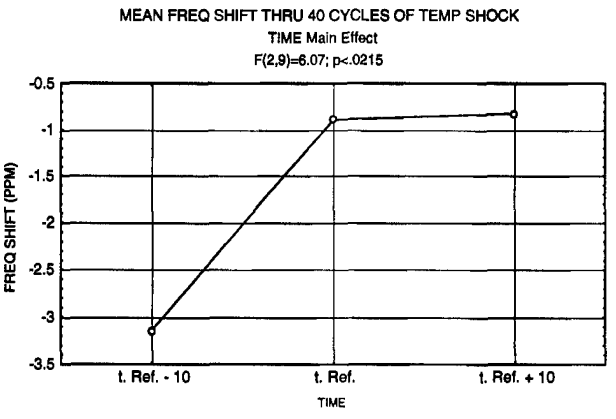
ANOVA analysis was performed for the frequency shift through 40 cycles of temperature shock.

Following is the ANOVA table:

MEAN TEMPERATURE SHOCK SHIFT ANOVA TABLE					
SOURCE OF VARIATION	SUM OF SQUARES	DEG. OF FREEDOM	MEAN SQUARES	F CALC	ALPHA RISK
PEAK TIME	21.10349	2	10.51745	F = 10.5/1.7 = 6.18	0.021
PEAK TEMP	1.4437	2	0.72185	F = 0.72/1.7 = 0.42	0.672
TIME X TEMP	14.6100	4	3.65250	F = 3.65/1.7 = 2.15	0.560
ERROR	15.6044	9	1.73382		
TOTAL	52.7616	17			

The peak time had a significant effect on the frequency through temperature shock.

Following are the effects plot and Duncan's table:



DUNCAN'S POST HOC TEST FOR PEAK TIME				
F. SHIFT IN TEMP SHOCK		-3.15 PPM	-0.89 PPM	-0.82 PPM
		t. Ref. - 10	t. Ref.	t. Ref. + 10
-3.15 PPM	t. Ref. - 10		0.016	0.065
-0.89 PPM	t. Ref.	0.016		0.934
-0.82 PPM	t. Ref. + 10	0.065	0.934	

The above table indicates that t. Ref. - 10 was significantly worse than t. Ref. with an alpha risk of 0.016, and worse than t. Ref. + 10 with an alpha risk of 0.065. It also indicates that there was no statistically significant difference between t. Ref. and t. Ref. + 10.

SEAM WELD VERIFICATION RUN

Data from the above experiments were used to developed the pre-age peak time and peak temperature that would minimize the frequency shift through both aging and temperature shock. These settings were confirmed by means of a verification run.

FIRST ROBOTIC SENSITIVITY DESIGNED EXPERIMENTS

After the high temp pre-condition experiments were completed, robotic adhesive dot placement was studied in two separate one-way experiments as follows:

ADHESIVE VOLUME (OR MASS) DOE				
MASSES				
w. Ref. - 0.4	w. Ref. - 0.2	w. Ref.	w. Ref. + 0.2	w. Ref. + 0.4
TWO REPLICATES FOR EACH OF THE ABOVE.				
ALL HAD A HEIGHT OF 19.0.				

ADHESIVE HEIGHT DOE		
HEIGHTS		
17.5	19.0	20.5
TWO REPLICATES FOR EACH OF THE ABOVE.		
ALL HAD A MASS OF w. Ref.		

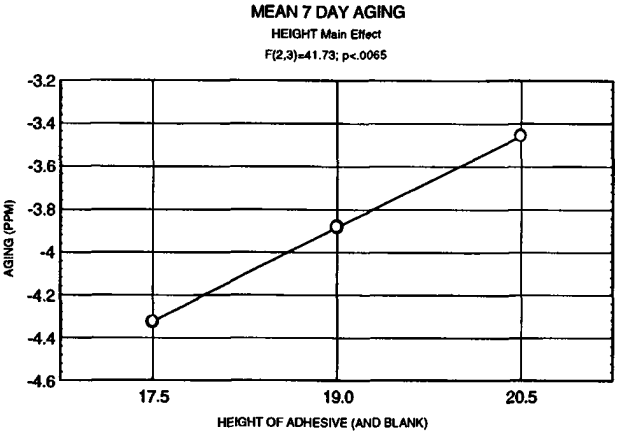
ANOVA analysis was performed for the frequency shift through 10 cycles of temperature shock.

Following is the ANOVA table:

MEAN TEMP SHOCK SHIFT ANOVA TABLE					
SOURCE OF VARIATION	SUM OF SQUARES	DEG. OF FREEDOM	MEAN SQUARES	F CALC	ALPHA RISK
HEIGHT	0.7483	2	0.37415	F = 0.37/0.009 = 41.7	0.0064
ERROR	0.0269	3	0.00897		
TOTAL	0.7752				

The above ANOVA table indicated that the height had a significant effect on the temperature shock frequency shifts.

Following is the effects plot and its associated Duncan's post hoc test table for 7 day aging:



DUNCAN'S POST HOC TEST FOR ADHESIVE HEIGHT				
MEAN 7 DAY AGING		-4.325	-3.885	-3.460
		17.5	19.0	20.5
-4.325	17.5		0.019	0.116
-3.885	19.0	0.019		0.021
-3.460	20.5	0.116	0.021	

The above Duncan's table indicates that the height of 19.0 was significantly different from 17.5 and 20.5

SECOND ROBOTIC SENSITIVITY DESIGNED EXPERIMENT

The robotic experiment just discussed treated the heights separately from the volumes. Because it would be preferable to include both factors together, a full factorial was designed to include both factors.

Following is the design of the new experiment:

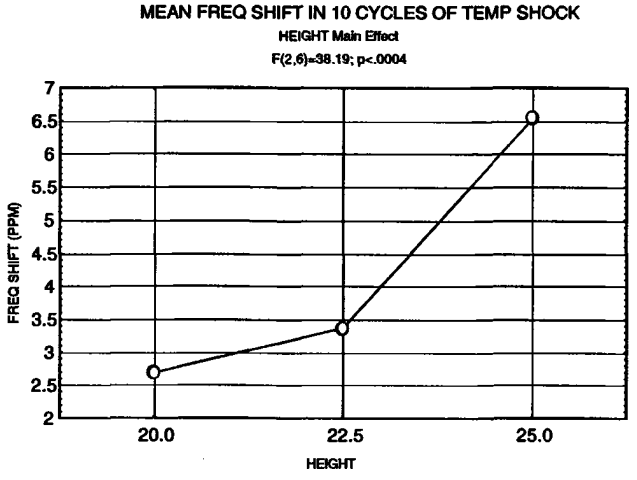
		ADH VOL.	
A		w. Ref.	w. Ref. + 0.4
D			
H	20.0	2 REPL	2 REPL
H	22.5	2 REPL	2 REPL
G			
T	25.0	2 REPL	2 REPL

ANOVA analysis on the frequency shift through 10 cycles of temperature shock resulted in the following:

MEAN TEMP SHOCK SHIFT ANOVA TABLE					
SOURCE OF VARIATION	SUM OF SQUARES	DEG. OF FREEDOM	MEAN SQUARES	F CALC	ALPHA RISK
HEIGHT	33.9326	2	16.9663	F = 17/0.4 = 38.2	0.0004
VOLUME	0.6769	1	0.6769	F = 0.7/0.4 = 1.52	0.2632
HGT X VOL.	0.1586	2	0.0793	F = 0.08/0.5 = 0.18	0.8408
ERROR	2.6652	6	0.4442		
TOTAL	37.4333	11			

The above indicated that the adhesive (or blank) height had a significant effect on the temperature shock.

Following is the effects plot and Duncan's post hoc test table:



DUNCAN'S POST HOC TEST FOR ADHESIVE (OR BLANK) HEIGHT				
F. SHIFT IN TEMP SHOCK		2.68 PPM	3.37 PPM	6.54 PPM
		20.0	22.5	25.0
2.68 PPM	20.0		0.1965	0.0012
3.37 PPM	22.5	0.1965		0.0007
6.54 PPM	25.0	0.0012	0.0007	

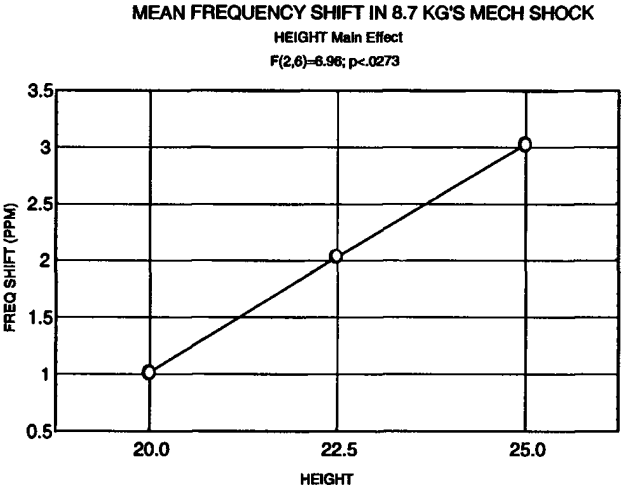
The above indicated that heights of 20.0 and 22.5 had significantly different frequency shifts from 25.0, but that the difference between 20.0 and 22.5 was not significant.

Following is the ANOVA analysis for adhesive (or blank) height and adhesive volume on mechanical shock frequency shift:

MEAN MECHANICAL SHOCK SHIFT ANOVA TABLE					
SOURCE OF VARIATION	SUM OF SQUARES	DEG. OF FREEDOM	MEAN SQUARES	F CALC	ALPHA RISK
HEIGHT	8.0604	2	4.0302	6.959	0.027
VOLUME	0.5376	1	0.5376	0.928	0.372
HGT X VOL.	1.1162	2	0.5581	0.964	0.434
ERROR	3.4746	6	0.5791		
TOTAL	13.1888	11			

The above table indicated that the adhesive (or blank) height had a significant effect on the 8.7 kg's mechanical shock frequency shift.

Following are the effects plot and Duncan's post hoc test table for the mechanical shock frequency shift.



DUNCAN'S POST HOC TEST FOR ADHESIVE (OR BLANK) HEIGHT				
F. SHIFT IN MECH. SHOCK		1.015 PPM	2.027 PPM	3.022 PPM
		20.0	22.5	25.0
1.015 PPM	20.0		0.109	0.045
2.027 PPM	22.5	0.109		0.114
3.022 PPM	25.0	0.045	0.114	

The above indicated that the frequency shift difference between heights of 20.0 and 25.0 was significant, but that the differences between 20.0 and 22.5 and between 22.5 and 25.0 were not significant.

CONCLUSIONS FROM ABOVE TWO ROBOTIC STUDIES

Combining the above two robotic studies led to the conclusion that the height of 17.5 resulted in bad aging. The height of 25.0 resulted in bad frequency shift in temperature shock and mechanical shock. The volume of adhesive did not effect the environmental frequency shift. Because of this, the optimum height setting would be between 20.0 and 22.5. The middle of the adhesive volume ranges (w. Ref.) would assure margin on both sides.

PROCESS VERIFICATION RUNS

A verification run was performed with the best settings of the following process variables:

Cure time	Adhesive height
Cure temperature	Adhesive volume
High temp peak temperature	High temp peak time

ROBOTIC SENSITIVITY DESIGNED EXPERIMENT WITH TWO DIFFERENT LOTS OF ADHESIVE

Some of the subsequent testing resulted in poor yields for Fo-150 kHz (ultimate attenuation). Poor condition of the adhesive lot was suspected to be one of the causes of the unacceptable yields. Because of this, an experiment was developed to determine if the poor yields were caused by the adhesive or by some of the robotic settings. At the same time, temperature characteristic testing was added as a second response variable.

Following was the design of the experiment (single replicate design):

		SUSPECT ADH BATCH				NEW ADH BATCH		
		PIN HEIGHTS						
			18.5	20.5	22.0	18.5	20.5	22.0
DOT	HGHTS							
		13.0						
		19.0						
		24.0						

Following is a table of the number of failures found for the new lot of adhesive for frequency shifts at -40 C and for the room temperature ultimate attenuation parameter (Fo-150 kHz).

		DEFECTS IN NEW LOT OF ADHESIVE						
		-40 C Frq. DEFFECTS				Fo - 150 kHz DEFECTS		
		PIN HEIGHTS						
		18.5	20.5	22.0		18.5	20.5	22.0
DOT	HGHTS							
	13.0	0	0	1		1	0	1
	19.0	1	0	1		1	0	1
	24.0	0	0	0		0	0	2

The above table indicates that the new lot of adhesive had only a small number of failures for both response variables. The 0% failure rate of the 20.5 pin height may be related to the fact that this was close to the optimum setting as recommended in the previous robotic sensitivity experiments.

Following is a table of the number of failures found for the suspect lot of adhesive:

		DEFECTS IN SUSPECT LOT OF ADHESIVE						
		-40 C Frq. DEFFECTS				Fo - 150 kHz DEFECTS		
		PIN HEIGHTS						
		18.5	20.5	22.0		18.5	20.5	22.0
DOT	HGHTS							
	13.0	2	3	5		2	0	0
	19.0	9	0	3		1	2	0
	24.0	6	4	2		1	0	1

The number of failures was small for the Fo - 150 kHz parameter, but was large for the TC at -40C. Only the optimum setting for dot height and pin height was defect free for the -40C temperature characteristic parameter.

The results suggested that there was an interaction between the batch of adhesive and the locations of the adhesive for the temperature characteristic test. Poor yields for -40C temperature characteristics were obtained only with the combination of both the suspected epoxy lot and non-optimum robotic settings

CONCLUSIONS

1. The high temperature pre-condition time was extremely important for optimizing the frequency shifts through the qualification environments.

2. Robotic sensitivity studies can help optimize process yields and frequency shifts in the qualification environments.

3. Optimized robotic settings can make the process unaffected by a wider set of adhesive conditions.

Reference

[1] D. C. Montgomery, Design and Analysis of Experiments, New York: John Wiley and Sons, 1991, ch. 7, pp. 219-221.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

INTERNAL HEATED QUARTZ RESONATOR WITH LOW SENSITIVITY TO AN ACCELERATION

Igor V. Abramson

Omsk Research Instrumentation Institute
P.O. Box 7357, 644020 Omsk, Russia

Abstract: The paper describes results of development of the internal heated quartz resonator novel design. Such a resonator is built on the composite heating the crystal plate technique and possesses an unique compromise of high frequency stability with 15 s warm-up time, 4 mA heating current and less than $5 \cdot 10^{-10}$ 1/g acceleration sensitivity. It's shown in the paper the performances of the oscillator based on the IHR are very attractive for various global radionavigation and communication mobile systems.

1. Introduction.

Recent development of global radionavigation and radiocommunication systems (GPS, INMARSAT, etc.) demands development of mobile OCXOs that meet following set of fundamental requirements [1]:

- high frequency versus temperature stability over a wide ambient temperature range;
- low ageing rate;
- short warm-up time;
- low power consumption;
- low phase-noise level;
- less than $5 \cdot 10^{-10}$ 1/g acceleration sensitivity;
- small oscillator volume.

At a glance at the requirements it is clear that some of them, such as short warm-up time, low consumption and high frequency stability, are difficult to impliment simultaneously when conventional approach to the OCXO building is used. As it's shown in a number of works [2-4] one of promising way to design mobile OCXOs uses the internal heated resonator (IHR) technology.

At the 1992 IEEE AFCS the author presented one of the most advanced IHR design - the composite heated resonator, in which the combination of heating the crystal plate with the film and transistor heaters is used providing surprising short warm-up time and very low consumption power along with high enough frequency versus temperature and time stability [5].

Present work describes some new achievements in the composite heating resonator characteristics resulting from essential modification of its construction and technology.

2. Basic performances of the IHR novel design.

The operation principle and the basic construction of the composite heated resonator are considered in [5,6]. As it is shown in the papers the fundamental advantages of the device - extremely low power consumption and very short warm-up time at high temperature stability result from redistribution of the heating power between the film and the transistor

0-7803-2500-1/95 $4.00 © 1995 IEEE

heaters at the start moment and in the stationary conditions. Owing to the start power is almost entirely dissipated on the crystal plate its temperature rises very fast reaching the SC-cut lower turnover temperature in several seconds. In the stationary conditions almost whole power dissipates in the transistor heater and the small part does in the film heaters that leads to negligible thermal gradients over the plate resulting in high frequency versus temperature stability. Placing the power transistor of the thermocontrol system inside the resonator volume provides also low consumption power of the resonator due to excellent isolation of the heated elements from an environment.

As compared with the previous composite heated resonator design the new one uses essentially modified crystal holder with lower thermal capacity and improved thermal isolation. The holder is placed in glass encloser of 18 mm diameter and 35 mm length. Instead of the glass encloser the metal one can be use provided it has at least five isolated leads and better than 0.1 Pa vacuum degree.

Fundamental characteristics of the developed IHR are displayed in fig. 1-3.

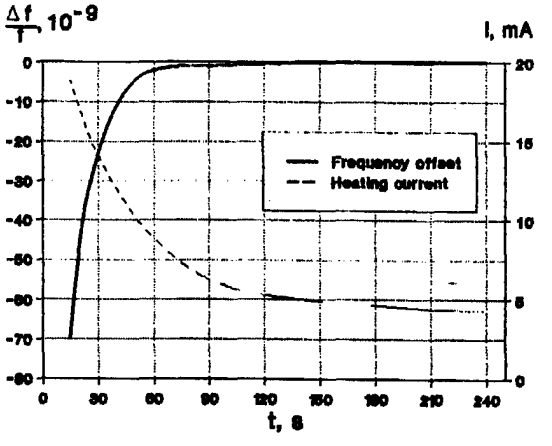


Fig.1. The composite heated resonator warming-up curves.

As one can see from the resonator transient process curves depicted in fig. 1 it takes less than 15 s to set the frequency with $1 \cdot 10^{-7}$ accuracy to the nominal frequency and about 30 s for $2 \cdot 10^{-8}$ accuracy. Thanks to lower thermal capacity of the holder the heating current through the resonator falls off very rapidly reaching after 2 min of operation 1.2 times of the steady state value. Thereat the whole energy consumpted from the voltage source during the warming-up is less than 1 J that makes the device attractive for the battery supplied oscillator.

The steady state heating current through the resonator is about 4 mA that is about the current consumpted by the oscillator amplifier circuit.

Temperature stability of the new resonator has been explored carefully exciting the SC-crystal plate on the thermosensive B-mode. As it follows from fig. 2 the frequency shift of the operating IHR excited on the B-mode over a temperature range -60 +70 °C is about 100 Hz that corresponds to ± 0.2 °C temperature shift. So small temperature variation can change the SC-cut resonator frequency by not more than $1 \cdot 10^{-9}$ that along with insignificant values of the plate thermal gradients allows to achieve high temperature stability of the resonator.

Typical frequency versus temperature

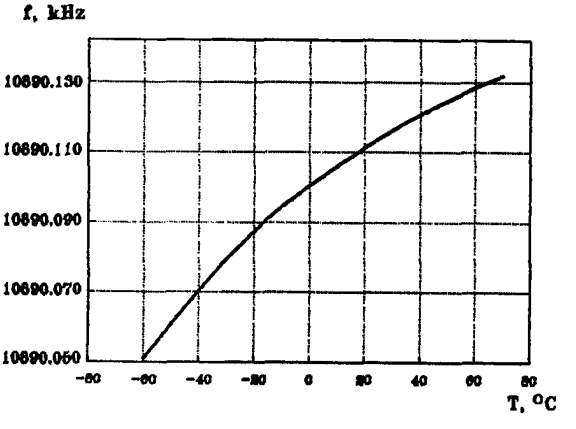


Fig.2. The B-mode Frequency of the IHR vs. ambient temperature.

instability of the IHR is depicted in Fig.3 to be about $5 \cdot 10^{-9}$ that corresponds to the stability of the best OCXOs.

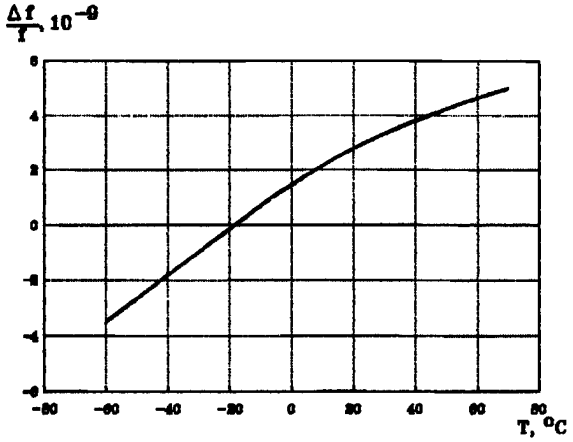


Fig.3. The IHR frequency vs. temperature stability.

Typical ageing curves of the developed IHR are depicted in fig.4. As it follows from the data the ageing rate bellow $2 \cdot 10^{-8}$ for the first month operation is observed on three samples and the rate of less than $1 \cdot 10^{-8}$ for second month is provided with all the IHRs.

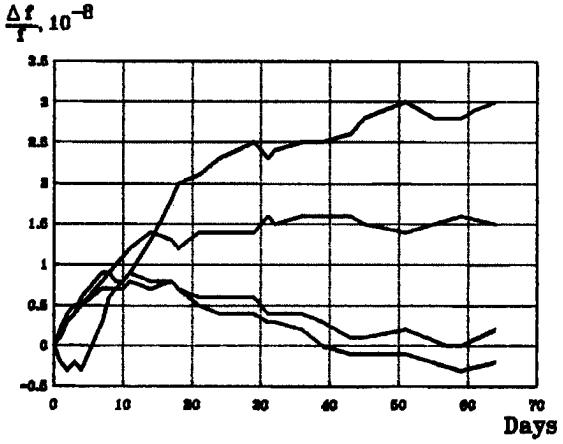


Fig.4. The typical ageing curves of the IHRs.

3. Acceleration sensitivity of the IHR

In the last few years reduction of the acceleration sensitivity of crystal resonators remains as one of the most important and complex problem. This problem is especially difficult to solve when IHRs designed due to limitations imposed on the plate geometry and sizes choice as well as on the holder construction.

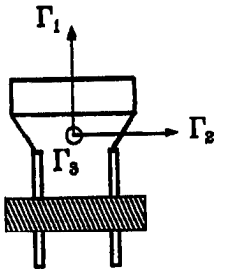


Fig.5. The crystal plate two-point mounting structure.

To find the most effective way to reduce the acceleration sensitivity of the IHR its physical origins have been explored theoretically and experimentally. It has been determined that the normal component of the sensitivity Γ_1 for the plate two-point mounting structure outlined in fig. 5 is produced mainly by compression-extention stresses in the plate arising from the plate flexure under the normal component of the acceleration. This component of the sensitivity is reduced when the plate has a close to the simmetric form and is fixed at the the frequency vs. stresses zero azimuth.

It the contrast, the in-plane component of the sensitivity along the mounting points - Γ_2 - is caused by thickness-shear stresses in the plate and can be reduced when the plate is mounted at the $\bar{c}_{666}$ non-linear constant zero azimuth. However, such arrangement of the mounting points may impact on the resonator long-term stability as does not coincide with the force-frequency coefficient zero orientation. To reduce this component without degradation of the long-term

stability a special mounting structure which doesn't transform the holder displacements into the plate thickness-shear stresses has been used in the resonator.

The in-plane component Γ_3 is appeared to be about zero when the plate is fixed symmetrically to its geometry and vibration area center. Non-zero values of this component observed usually in practice should be related to the deflections from the structure ideal configuration.

As a result of the researches the IHR G-sensitivity has been reduced by a factor of 3 - 5 as compared with previous design. Typical figures of the G-sensitivity for 4 samples of the IHR are shown in table 1.

Table 1.

No. THR	G-sensitivity, 10^{-10} 1/g			
	Γ_1	Γ_2	Γ_3	$\Gamma = (\Gamma_1^2 + \Gamma_2^2 + \Gamma_3^2)^{1/2}$
159	2.0	3.5	0.5	4.0
160	1.5	2.5	1.5	3.3
161	1.0	1.5	1.5	2.4
162	2.0	4.0	2.0	4.9

As it follows from the data the most values are related to the in-plane component along the mounting points Γ_2 and the least values are related to the in-plane component Γ_3 .

All the samples have the sensitivity level below $4 \cdot 10^{-10}$ 1/g for "the worst" direction, that is close to the results accomplished with use of the QAS crystal technology [7].

4. Performances of the oscillator based on the composite heated resonator.

Developed composite heated resonators have been involved in the oscillator design that allowed to reach combination of performances to be unattainable for any other kinds of OCXOs (table 2).

Table 2.

Characteristics	Data
Operating frequency range, MHz	8-13
Temperature range, °C	-60 +70
Frequency instability:	
temperature	$\pm 3 \cdot 10^{-8}$
short-term (1 s)	$<1\ 10^{-11}$
long-term per day	$(1-5)\ 10^{-10}$
Warm-up time, s	
to $1 \cdot 10^{-7}$	<15
to $2 \cdot 10^{-8}$	<30
Power consumption	
at the start moment, (before 5 s), W	1
In the steady state (in 3 min after switch on, mW	150
G-sensitivity (maximum), 10^{-10} 1/g	4
Phase-noise level, dBc at offset:	
100 Hz	-150
1000 Hz	-155
10000 Hz	-160
Volume, cm^3	45

Temperature instability of the oscillator to be 3-5 times greater than that of IHR itself is mainly caused by temperature instability of the oscillator circuit to be placed out of the thermostated volume. Further reduction of the temperature instability can be reached recluding

the oscillator curcuit components selecting unwanted B-mode that is possible when the lateral excitation of the SC-cut crystal is used.

On the data in table 2 one can conclude that the OCXO built on the developed IHR to a large extend corresponds to the requirements of up-to-date navigation and communication mobile systems that makes it promising for such applications.

Acknowledgements

The author expresses a gratitude to Mr. A. Kutalev for a help in testing the IHR in the OCXO design. The author also thanks Miss N. Kuchatova for an assistance in experimental data treatment and preparation of this paper.

References

[1] G. Marotel, G.Caret, J.P. Aubry "An Oscillator for the GPS Program", in Proc. 40th Ann. Freq. Cont. Symp., 1986.

[2] Ya.L. Vorokhovsky and B.G. Drakhlis, "High Stability Quartz Oscillators on Internally Heated Quartz Resonators with AT and SC Cuts", in Proc. 45th Ann. Symp. on Freq. Cont., USA, 1991, pp. 447-451.

[3] Ya.L. Vorokhovsky, I.G. Petrosian, B.G. Drakhlis, A.A. Volkov "The Further Development of Internally Heated Quartz Resonators and the Oscillators on Their Basis for Use in Up-to-date Search-and-Rescue and Communication Systems", in Proc. 1992 IEEE Freq. Cont. Symp., pp. 505-512.

[4] B. Long and G. Weaver, "Quartz crystal oscillators with direct resonator heating", in Proc. 45th Ann. Symp. on Freq. Cont., USA, 1991, pp. 384-392.

[5] I.Abramson, A. Dikidzhi "Improvement of Characteristics of Quartz Resonator-Thermostate with Direct Heating Piezoelement", in Proc. 1992 IEEE Freq. Cont. Symp., pp. 499-504.

[6] I. Abramson "OCXO Design Using Composite-Heating of the Crystal Resonator", IEEE Trans. Ultrason., Ferroelec., and Freq. Cont., vol.41, no.2, pp.284-287, March 1994.

[7] M. Zafrany and oth. "G-sensitivity of High Frequency Quartz Crystal Resonators", in Proc. of the 8th EFTF, Germany, 1994, pp. 93-102.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

FREQUENCY-TEMPERATURE CHARACTERISTICS OF QUARTZ CRYSTAL UNITS OF DIFFERENT CUTS OPERATING OVER A WIDE TEMPERATURE RANGE INCLUDING HELIUM TEMPERATURES

A. G. SMAGIN

Crystal Co., 105425, Moscow, 9th Parkovaya str., 47-1-29, Russia

Abstract

Investigation of frequency vs temperature dependence of quartz crystal elements of singly- and doubly-rotated cuts during excitation of B- and C-type vibrations in them has been carried out over the temperature range from 4 to 400 K. The plate orientation accuracy was ±30". Isonormal quasitransversal vibrations and pure shear vibrations have been exited in metallized piezo-electric elements at frequency 5 MHz and 10 MHz.

When setting up equations for the electronic computer it is necessary to make into account the change of curvature sign due to relaxation and to make operations with generalized frequency-temperature characteristics (FTCs). FTCs of yxbl/+15°-35°19' crystal units (type B-vibrations) have extrema at T=86 K and 229 K. FTCs of crystal units of the same cut in which vibrations of type C have been excited, have no extrema except for the maximum at T → 0 K. Crystal units of other cuts have such FTC features as maxima, minima and flat portions with the inflection.

Introduction

Acoustic and frequency system based on quartz and quartz resonators operate in objects of new engineering at temperatures of liquid nitrogen, hydrogen and helium.

Quartz crystal resonator temperature lowering causes substantial decrease in processes velocity, stipulating surface, contact and structural irreversible changes in crystal and its associated device [1].

Resonator ageing decreases by exponential law with temperature lowering and attains highly small value at T → 0 K [2].

Crystal oscillator designing with resonators ovenized at the temperature of liquid nitrogen and helium makes it possible to obtain of extremely high short-term and long-term stabilities. This problem can be solved only by searching for quartz crystal cuts having zero TCFs at nitrogen and helium temperatures, because with high stability, the frequency variation at temperature variations assumes ever greater importance.

Experimental technique

For reducing FTCs sample spread of crystals in one cut without macroscopic defects have been chosen for manufacturing piezoelectric elements and thorough control of their optical and structural homogeneity has been made. Quartz crystal plates oriented at X-ray goniometer with the error of ±30". Samples having maxima caused by relaxation have not been considered, because relaxation distoris substantially (statistical treatment of results) FTC of quartz.

The set of equipment for crystal unit FTC measurement consists of a vacuum cryostat with coaxial terminals and apparatus for quartz excitation and vibration frequency and temperature measurements. Evacuated cavity to be measured represents a massive red copper shell inside of which two cylindrical thermostating screens are located. Quartz crystal units were placed inside of an isothermal copper screen with silver clad and polished internal surface. The isothermal screed tu was hanged at kapron threads inside of a vacuum shell, immersed into a Dewar with a cooling agent. At the external surface of the screen temperature wire sensors were wound, the phosphorous bronse sensors served as arms of a bridge in the temperature regulating circuit at the internal surface differential thermocouples were located for temperature gradient control along the height and the diameter of the screen. With great difference from cooling agent temperature (besides cooling agents boiling at normal and lowered pressure ice scrumbled in small pieces and drowned in water has been used) at the external screen an intermediate temperature has been maintained, and the regulation has been effected manually with the accuracy of ±0,1° by bridge disbalance, into the arm of which a temperature sensor of the screen has been switched. The temperature of internal isothermal screen was maintained by automatic temperature regulator with

0-7803-2500-1/95 $4.00 © 1995 IEEE

the accuracy of ±0,01°. The temperature was measured by quartz frequency thermometers specially fabricated by us which have been gradually by means of reference resistance thermometers - those platinum and germanium ones (TSPN-1 and TSGe-2) with the absolute accuracy of ±0,01°.

Temperature gradients within the volume of operating chamber did not exceed ±0,025°. Frequency measurement results are highly critical to sample temperature rate of change (rates <0,1° min are permissible)&

The procedure described enabled us to increase temperature measurement accuracy and reduce the time of exposure at fixed temperatures, because the heat inertion time of resonators under investigation and that of temperature sensors have been equal. Moreover, the measuring circuit has been simplified considerably: FTC and temperature have been measured by using the same apparatur. Samples excitation at the scries resonance frequency in oscillator operating over a wide range has been carried out with the error of $\pm 1 \times 10^{-7}$.

Theory.

The main equation, describing elastic waves motion in anisotropic medium is Cristoffel equation:

$$(\Lambda - \lambda)\,\bar{u} = 0, \qquad \lambda = v^2,$$

where tensor of the rank two

$$\Lambda = \lambda_{ijlm}\, n_e\, n_j = (c_{ijlm}/\rho)\, n_e\, n_j,$$

$\bar{u}$ is the displacement vector, λ is the effective reduced elastic constant in the direction n, v is the phase velocity, n_e and n_j are direction cosines of the wave normal. Solution of the equation contains effective elastic constant values for three waves propagating along the wave normal:

$$\lambda_n = c_n/\rho = v^2 = \omega^2/k^2, \qquad \omega = 2\pi f$$

is the vibration eigen frequency,

$\bar{k}$ is the wave vector, ρ is the crystal density. Taking into account boundary conditions the vibration frequency is:

$$\omega_m = \frac{\pi}{\ell}\sqrt{\lambda_n^E}, \qquad m = 1,2,3$$

Elastic vibrations of three isonormal waves, propagating in the crystal are designated as types of vibration A, B and C ($\omega_A > \omega_B > \omega_C$). In a general case vibrations of type A are quasilongitudinal, vibrations of types B and C are quasitransversal. So far as frequencies are non-linear functions of temperature, temperature variations of quartz crystal frequency and all the values connected with their calculation, it is customary to describe them by power series of the kind [3]:

$$\frac{\omega - \omega_0}{\omega_0} = \frac{\Delta\omega}{\omega_0} = \sum T^{(n)}\omega\,(T - T_0)^n,$$

$$T^{(n)}\omega = \frac{1}{n!\,\omega_0}\left(\frac{\partial^{(n)}\omega}{\partial T^{(n)}}\right)_{T=T_0},$$

where $T^{(n)}\omega$ are temperature coefficients (TCs) of frequency defining completely the character of dependence $\omega = \omega(T)$. By differentiating (3) over temperature, it is possible to obtain relationships of temperature coefficients frequency and TCs of effective reduced elastic constants

$$T_\lambda^{(1)} = 2T\omega^{(1)} + 2T_\ell^{(1)},$$

$$T_\lambda^{(2)} = 2\left[T\omega^{(2)} - \frac{1}{2}\left(T\omega^{(1)}\right)^2\right] + 2T_\ell^{(2)} + \frac{1}{2}\left[\left(T\lambda^{(1)}\right)^2 - 2\left(T_\ell^{(1)}\right)^2\right],$$

$$T_\lambda^{(3)} = 2\left[T\omega^{(3)} - T\omega^{(2)}T\omega^{(1)} + \frac{1}{3}\left(T\omega^{(1)}\right)^3\right] + 2T_\ell^{(3)} + \left[T_\lambda^{(2)}T_\lambda^{(1)} - 2T_\ell^{(2)}T_\ell^{(1)}\right] - \frac{1}{3}\left[\left(T_\lambda^{(1)}\right)^3 - 2\left(T_\ell^{(1)}\right)^3\right],$$

where $T_\lambda^{(n)} = T_\ell^{(n)} - T_\rho^{(n)}$,

$T_\rho^{(n)} = -2\left(\alpha_{11}^{(n)} + \alpha_{33}^{(n)}\right)$ is temperature coefficient of density, α_{11} and α_{33} are the components of thermal expansion tensor:

$$T_\ell^{(n)} = \alpha_\ell^{(n)} = \alpha_{11}^{(n)} + n_3^2\left(\alpha_{33}^{(n)} - \alpha_{11}^{(n)}\right)$$

are temperature coefficients of the plate (along the direction of wave normal). By knowing TCs of elastic constants [4] and density, and also components of thermal expansion tensor of quartz, it is possible to calculate temperature coefficients of frequency of crystal elements of any cut.

Exact solution of Cristoffel equation for an anisotropic crystal has a complicated appearance (Kardano formulae) and is practically not applicable to the general investigation of frequency vs temperature dependence. The use of cumbrous exact formulae is not resonable, and it necessary therefore to use sufficiently simple approximated relations, characterizing basic properties of elastic waves in a crystal. The use of the comparision method of a real crystal with the transversal-isotropic medium nearest to it in its elastic properties enables to simplify calculations considerably [5]. For transversal-isotropic medium (hexagonal crystal) the Cristoffel equation with any direction of wave normal breaks down into a linear and quadratic equations and allows therefore to obtain a comparatively simple general solution.

Calculation of velocities and displacements of three isonormal waves, propagating in a trigonal quartz is effected with a necessary accuracy. However, the formulae of the method described are also sufficiently cumbrous in spite of the transversal-isotropic medium use. Therefore all the calculations have been made by means of a computer.

Experimental results. Generalized frequency-temperature characteristics

Investigation of frequency-temperature dependence of quartz crystal resonators was carried out over the temperature range from 4 K to 400 K: over the range 4 K to 20 K with the step of 1°, over the range 20 K to 400 K with the step of 2.5°. At 7 resonators each of singly- and doubly-rotated cuts: yxl/+35°17' (vibrations of B and C types), yzb/+5° (vibrations of C-type), yxl/-48°5', (vibrations of B-type), yxbl/+12°30'/-33° (vibrations of C-types), yxbl/-19°06'/+34°05' (vibration of B- and C-types), yxbl/+15°/-35°20' (vibrations of B- and C-types) were measured. The orientation accuracy of quartz plates was ±30". Isonormal, quasitransversal and pure shear vibrations were excited in metallized piezoelectric elements with the diameter of d=12.5 mm at frequencies 5 MHz and 10 MHz, with the relation of diameter to length being $d/l \approx 40\text{-}55$, which serves a good approximation to the electrical boundary conditions of electrical field constancy. Measurements of dielectric permeability and piezoelectric module of quartz also confirm, that errors inserted due to non-consideration of piezoelectric errors in equations of motion, are small and are approximately constant over all the temperature range investigated. Taking into account the relation of crystal plate diameter to thickness, one can consider, that within the limits of the accuracy attained with regard to determination of temperature coefficients of frequency, $T\omega^{(n)}$ the Cristoffel theory is applicable.

Resonator frequency variation at temperatures above 15 K is determined not only by temperature dependence of its parameters, but also by change of the effective elastic constant caused by relaxation process of elastic vibration attenuation. While working out equations for the computer it is necessary to take into account the change of curvature sign, caused by relaxation and perform operations with generalized FTCs.

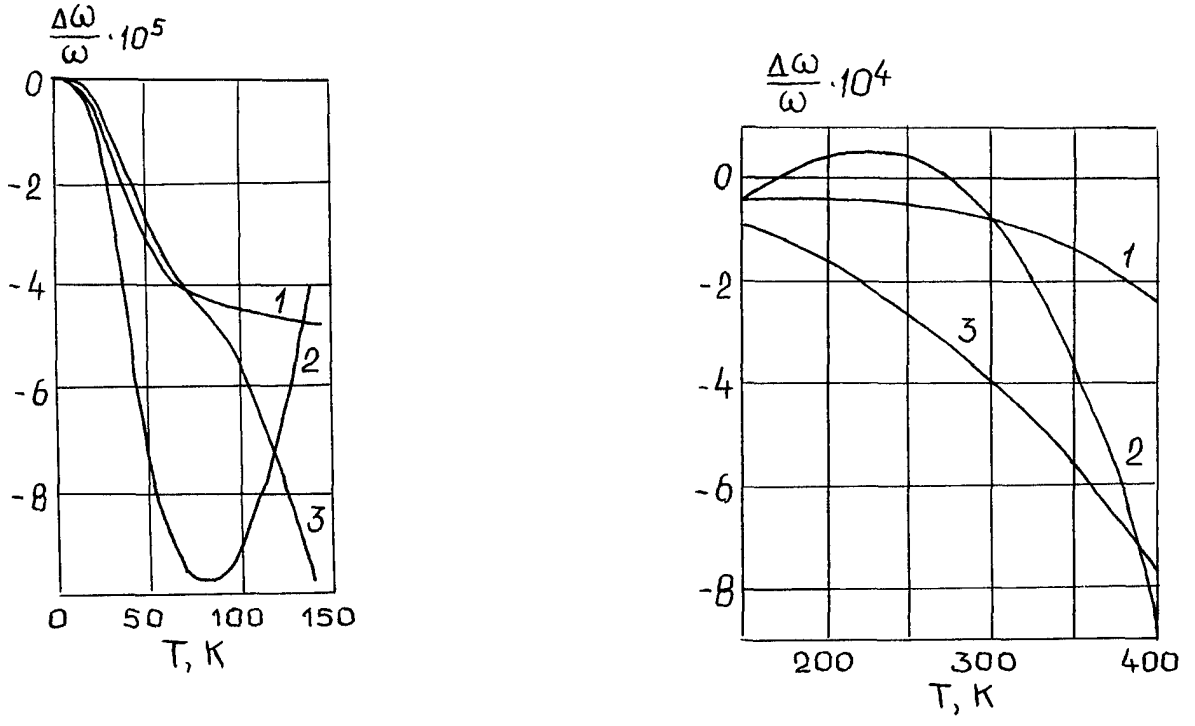
Figures 1-3 confirms theoretical representations regarding crystal elastic constants behaviour in the vicinity of T = 0 K. With the approximation T → 0 K the internal energy of a solid tends to a constant value, due to zero vibrations of a crystal lattice. Variation velocity of elastic constants vs temperature representing the second derivatives of internal energy over strains and consequently, the relative frequency variation of resonators tends towards zero at T → 0 K.

Resonators of different cuts have FTC peculiarities of the type maxima, minima, gently sloping portions with inflection (Fig. 1-3).

Resonator FTCs of yxbl/+12°30'/-33° cut (type C-vibrations) besides of the extremum at T → 0 K have a long gently sloping portion with the inflection at T=172.5 (Fig. 1, curve 1). Resonator FTCs of yxbl/+15°/-35°20' - cut (type B-vibrations, Fig. 1, curve 2) have extrema at T=86 K and T=229 K. Resonator FTCs of the same cut, in which type C vibrations have been excited, does not has an extremum (curve 3) except for the maximum at T → 0 K. Resonator FTCs of yxbl/-19°06'/+34°15' - cut (type C vibrations) have exterma at T = 311 K and T = 394 K (Fig. 2, curve). Resonator FTCs of yxl/+35°17' - cut (type C vibrations) have extrema at T = 264 K and 345 K (curve 5), and resonator FTCs of the same cut (type B vibrations) have extrema at T = 80 K and 336 K (curve 6). Resonator FTCs of yzb/+5° - cut (type B vibrations) and of yxbl/-19°06'/+34°05' - cut (type B vibrations) have no extrema (Fig. 3, curves 7, 8) except for maxima at T → 0 K. The last ones possess a rather large TCF over all the temperature range, and the cut yzb/+5° has $TCF=10^{-4}$ K^{-1} over the temperature range 300-400 K, the maximum one for all the known cuts, which makes it usable for development of high-sensitive temperature sensors.

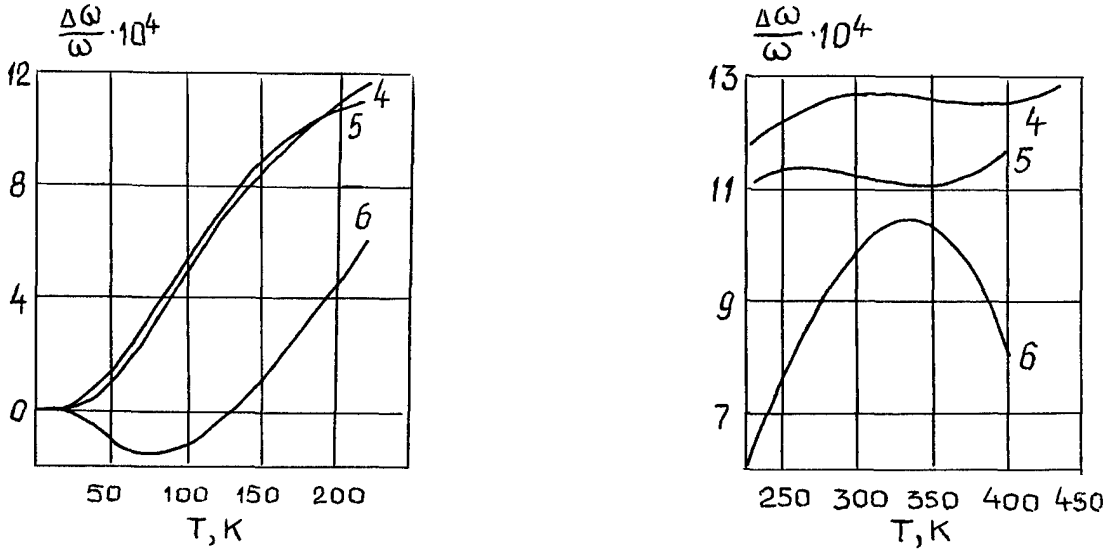
All the piezoelectric element cuts investigated have maxima or minima in the vicinity of T → 0 K.

The measurement of resonator FTCs of a single cut was averaged under supposition that measurements in the sampling are subordinate to the laws of normal distribution. By several cuts FTCs have been measured in the case of two isonormal waves, i.e. the number of equations in the sets is more than the number of unknowns. This allows to hope for partial error compensation, because in the contrary case the sum of incidental errors will enter completely in the roots of equation set. Due to in cidental errors availability, the sets are incompatible, and the most probable values of unknowns are received from the conditions of minimality of squares sum of nonbindings.



(a) (b)

Fig. 1 (FTCs of crystal units of different cuts over the temperature range from 4 K to 420 K.
1 - yxbl/+12°30'/-33° - cut (type C vibrations;
2, 3 - yxbl/+15°/-35°20' - cut (type B and C vibrations).



(a) (b)

Fig. 2 FTCs of crystal units of different cuts over the temperature range from 4 K to 437 K.
4 - yxl/+19°06'/-34°15' - cut (type C vibrations;
5, 6 - yxbl/+35°17' - cut (type B and vibrations).

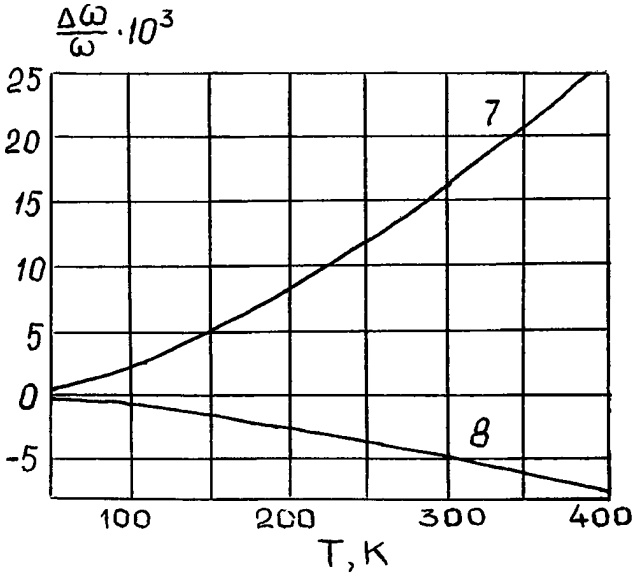


Fig. 3 FTCs of crystal units of different cuts over the temperature range from 4 K to 400 K.
7 - yxl/+5° - cut (type C vibrations;
8 - yxbl/-19°06'/+34°05' - cut (type B vibrations).

As measurement results show, the function $\omega = \omega$ (T) over the temperature range from 4 K to 400 K reveals such a number of curvature sign changes, which cannot be described by polynominal of degree three. Approximation accuracy of experimental dependencies can by elevated either by increasing approximating polinominal degree, or by splitting all the approximation range in several subranges. Tentative selection of approximating functions by means of a computer has shown that beginning with the terms of degree four checked by Student's distribution quantiles $t_{0.95}$, the significance level of coefficients in the equation of approximated regression turns out to be small compared to the mean weighted dispersion value, connected with incidental observations errors.

Great accuracy in the description of experimental frequency-temperature dependencies is achieved only at the expence of splitting approximation range into subranges: 4 K to 20 K, 20 K to 77.5 K, 77.5 K to 200 K, 200 K to 400 K. Averaged FTCs, approximated by poly nomials of degree three, have coefficients of approximated regression equation.

Experimental analysis

Experimental technique and measuring equipment errors analysis shows, that incidental errors of FTC measured values represent a sum of a large number of independent incidental errors, the influence of any of which at the total sum is small. These errors are: presence of temperature gradient within the volume of a tetrometric assembly, crystal element orientation inaccuracy sample FTC difference of one cut due to impurity influence ets. One can consider as a reult that the incidental error values of measurements made have a distribution near to a normal one. The measurement results are quite characterized by an average value of sampling (frequency maesurements in a sample are equiprecise with the spread of the observed values s^2). For the sprea evaluation it is possible to use obresvations over all the range of temperature measurements. With this, although the average is changing, but the components of incidental error and their contribution do not change practically. The use of mean weighted dispersion for all the measurement range allows to elevate considerably the evaluation accuracy of incidental errors, although in certain cases a comparatively small sampling has been used.

By analysing FTC of the investigated cuts, it is possible to make a conclusion, that the most perspective for the search of optimum angle of cuts with zero TCF at nitrogen temperatures are the cuts yzb/-α with small variations around -48° (type B vibrations) and also yxbl/+ α/-β of the first and the second rotation angles around the values α=12°30', β =-33°, which ensure a very broad gently sloping FTC portion. Further investigation of crystal unit FTCs of different cuts at low temperatures will make it possible to calculate quartz crystal parameters which have not been measured in this temperature region before and as a consequence, to find zero TCF cuts at nitrogen and helium temperatures.

Acknowledgement

Author thanks B. G. Milstein for the help in making measurements.

References

[1]. Smagin A.G. Precision Quartz Resonators. Phisical Grounds, M.: Publ. House of Standards, 1964, 259 p.p. (in Russian).
[2]. Smagin A.G. Scientific session devoted to centenary of A.S.Popov birth. M.: Publ. House VNTOR and E., p.p. 88, 1959; Proc. of VNIISIMS, v. 6, 1962 (in Russian).
[3]. Bechmann R., Ballato A.D., Lukaschek T.J. Proc. IRE, v. 50, p.p. 1812, 1962.
[4]. Smagin A.G., Milstein B.G. Crystallography, v. 20, 1, p.p. 87-94, 1975.
[5]. Fiodorov F.I., Theory of elastic waves in crystals, M.; Publ. House "Science", 1965 (in Russian).
[6]. Smagin A.G., Apparature and technique experiment, n., 6, p.p. 157-159, 1975.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A HIGH SENSITIVITY MECHANICAL LOSS SPECTROMETER

Richard A. Kant and Carmine Carosella

Naval Research Laboratory, Washington, DC 20375
Internet: kant@nrlfs1.nrl.navy.mil

Abstract

NRL scientists have developed an instrument that has extreme sensitivity to energy dissipation in mechanical resonators. It has many potential applications, including detecting and identifying low levels of both chemical impurities and structural defects. While previous instruments measured loss in a resonator to one part in ten, the new instrument measures mechanical loss to one part in 10^4 or better. Thus, the new spectrometer can separate low level signals from background signals that are several orders of magnitude larger.

Background

Chemical and structural defects in solids have a profound influence on the behavior and properties of solids including electrical and thermal conductivity, corrosion resistance, hardness and fatigue life. For many years, researchers have used mechanical loss (or internal friction) spectroscopy to study a wide range of defects including substitutional impurities[1], vacancies[2], interstitials[3], dislocations[4] and grain boundaries[5]. Loss spectroscopy has also proven valuable for investigations of radiation damage[6], diffusion[7], and thermal processes[8]. Today,

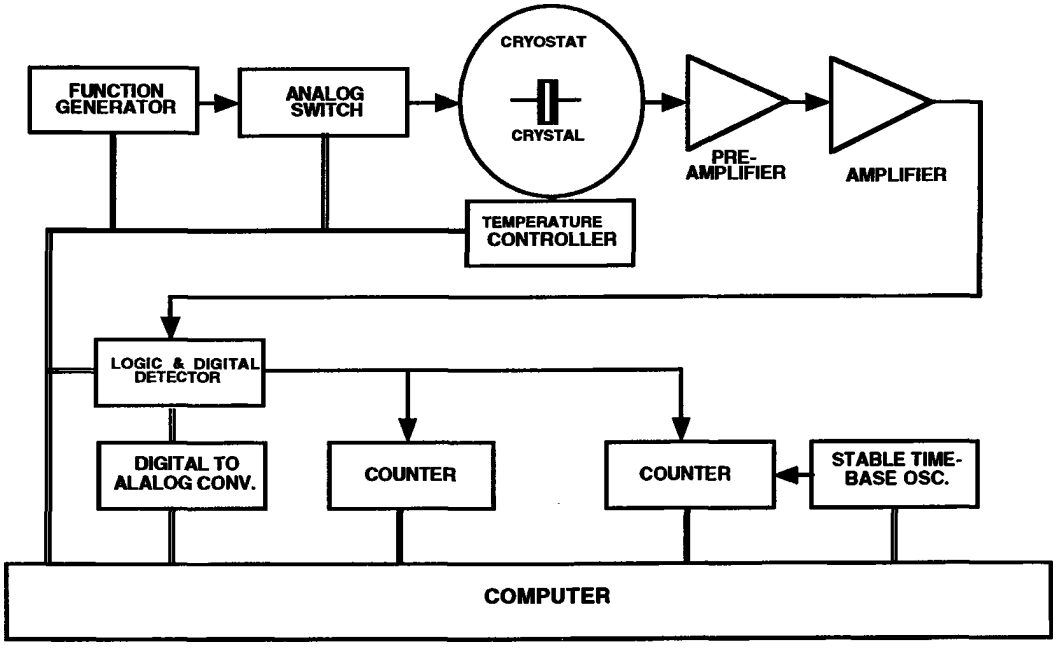


Figure 1. Block diagram of mechanical loss spectrometer.

U.S. Government work not protected by U.S. copyright

there is a renewed interest in loss measurement techniques because of the discovery that the sensitivity and selectivity of chemical sensors are greatly enhanced by information from loss measurements. In addition, researchers have recently begun to use loss measurements in tribology to study atomic level interactions at interfaces between sliding surfaces. However, the sensitivity of conventional techniques limits the resolution in these new applications and the small signals of interest have been difficult or impossible to detect. A new detection circuitry developed at NRL provides a thousand fold increase in resolution and therefore makes it possible to isolate and measure effects that were previously undetectable.

Loss spectroscopy is based on the fact that defects absorb energy produced by external stresses and dissipates it as heat. A loss spectrum is made by measuring energy loss as a function of temperature while mechanically stressing a sample at a fixed frequency. Peaks appear in a spectrum when the relaxation time of a defect is comparable to the period of the applied stress. In order to make a loss measurement, a resonator is made to vibrate near a resonant frequency. In the case of quartz crystal resonators, this is accomplished by delivering the driving signal to metal electrodes deposited directly on the quartz. When the drive signal is switched off, the amplitude of the oscillations decreases due to energy dissipated by defects in the resonator. The loss per cycle is then determined from the decay rate of the vibration amplitude. A detector circuit measures the decay of peak amplitude as a function of the number of cycles executed and the loss is calculated from the slope of a fit to the logarithm of the decay signal. We achieve extremely high sensitivity by means of a new detector circuit and the associated software.

The NRL loss spectrometer

Fig. 1. is a block diagram of the NRL high resolution loss spectrometer. A computer controls all operations of the instrument and performs data acquisition and analysis. Specimens are mounted in a vacuum cryostat and maintained at any temperature from 4 K to 600 K, to within ±1 mK. Signals from the resonator are amplified and fed to a digital detector circuit invented at NRL which, in turn, gates two counters. One counter determines the number of vibrations executed between each of a number of threshold amplitudes and the other records the elapsed time for determining the resonant frequency.

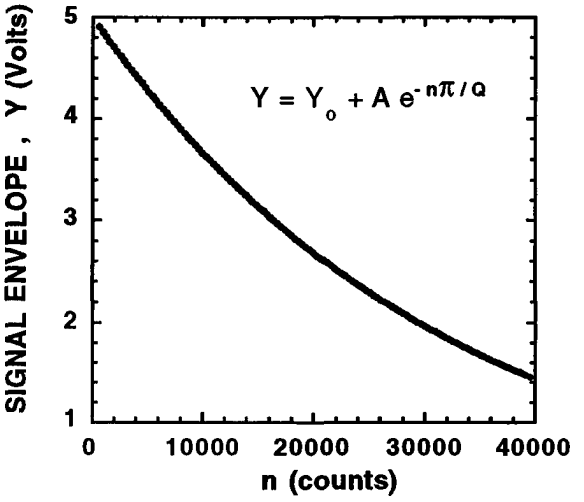


Figure 2. Experimental data for a typical decay envelope.

If a single relaxation process dominates the behavior of the decay, then the experimental data are indistinguishable from a least squares fit of the data (see Fig. 2). Figure 3 shows that

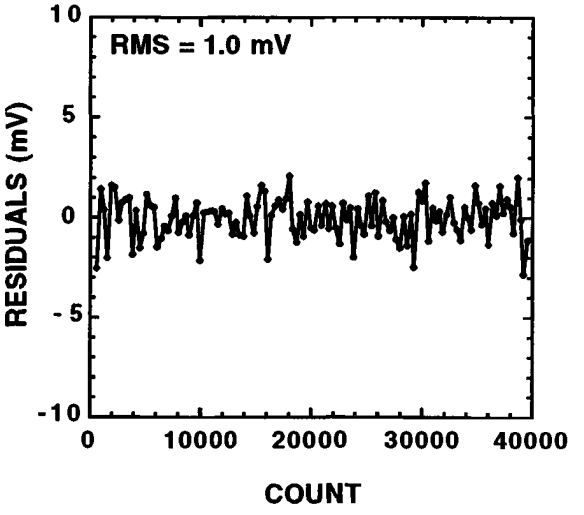


Figure 3. Deviation of experimental data without interfering mode present.

the experimental data points differ from the theoretical curve by less than one part in a thousand. Changes in the deviation from the best fit or unusually large deviations indicate the presence of multiple relaxation phenomena and additional analysis is needed. For example, figure 4 shows modulation of the envelope observed when a second, interfering, mode is present. Fits of selected portions of each decay curve can give multiple loss rates.

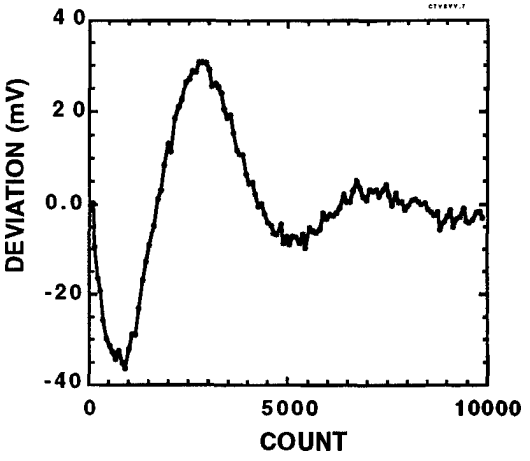


Figure 4. Modulation of the envelope indicates presence of a secondary vibration mode.

Spectrometer Performance

The high resolution achieved with the NRL loss spectrometer is illustrated in Figures 5 and 6. These spectra are from a high purity, low defect density quartz and are normalized to the peak loss shown in Figure 5. The resolution of figure 5 is consisten with the output from a conventiona spectrometer and reveals no small features or structure.

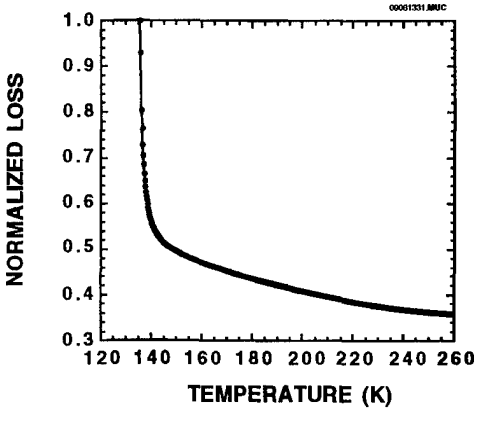


Figure 5 Low resolution The loss spectrum appears relatively featureless at temperatures above 160 K.

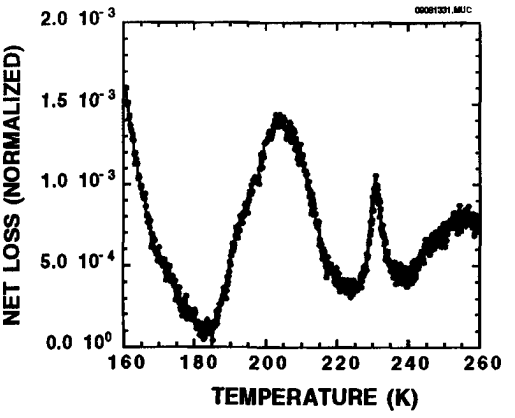


Figure 6. Fine structure in the loss spectrum revealed by high resolution spectrometer.

By contrast, Fig. 6. is a high resolution version of a portion of the same spectrum with the background subtracted. Note that the amplitude of the loss peak near 230 K in Fig. 6 is a small fraction (7×10^{-4}) of the peak loss in figure 5 and yet this peak is easily resolved from the overall background loss in that portion of the spectrum. This illustrates the sensitivity achieved with raw data from the NRL mechanical spectrometer. Additional sensitivity is available using filtering and signal averaging.

Relaxation phenomena with very large activation energies have high-temperature loss peaks, but raising the temperature too high can significantly alter or damage the host material. High resolution spectroscopy reveals structure not resolvable by conventional spectrometers.

The time dependence of loss at a number of temperature points well below the loss peak will yield, for example, the high temperature activation energy. The high sensitivity of the NRL spectrometer make it well suited to make

such measurements. In fact, we have measured loss changes as small as 10^{-9} per hour using the new instrument. The study of such high activation energy relaxation processes provides a way to increase our understanding of long term instabilities in quartz and other technologically important materials. In summary, the new high sensitivity NRL loss spectrometer opens the way to a the study of a new range of defect densities and dynamical effects and it may serve as the basis for a new generation of ultra sensitive chemical sensors.

References

1. N.E. Byer and H.S. Sack, (1966). *Phys. Rev. Lett.* **17**, 72.

2. A. Scholtz and A. Seeger, (1963). *Phys. Status Solidi.***3**, 1480.

3. B.S. Berry, (1970). *J. Phys. Chem. Solids* **31**,1827

4. P.G. Bordoni, (1954). *J. Acoust. Soc. Amer.* **26**, 495

5. T.S. Ke, (1947). *Phys. Rev.* **71**, 533.

6. R.M. Stern and A.V.Granato, (1962). *Acta Met.* **10**, 358

7. J.L. Snoek, (1942) *Physica* .**9**, 862

8. H.R. Randall, F.C. Rose, and C. Zener. (1939). *Phys. Rev.* **56**, 343

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Microresonator Sensor Arrays

John R. Vig*, Raymond L. Filler* and Yoonkee Kim**
*U.S. Army Research Laboratory, Fort Monmouth, NJ 07703-5601
**National Research Council - Resident Research Associate
EMail: JVig@ARL.mil

Abstract

A high-sensitivity sensor and sensor array is proposed which can be used for infrared (IR) detection and imaging, and for chemical, biological agent and other types of sensing. The sensing elements consist of microresonators. Calculation shows that an array of microresonators in the 200 MHz to 1 GHz range can be the basis of an uncooled IR imaging system with a noise equivalent temperature difference, NETD, of < 0.01 K which is superior to the (actual) performance of state-of-the-art uncooled imaging arrays and is competitive with cryogenically cooled imaging arrays. The frequencies of microresonators are sensitive to both the mass loading and the temperature changes produced by adsorption/desorption and chemical reactions. That the two frequency changes are independent can be exploited for chemical and biological sensors which are both sensitive and selective. The design and fabrication problems to be overcome before such microresonator arrays can be realized are discussed.

Introduction

Major advances have been made in the miniaturization of most types of electronic components during the past 30 years. Although small resonators have been developed during this period (e.g., tuning forks, strip AT-cuts, etc.), the dimensions of precision quartz resonators have not been reduced significantly. That fundamental mode 1 GHz AT-cut and SC-cut quartz resonators can be fabricated by means of chemical polishing techniques has been demonstrated [1-4], however, these resonators are not useful for precision frequency control applications. The primary reason is that such resonators are extremely sensitive to mass loading which makes frequency adjustment difficult, and degrades the stability, especially the aging. For example, a 1.6 GHz fundamental mode AT-cut resonator [2] is 1 micrometer thick, which corresponds to about 2,000 molecular layers. Therefore, a thickness change of a single molecular layer, e.g., due to contamination adsorption, changes the frequency by 1 part in 2,000, i.e., by 500 ppm.

For sensing, however, the absolute frequency and the aging are not important. The important factors are the frequency change caused by a measurand, and how accurately that frequency change can be measured. In this paper, the potential of microresonator arrays as IR sensors, and as chemical and biological sensors is explored. It is shown that, combining the low noise characteristics of quartz crystal oscillators, the steep frequency vs. temperature characteristics of resonators made of certain cuts of quartz, and the small thermal mass and high thermal isolation capability of microresonators can result in high performance sensors and sensor arrays.

Infrared Sensors

Infrared (IR) sensors [5-9] require no visible light in order to "see" in the dark. They operate by sensing the IR emitted by all objects. IR sensors can be divided into two categories, photon detectors, which are usually cooled, and thermal detectors, which may, but need not, be cooled. The sensing elements in photon detectors may be photoconductors, Schottky diodes, quantum well infrared photodetectors (QWIP), or photovoltaic devices. In the widely used thermal detectors, the elements have been resistive bolometers, ferroelectric bolometers, pyroelectric devices, or thermocouples. In this paper, it is shown that microresonator-based IR sensors can, potentially, equal or surpass the performance of other

U.S. Government work not protected by U.S. copyright

types of IR sensors, whether they be photon detectors or thermal detectors. The microresonator IR sensor belongs in the category of thermal detectors.

The performance of any thermal detector is determined mainly by: 1) the steady-state response, given by:

$$\Delta T = \frac{P}{G} \quad (1)$$

where P = power absorbed from the heat source, ΔT = the temperature rise due to P, and G = the thermal conductance from the sensitive element to a heat sink; 2) the noise of the detector relative to the signal produced by ΔT, which limits the degree to which the ΔT can be resolved; and 3) the time constant τ_T of the detector element, given by:

$$\tau_T = \frac{C}{G} \quad (2)$$

where C is the heat capacity of the element.

For an IR sensor array, the most important parameters are the microresonators':
1) temperature coefficient of frequency
2) noise
3) thermal conductance to a heat sink
4) heat capacity
5) IR absorbance
6) dimensions, and
7) fill factor (*i.e.*, the ratio of sensing area to total area of the array).

Temperature Coefficients

Microresonators may be fabricated from a large variety of (bulk or thin film) piezoelectric materials, especially since the materials need not possess temperature-compensated cuts. Although materials other than quartz (e.g., ZnO, GaAs, etc.) may, eventually, prove to be more suitable than quartz for microresonator sensors, the discussion that follows shall be confined to quartz microresonators.

The possibilities of using quartz resonators as precision thermometers have been described since at least 1962 [10-21]. The ability of quartz thermometers to measure small temperature changes was discussed by Smith and Spencer in 1963 [13]. They pointed out that it is possible to sense temperature changes of microkelvins. Quartz resonators have also been proposed for IR sensing [15,16,18,20]. The typical dimensions of the resonators in the quartz thermometers and IR sensors were about 1 cm in diameter and a fraction of a millimeter in thickness, *i.e.*, they were not microresonators.

Quartz resonators' frequency can vary with temperature monotonically with a slope of about 10^{-4}/°K. Due to the low noise capabilities of crystal oscillators, the frequency noise limitation for resolving temperature changes is, for example, $\leq 10^{-12}$ for a low-noise 10 MHz resonator. Therefore, the noise of such a resonator corresponds to temperature fluctuations of ≤10 nanokelvins.

The resonator cut for microresonator sensors may be chosen from among a wide variety. Some of the possibilities are the: AC-cut (20 ppm/°C [10]), LC-cut (35.4 ppm/°C) [14], Y-cut (~90 ppm/°C [10]), SC-cut (b-mode: -25.5 ppm/°C [22], or dual-mode: 80 ppm/°C to >100 ppm/°C [23,24]), NLSC-cut (~14 ppm/°C) [25], or any other cut that can be made thin and have a well-behaved temperature sensitive mode. The NLSC-cut, the dual-mode SC-cut and other thermometer cuts from the stress-compensated-cut locus [26] will be especially advantageous when the pixels are switched on and off for short periods, e.g., when the array is divided into subarrays and one oscillator circuit per subarray is used to sequentially excite the resonators in the subarray; or when the microresonators are illuminated intermittently, as would be the case when the microresonators are used in a scanned array.

Noise

The noise floor (*i.e.*, the Allan deviation floor) of a state-of-the-art AT-cut or SC-cut resonator can be described by $\sigma_y(\tau) \approx (2\times10^{-7})\ Q^{-1} \approx 1.2\times10^{-20}\ Hz^{-1}\cdot f_0$ [27-29], where $\sigma_y(\tau)$ is the Allan deviation floor and f_0 is the resonator frequency (in Hz). The best (bulk acoustic wave, BAW) noise performance has been achieved with overtone resonators. Microresonator

sensor arrays seem most promising when the resonators are in the 200 MHz to 1.0 GHz range, as is shown below. Noise data are available for 400 MHz to 900 MHz surface acoustic wave (SAW) resonators, and for these, at the noise floor, $\sigma_y(\tau)$ was in the range 1×10^{-11} to 1×10^{-10} [30] (after converting the data from the frequency domain to the time domain). For 100 MHz 3rd overtone and 160 MHz 5th overtone BAW (SC-cut) resonators, the noise floor corresponds to 1.2 to 1.9×10^{-20} $Hz^{-1}\cdot f_0$ [31].

The noise of microresonator sensors may be adversely affected by size effects [32,33] and by contamination adsorption and desorption [34]. For purposes of calculating the properties of microresonator sensors, we assume that the noise floor will be 10× worse than 1.2×10^{-20} $Hz^{-1}\cdot f_0$, *i.e.*, we assume that the noise will be 1.2×10^{-19} $Hz^{-1}\cdot f_0$. This predicts, for example, that the noise floor of an 800 MHz microresonator will be 1×10^{-10}.

A noise contribution which is negligibly small in ordinary resonators, but is the dominant noise contribution in microresonators below about 2 μm in thickness, is the temperature fluctuation noise. This noise is due to the quantum nature of heat exchange, i.e., that heat exchange takes place by the absorption and emission of photons and phonons. The smaller the heat capacity (i.e., the higher the frequency) of a microresonator, the larger the temperature fluctuations due to this noise source. It can be shown that the mean temperature fluctuation, ΔT, of any object due to this noise is given by

$$\Delta T = \sqrt{\frac{\kappa_B T^2}{C}} \qquad (3)$$

where C is the heat capacity and κ_B is Boltzmann's constant [5,8]. This temperature fluctuation noise manifests itself as frequency noise, via the microresonators' temperature coefficient, and adds to the other oscillator noise sources, as is shown in Appendix A.

Frequency Measurement

The power required to simultaneously excite all microresonators in a large array may be excessive in many applications. Dividing the array into subarrays and exciting the resonators sequentially, with one oscillator per subarray, can reduce the power requirement to manageable levels. One resonator that is shielded from the IR source (or from the adsorbing species when the array is used for chemical or biological sensing) can serve as a reference, and the frequency differences between that reference and the other resonators can be measured rapidly.

In a conventional crystal oscillator, how rapidly the frequencies may be measured is a function of the resonators' noise characteristics. The shortest measurement time, τ, which one may use for frequency measurement without a loss of resolution is the τ at the "knee" of the $\sigma_y(\tau)$ vs. τ curve, i.e., it is the lower τ limit of the noise floor. At this limit, other noise processes, such as white phase noise or Johnson noise (which is white FM noise) begin to exceed the flicker of frequency noise, so that $\sigma_y(\tau)$ starts to increase as τ decreases. The τ at the knee of the $\sigma_y(\tau)$ vs. τ curve is a function of the drive level, so that in the design of a microresonator array, it will be necessary to balance the drive level, self-heating, and the measurement time.

For example, for a 500 MHz SAW resonator-oscillator, the knee of the $\sigma_y(\tau)$ vs. τ curve was at about 10^{-3} s when a 40 MHz high-frequency cut-off (f_c) was used for measuring at small τ's [35]. Assuming that with the same f_c, the knee of a 500 MHz microresonator's $\sigma_y(\tau)$ vs. τ curve would be at the same τ, and using an f_c of ~40 kHz, one may lower the knee to a τ of about 30 μs (because $\sigma_y(\tau)$ scales as $(f_c)^{1/2}$ in the white phase noise region) [36]. Therefore, except for the effect of dead time which is described in Appendix A, one can, in principle, sequentially excite each resonator in a 16×20 subarray with one oscillator in a 33 ms video frame time, so that 252 oscillators (and counters) would be required to excite (and measure) a 240×336 array. Measuring the 320 resonators in 33 ms allows 103 μs per reso-

nator for the switching, startup and measurement of each resonator.

Whether or not one may lower the τ to 30 μs without increasing the noise depends on the resonator frequency the drive level and the dead time. For example, measuring a 500 MHz microresonator's frequency at a 1 microwatt drive level and with τ = 30 μs increases the oscillator noise by about a factor of 10 because, as is shown in Appendix A, the Johnson noise becomes significant at low drive levels and short measurement times. However, at 500 MHz (and higher frequencies), the temperature fluctuation noise is larger than the resonator flicker noise, so the total noise increases by less than a factor of 10 over the case when the Johnson noise is negligible (i.e., when flicker noise is the dominant oscillator noise source). Therefore, in designing a scanned array, one must balance the number of resonators per subarray with the drive level and the NETD. The larger the number of resonators in a subarray, the higher the NETD for a given drive level.

Thermal Isolation

From Eq. (1) it follows that, to maximize the response of an IR detector to a given IR source, the thermal conductance, G, from the sensitive element to a heat sink must be minimized. On the other hand, Eq. (2) shows that minimizing G for a given C also maximizes τ. Reducing C reduces τ, but increases the temperature fluctuation noise, as can be seen from Eq. (3). Therefore, in designing a microresonator array, one must find the proper balance between G and C.

The thermal isolation of microresonators may be controlled in a variety of ways. The ultimate isolation can be achieved by levitating [37-40] the microresonators in a vacuum and combining the levitation with electrodeless excitation [29]. In such an arrangement, heat exchange would be by means of radiation only. The resulting long time constants and the extra complexity would make such an array unsuitable for most applications (although the time constant could be controlled by admitting a controlled amount of, e.g., helium).

A controlled amount of thermal isolation may also be produced during the etching of the microresonators. For example, one may form thin and narrow bridges, or thin and narrow rings to surround the resonators, the rings being connected to the resonators (and to each other, for multiple rings) by bridges. Figure 1 illustrates square resonators isolated by means of bridges at each corner. Figure 2 illustrates circular resonators, each isolated by means of three evenly spaced bridges. A third way to produce the thermal isolation and mechanical supports is to use free-standing thin film strip supports [41]. As, for example, an 800 MHz microresonator's mass is only about 30 nanograms, the supports need not be massive.

For purposes of calculating the figures of merit of microresonator IR sensor arrays, bridge supports are assumed. The bridges' dimensions are selected to provide a G that results in a time constant of 10 ms, i.e., τ is fixed at 10 ms, and G is a variable that is a

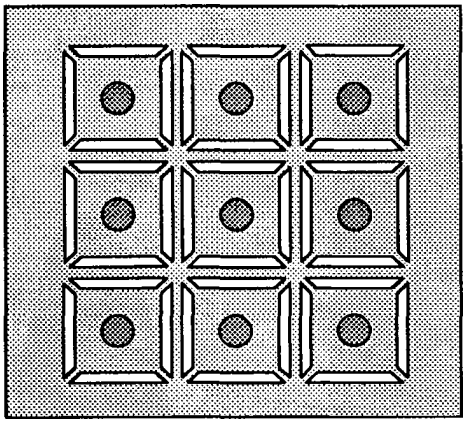


Figure 1. Thermal isolation of square microresonators.

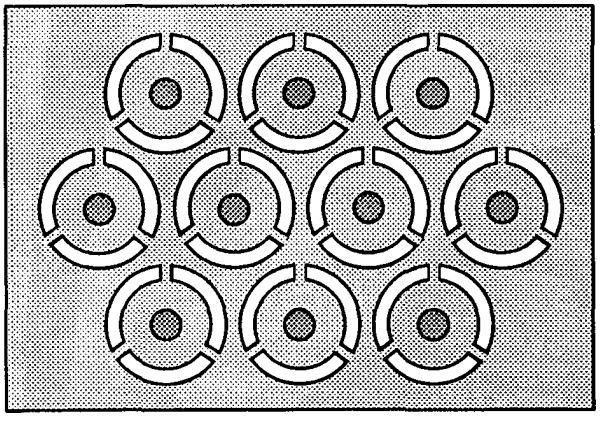


Figure 2. Thermal isolation of circular microresonators.

function of resonator frequency. The calculation of bridge dimensions is described in Appendix A.

The IR Absorption of Microresonators

Quartz is a good absorber of IR near 9, 12, 20, and 26 μm [42]. A theoretical calculation over the 8-14 μm wavelength range was performed using refractive indices in Palik [43]. It indicates that quartz of 2.5 μm thickness has about 50% absorption peaks near 9.5 μm and 12.5 μm. The average energy that would be absorbed by a 2.5 μm thick quartz plate in the 8 μm to 14 μm IR band is about 17%. The overall IR absorption over the wavelengths can be increased by applying IR absorbing coatings.

A wide variety of IR absorbing coatings is available [44-50]. Thin films that have a sheet resistance which is half the impedance of free space, i.e., 188 Ω per square, can absorb 50% of incident IR radiation [44]. (The impedance of free space is 377 Ω.) At such an optimum film thickness, 50% of the incident IR radiation is absorbed, 25% is reflected, and 25% is transmitted. The absorption can be increased to > 50% by creating multiple passes of the IR through front of the enclosure and a reflecting film on the resonator, e.g., by having an antireflection film on the back side. Making the enclosure into a resonant optical cavity is one way to obtain high absorption. According to Lang, et al. [45], at a critical thickness of 17 nm (34 μg/cm^2), a vacuum evaporated gold film has about a 0.5 absorption. If the film is thicker, it reflects the IR (the electron mobility is high and the electric field parallel to the surface vanishes); if the film is thinner, it is transparent (there are not enough electrons with which to interact), and at the critical thickness, the electrons interact with the incoming wave but are not free to move, hence the optimum absorption. An 8 nm nichrome film is also a good absorber, and has been used on some thermal detectors.

A three-layer absorber stack consisting of a metal film of 377 Ω/square on the side facing the IR, a dielectric layer of refractive index n and thickness d = $\lambda_{max}/4n$, and an IR reflecting film on the back side has been shown to be capable of > 95% absorption at λ_{max} [46]. For example, an effective absorption of 90% can be achieved from 8-13 μm when λ_{max} = 10 μm, and nd = 2.5 μm. The sheet resistance of the front surface film is not critically important; 95% or greater absorption has been achieved within the range of 300-600 Ω/square. The absorption outside the ±2.5 μm bandwidth decreases rapidly, especially on the shorter wavelength side, e.g., the absorption drops to about 10% at 5 μm.

A good absorber system may consist of IR absorbing coatings on either or both major surfaces of the resonators and a good IR reflector on the inside surface of package (on the side away from the IR source, of course). This structure, which provides an optical cavity to capture incident IR energy, can be optimized to obtain maximum IR absorption. Ellipsometry theory [43] may be applied to analyze this multilayered structure. However, the material constants of IR absorbing films strongly depend on fabrication conditions and, therefore, cannot be predicted accurately from the literature. Experiments will need to be performed to design an optimal structure.

Porous films are better absorbers than well-ordered films; e.g., gold evaporated thermally in a 100 Pa nitrogen atmosphere results in a film structure which is "a conglomerate of very fine needles with linear dimensions of about 25 nm [45]." Such films have an absorption near 1 at 250 to 500 μg/cm^2, and about 0.7 at ~40 μg/cm^2. Similarly deposited silver films show a higher absorption at the same mass than the gold films; absorption is > 99% at 80 μg/cm^2. Lang cautions that electromigration is a potential problem for silver, however, this ought not to be a problem for a resonator, where the films can be kept free of DC potentials.

The IR absorbing film can be deposited so as to contribute to energy trapping. Mass loading that is equivalent to about 1/50 of the plate thickness could provide the proper energy trapping [51]. (At 800 MHz, for example, 1/50 of the plate thickness is ~40 nm of quartz-equivalent mass loading.) Additional absorbing film can be outside the active area of the resonator. If porous gold is too heavy to use in the active area, one can select an appropriate film from a

variety of lighter films that are available, e.g., porous aluminum, nichrome, iron, magnesium, copper, carbon, etc. Aluminum is suitable from the mass loading point of view, but it has high reflectance. However, its reflectance drops rapidly below 10 nm of film thickness; for 8 μm IR radiation, the transmission vs. thickness and the absorption vs. thickness curves cross at ~6 nm film thickness and the values are ~30% for each [48]. If the back side of the enclosure is reflecting (e.g., an Al film > 15 nm thick has ~100% reflectance to IR wavelengths up to at least 50 μm), then two passes of the IR result in ~64% absorption. For 1.6 GHz AT-cut resonators, ~40 nm thick aluminum electrodes were about the lowest mass electrodes which could be used [2,52]. Below 40 nm, the electrodes' resistivity was too high.

Simultaneously obtaining sufficient energy trapping and high IR absorption presents a challenging design and fabrication problem. Some possibilities are: using a ring electrode [51,53,54] which surrounds an IR absorbing film on the side facing the IR radiation and either a ring or a solid (and reflecting) electrode on the back side; using other arrangements of electrodes with opening(s) for allowing the IR radiation to pass through; using lateral field excitation [55-57] such that the exciting electrodes are on the side of the resonators facing away from the IR source and the IR absorbing film is on the front side; using electrodeless excitation [29,58,59], with the electrode facing the IR source being a ring electrode [51]; and etching energy trapping mesas into the microresonators [51].

The film thickness must be controlled in order to obtain maximum absorption. Control can be achieved by either the usual monitoring of the frequencies during deposition, or by monitoring IR transmission, e.g., with an IR LED on one side and a photodetector on the other side of the resonator (wafer).

At the higher value (or with a thicker than 1.6 GHz plate), the quartz plate itself could become a quarter wave plate for ~10 μm IR, and thereby help produce a strong absorption at the 10 μm range. At the lower value, and for thinner plates, the reflector can be deposited on the inside surface of the back of the enclosure such that the quartz plus the gap between the quartz and the back of the enclosure form a quarter-wave dielectric layer.

IR Detector Performance Measures

The noise-equivalent power, NEP, of a detector is defined as the amount of incident root mean square (rms) signal power which causes an output signal-to-noise ratio of unity.

The two commonly used performance measures of IR detectors and IR imaging arrays, respectively, are D* and NETD (also called NEΔT).

The specific detectivity, D*, is defined by

$$D^* = \left(\frac{\sqrt{A \cdot B}}{NEP}\right) \tag{5}$$

where A is the area of a pixel (i.e., of a microresonator), and B is the measurement bandwidth. The unit of D* is cm $Hz^{1/2}$/Watt.

The noise-equivalent temperature difference, NETD, is defined as the change in temperature of a large blackbody source within the field of view of an IR imaging system for which the detected signal is equal to the system noise. Figure 3 shows NETD as a function of microresonator frequency for two measurement conditions: (a) 30 μsec τ and 30 Hz frame rate ($r \sim 1000$) and (b) 30 msec τ and 30 Hz frame rate ($r \sim 1$) (See Appendix A for the definition of r and the details of the calculations of D* and NETD). Also shown in Figure 3 are the individual contributions to NETD of the oscillator noise and the temperature fluctuation noise. Figure 4 shows NETD as functions of G (and τ_T) for a 600 MHz array, down to the radiation thermal conductance limit of 1.3×10^{-4} K. Table 1 shows a summary of NETD, quartz wafer area needed for a 240×336 array, and the bridge length needed for a 10 ms time constant for 100 MHz, 200 MHz, 600 MHz, 1.0 GHz, 1.4 GHz and 1.8 GHz arrays. In the calculations, the following assumptions were made: T = 300 K, τ_T = 10 ms (except in Figure 4, where G, and therefore τ_T, is a variable), diameter to thickness ratio = 80, frequency constant = 1.6 MHz·mm, temperature coefficient =

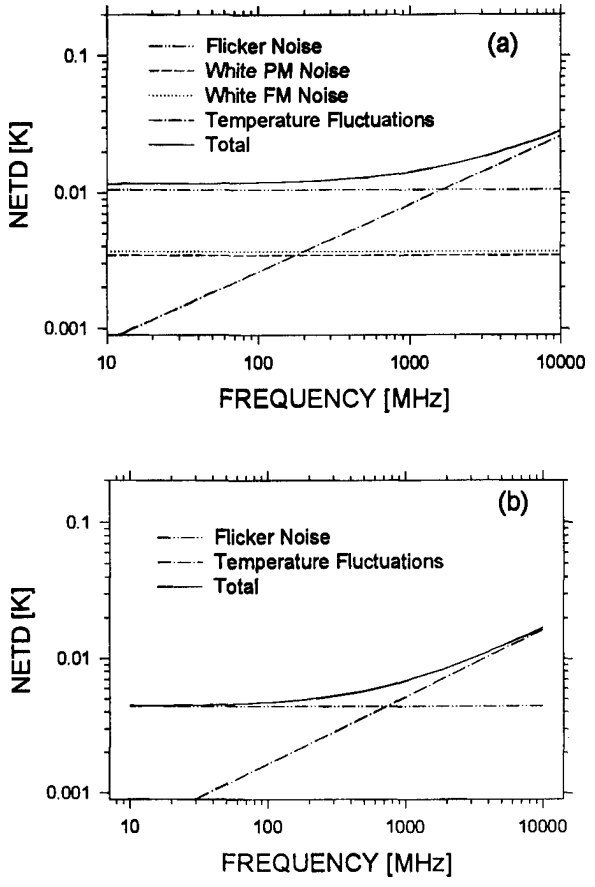


Figure 3. NETD as a function of microresonator frequency at (a) r=1000, (b) r=1. White PM and white FM noise contributions are negligibly small for r=1.

50 ppm/K, oscillator noise at the noise floor = 1.2×10^{-19} $Hz^{-1}\cdot f_0$, no chopping, transmittance of the atmosphere between the target and the sensor = 0.8, 56% IR absorption, 80% fill factor, f/1 optics, and the wavelength range of 8-14 μm. Of course, cooling the array improves the performance (as it does for all IR detectors). However, using subarrays at 600 MHz and assuming that flicker noise is not affected by temperature, the NETD is reduced from 0.013 K at T = 300 K to only 0.012 K at T = 175 K (which can be achieved with thermoelectric coolers).

For comparison, the best uncooled IR imaging arrays that have been reported are the silicon microbolometer array [60], with NETD = 0.04 K, and the ferroelectric bolometer array [61], with NETD = 0.05 K. The best available cryogenically cooled IR imaging

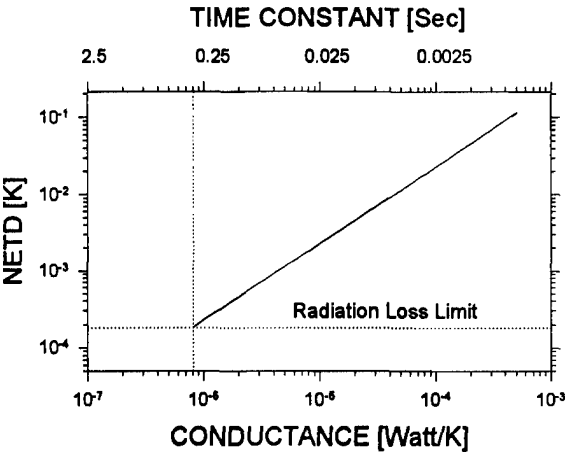


Figure 4. NETD as a function of G (and τ) for a 600 MHz microresonator array at r = 1.

Table 1. Comparison of key parameters of 100 MHz, 200 MHz, 600 MHz, 1.0 GHz, 1.4 GHz and 1.8 GHz arrays with 240×336 pixels.

Frequency [MHz]	NETD [K]		Wafer size [cm×cm]	Bridge length [μm]
	r=1000	r=1		
100	0.011	0.0047	-	1.3
200	0.012	0.0050	-	2.7
600	0.013	0.0059	6.4×9.0	8.1
1000	0.014	0.0068	3.9×5.4	13.6
1400	0.015	0.0076	2.8×3.9	19.4
1800	0.016	0.0083	2.2×3.0	25.3

arrays (HgCdTe, InSb, PtSi and GaAs/AlGaAs) exhibit NETD = 0.01 K [62]. Typical thermal imaging cameras exhibit NETDs of 0.02 to 0.2 [63]. (To be fair, it must be pointed out that the NETD values used for comparison are measured values, whereas the microresonator array NETD values are calculated ones, based on some known and some estimated properties of quartz microresonators, i.e., no microresonator arrays have yet been built. The calculated NETDs assume that the design and fabrication problems can be solved without degrading the NETD.)

Chemical and Biological Agent Sensing

Quartz crystal microbalances (QCM), which exploit the dependence of quartz resonator frequency on mass loading, have been used for a wide variety of applications [64]. These have included thickness monitoring in thin film deposition, adsorp-

tion/desorption studies, humidity sensing, oxidation studies, corrosion studies, chemical and biological sensing [65], etc. Calorimeters [66,67] have been used to sense the heat produced during adsorption, desorption, and chemical reactions on surfaces.

Because resonators can be sensitive to both mass and temperature changes, a resonator can act as a QCM and calorimeter, simultaneously. By applying different thin films to the different resonators in an array and observing the pattern of frequency changes due to an unknown that is admitted into the resonator array enclosure, it ought to be possible to detect and identify chemical and biological agents. The frequency changes of individual resonators will be the sum of the changes due to mass loading and due to the temperature changes resulting from the heats of adsorptions or reactions. Dual mode SC-cut resonators [24] may be especially useful for this application, as one of the c-modes can be used for sensing the mass loading, while the beat frequency is used as the temperature sensor.

Figure 5 shows an example of the detectability limits as a function of resonator frequency for an adsorption process with an activation energy of 25 Kcal/mol. The calculation details are shown in Appendix B. The detectability limits are in units of percent of a monolayer, and the example assumes that: the density of the adsorbed layer is the same as the density of quartz, the measurement time is equal to τ_T, and the noise is unaffected by the presence of the agents to be sensed (which may be correct at low pressures, but not at high gas pressures or in liquids which would significantly degrade the resonators' Q).

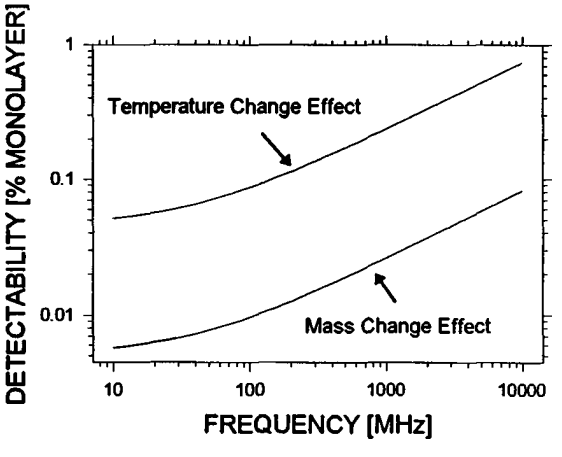


Figure 5. Example of detectability limits as a function of resonator frequency for mass change and temperature change due to an adsorption process with an adsorption energy of 25 Kcal/mol.

Whereas, for IR imaging arrays, large arrays are desirable, for chemical and biological sensing, small arrays can be useful, and such arrays can be produced at frequencies where resonators are currently being mass-produced at low cost. Therefore, it follows from Figure 5 that microresonator arrays for chemical and biological sensor applications are within the state of the art. Based on the prices of mass-produced small resonators (e.g., the tuning-forks), microresonator arrays could be inexpensive enough to be used just once, e.g., in medical diagnostic applications, and then be either discarded or recycled.

Possible other applications of microresonator arrays, using the appropriate absorbing coatings, include: UV, X-ray, and millimeter wave detection and imaging, and multifunction sensor arrays.

Producibility and Other Challenges

Small arrays are producible with currently available technology. For example, strip AT-cut resonators are currently being produced by photolithography and etching techniques at frequencies up to 250 MHz [68]. Arrays of 250 MHz and even lower frequency resonators can be useful for scanned IR imaging arrays and for chemical and biological sensors. However, the lower the frequency, the larger is each resonator, and the fewer the resonators per wafer, as is illustrated in Figure 6. The example in Figure 6 shows the wafer size needed for a 240x336 array (i.e., 80,640 resonators per wafer). As the largest quartz wafers currently available are 10 cm×10 cm, it is clear that, for 240x336 arrays, the frequencies of the microresonators must be above about 500 MHz.

Several problems must be overcome before large arrays of microresonators may be realized. The techniques being used for producing 250 MHz strip AT-

cut resonators will need to be extended to higher frequencies, or new methods will need to be developed. Low defect density quartz will need to be used in order to achieve high yields. Maintaining sufficient surface parallelness is just one of the fabrication challenges. The arrangement of IR absorbing films and resonator electrodes will require careful design, as was discussed earlier.

Plating the microresonators to frequency is not a problem because the exact frequencies are not important. Moreover, not having all the microresonators at the same frequency will be an advantage because it will reduce the potential crosstalk among resonators. Packaging, making the necessary interconnections, dealing with the reactances of the structure, minimizing crosstalk, and making, driving, and measuring the resonators in the array so that they exhibit sufficiently low noise are challenging design and fabrication problems.

Summary and Conclusions

Calculations based on the properties of quartz resonators indicate that high frequency microresonator arrays can, potentially, result in high performance sensors. In optimizing the performance of a microresonator sensor array, one must find the proper balance between the thermal conductance from each microresonator to a heat sink and the resonators' heat capacity (which is determined by the resonator frequency). Another important balance that needs to be made is between resonator drive level (and the resulting self-heating) and the noise characteristics (i.e., the optimum frequency measurement method).

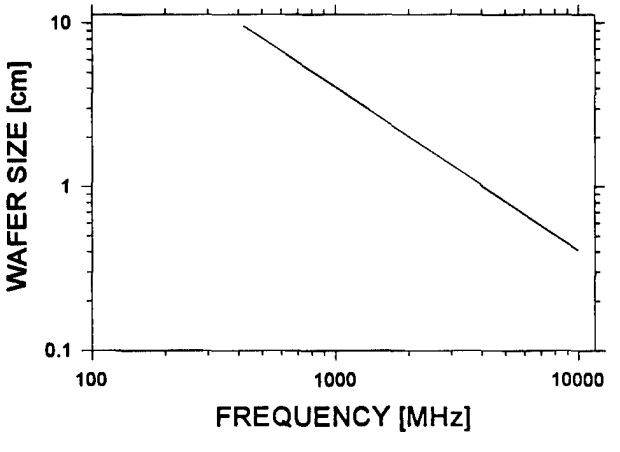


Figure 6. Wafer size required for a 240×336 pixels array as a function of resonator frequency, assuming a square wafer, a diameter to thickness ratio of 80, an 80% fill factor, and a frequency constant of 1.6 MHz mm.

For IR imaging arrays, one must also find the balance among performance, wafer size, and ease of fabrication. The calculated NETD at 300 K is ≤ 0.01 K when the resonator frequency is ≤ 1 GHz. At lower frequencies, the performance and ease of fabrication improve, but the number of microresonators per unit area decreases. The measured NETDs of uncooled IR imaging arrays made with other technologies are ≥ 0.04 K. (The calculated microresonator array NETDs assume that the design and fabrication problems can be solved without degrading the NETD.)

Because resonators can be sensitive to both mass and temperature changes, a resonator can act as a highly sensitive quartz crystal microbalance and calorimeter, simultaneously. By applying different thin films to the resonators in a resonator array and observing the pattern of frequency changes due to an unknown that is admitted into the resonator array enclosure, it is possible to detect and identify chemical and biological agents. The frequency changes of individual resonators will be the sum of the changes due to mass loading and due to temperature changes resulting from the heats of adsorptions or reactions. Calculation shows that small fractions of a monolayer of chemical and biological agents can be detected and identified with microresonator arrays.

Although the design and fabrication of microresonator arrays present many challenges, if the obstacles can be overcome, microresonator arrays can, potentially, result in high performance sensors.

Acknowledgments

The authors acknowledge and thank Paul Kruse for his helpful comments on a preliminary report on this work and on a draft of this paper, and for his book [5] from which we learned some of the basics of IR detector technology; Fred Walls and Thomas Parker, NIST, Charles Greenhall, JPL, and David Allan, Al-

lan's TIME, for their helpful comments on the noise aspects; Prof. Harry Tiersten of RPI for suggesting the use of ring electrodes and for analysis of the effects of displacement of levitated electrodeless resonators [69]; and Al Benjaminson for suggesting circuitry which may be used to excite the microresonators.

Appendix A - IR Sensor Calculations

The theoretical calculation for various figures of merit of IR detectors as a function of resonator frequency is shown in this appendix.

The thermal time constant of the detector, τ_T, is given by [5]

$$\tau_T = \frac{C}{G} \tag{A.1}$$

where C is the heat capacity of the detector and G is the thermal conductance due to both conduction via the mounting structure and radiation from the surface of the resonator.

The heat capacity of the quartz microresonator, C, is

$$C = c_v V \tag{A.2}$$

where c_v = 2.08 W·sec/cm^3·K is the specific heat of quartz and V is the volume of the resonator.

When the mounting structure consists of quartz bridges,

$$G = \frac{k_q nwt}{L} + 4\sigma T^3 (2A) \tag{A.3}$$

where k_q = 0.07 W/cm·K is the assumed thermal conductivity of the quartz bridges, $\sigma = 5.67\times10^{-10}$ W/cm^2·K^4 is the Stefan-Boltzmann constant, n, w, t, and L are the number, width, thickness, and length of the bridges, and A and T are the area and temperature of the resonator, respectively. The first term is the contribution due to conduction and the second term is due to radiation. The factor 2A is included in the second term because the resonator radiates from both the front and back surfaces.

The common practice in video application is to make the time constant be one-third the frame time, *i.e.*, to make the thermal time constant ~10 msec for a video frame rate of 30 Hz [70]. The bridge length required to force the thermal time constant of the detector to be τ_T is, using Eqs. (A.2) and (A.3) in Eq. (A.1) and rearranging:

$$L(\tau_T) = \frac{k_q nwt}{\frac{c_v V}{\tau_T} - 8\sigma T^3 A}. \tag{A.4}$$

The calculated value of L for a 10 msec time constant is plotted in Figure A.1. For purposes of this calculation, it is assumed that the shape of the resonator is square, the frequency-thickness constant is 1.6 MHz·mm, the diameter to thickness ratio is 80, there are 4 identical bridges with square cross-section (*i.e.*, w = t = the thickness of the resonator), and T = 300 K.

In these temperature sensitive microresonators, since the mass is extremely small and the temperature coefficient is large, background thermal fluctuation noise is significant. The mean square value of temperature fluctuations, $\overline{\Delta T^2}$, of any system is given by [5]

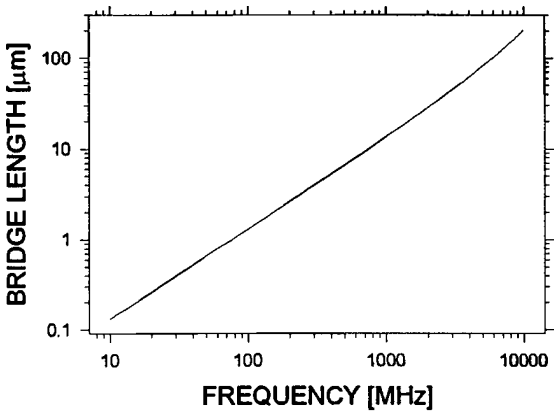


Figure A.1. Bridge length for response time of 10 millisecond as a function of resonator frequency.

$$\overline{\Delta T^2} = \frac{\kappa_B T^2}{C} \qquad (A.5)$$

where $\kappa_B = 1.38\times10^{-23}$ W·sec/K is Boltzmann's constant.

The differential equation describing the heat transfer of the detector is given by [5]

$$C\frac{d(\Delta T)}{dt} + G\Delta T = P(t) \qquad (A.6)$$

where P(t) represents the power fluctuations from a source. We assume that the spectrum of P(t) is independent of frequency (*i.e.*, that the noise is white noise). The steady state solution for the modulus of Eq. (A.6) is

$$\overline{\Delta T_f}^2 = \frac{S_P \Delta f}{G^2+4\pi^2 f^2 C^2}. \qquad (A.7)$$

where S_P is a constant. The mean square value of the fluctuations over all frequencies is obtained by integrating Eq. (A.7):

$$\overline{\Delta T^2} = \int_0^\infty \frac{S_P df}{G^2+4\pi^2 f^2 C^2} = \frac{S_P}{4GC}. \qquad (A.8)$$

Thus, equating Eqs. (A.5) and (A.8) gives the value of S_P to be $4\kappa_B G T^2$.

Substituting the value of S_P into Eq. (A.7) leads to the spectral density of the noise due to temperature fluctuations:

$$S_{\Delta T}(f) = \frac{4\kappa_B T^2/G}{1+4\pi^2 f^2 C^2/G^2}. \qquad (A.10)$$

This can be regarded as a white noise source with spectral density of $4\kappa_B T^2/G$ filtered by a single-pole low pass filter with cut-off frequency $f_{c0} = 1/(2\pi\tau_T)$, as follows

$$S_{\Delta T}(f) = \frac{4\kappa_B T^2/G}{1+f^2/f_{c0}^2}. \qquad (A.11)$$

The power spectral density of thermal fluctuation induced frequency noise is the power spectral density of temperature fluctuations multiplied by the square of the temperature coefficient, α_T. i.e., $S_y(f) = S_{\Delta T}(f)\,\alpha_T^2$.

We neglect the noise contribution of the radiative exchange between the detector and its surroundings because, in the structure under consideration, the conduction is the dominant thermal transport mechanism.

In addition to the temperature fluctuation noise, two noise processes are typically observed: [35]

1. Flicker of frequency where the power spectral density of frequency fluctuations follows a 1/f characteristic, i.e., $S_y(f) = h_{-1}/f$, where h_{-1} is a constant related to the Allan variance of the noise floor, σ_y^2, and σ_y is assumed to be $1.2\times10^{-19}\ \mathrm{Hz}^{-1}\cdot f_0$ where f_0 is the resonator frequency in Hz [35].

$$h_{-1} = \frac{\sigma_y^2}{2\ln(2)} = \frac{(1.2\times10^{-19}\,\mathrm{Hz}^{-1}\cdot f_0)^2}{2\ln(2)} \qquad (A.12)$$

2. White Phase modulation where the power spectral density of frequency fluctuations follows a f^2 characteristic, i.e., $S_y(f) = h_2 f^2$, where h_2 is a constant taken to be the value necessary to make the "knee" of the Allan deviation curve occur at 1 msec, τ_{knee}, for a 500 MHz device with a 40 MHz cut-off filter, f_c.

$$h_2 = \frac{(1.2\times10^{-19}\,\mathrm{Hz}^{-1}\cdot f_0)^2 (8\pi)(\tau_{knee})^2}{3 f_c} \qquad (A.13)$$

There is also white frequency noise (Johnson noise) which is independent of offset frequency i.e., $S_y(f) = h_0$. White frequency noise has a power spectral density which depends on the loaded Q, Q_L, and the power dissipated in the resonator, P.

$$h_0 = \frac{\kappa_B T}{Q_L^2 P} \qquad (A.14)$$

This is usually not observed because its level is usually below the other two noise processes. The relatively low power dissipation and the necessity to

make rapid measurements, may make white frequency noise a significant factor for microresonator arrays.

Figure A.2, shows the four contributions for a 500 MHz resonator using the noise floor and "knee" assumptions stated earlier. The sum of these four noise processes is shown in Figure A.3.

The specific detectivity, D*, being a function of the modulation frequency, is analogous to the familiar phase noise spectrum. The frequency dependence of D* can be expressed as

$$D^*(f) = \frac{R(f)A^{\frac{1}{2}}}{S_y(f)} \qquad (A.15)$$

where $S_y(f)$ is the power spectral density of frequency fluctuations in the detector. The responsivity, R, is defined by [5]

$$R(f) = \frac{\eta\alpha_T}{G(1+(2\pi f)^2\tau_T^2)^{1/2}} \qquad (A.16)$$

where η is the optical absorptance multiplied by the fill factor, and $\alpha_T = \frac{1}{f_0}\frac{df}{dT}$ is the temperature coefficient of the resonator frequency. The units of R are watt^{-1}.

Figure A.4 is D*(f) for three resonator frequencies all with τ_T = 100 msec. Figure A.5 is the thermal conductance required to obtain 10 msec and 100 msec time constants as a function of resonator frequency (with a constant diameter to thickness ratio of 80). The temperature rise due to drive level caused self heating can be readily determined from Figure A.5 and Eq. (1).

Figure A.6 is D* as a function of G (and τ_T) for 600 MHz at the frequency, where D* has a maximum

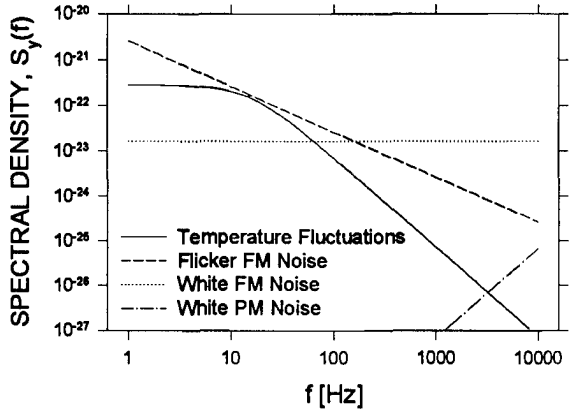


Figure A.2. Power spectral density, $S_y(f)$, vs offset frequency, f, of white phase, white frequency, flicker of frequency, and thermal noise processes for a 500 MHz resonator.

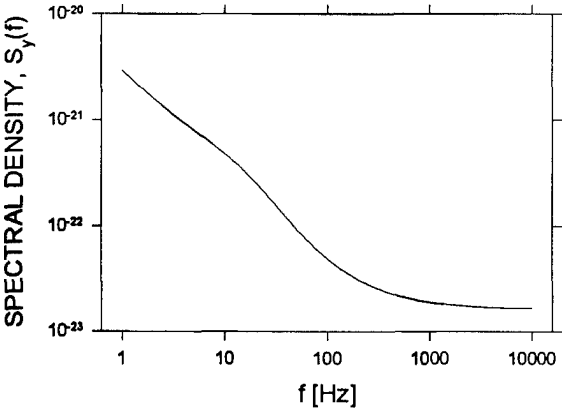


Figure A.3. Sum of white phase, white frequency, flicker of frequency, and thermal noise processes.

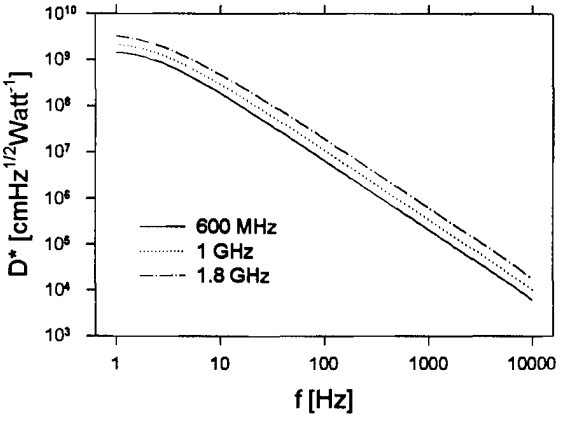


Figure A.4. Detectivity D* as a function of modulation frequency f and resonator frequency with τ = 100 msec.

value, $f=1/(2\pi\tau_T)$ [5]. For comparison, the values of D* of other uncooled thermal detectors are quoted: 1-3×10^8 at 15 Hz for a thermistor bolometer, 1.1×10^8 at 100 Hz for a ferroelectric bolometer [6], 10^9 at 10 Hz for a pyroelectric detector, $\sim3\times10^9$ at 2 Hz for a spectroscopic thermopile, and $\sim6\times10^9$ at 2 Hz for an alanine doped triglycine sulphate (TGS) detector [71].

Although the noise spectrum is a useful analytical tool, when we make real measurements we actually integrate the spectrum over some band of frequencies. A measure of the dispersion, i.e., noise, of the measurements is the variance (or its square root, the deviation).

In general, the variance of a noise process (or the sum of processes) is given by [72]

$$\left\langle \sigma_y^2(N,r,\tau)\right\rangle = \frac{N}{N-1}\int_0^\infty S_y(f)\frac{\sin^2(\pi f\tau)}{(\pi f\tau)^2}\left[1-\frac{\sin^2(\pi rfN\tau)}{N^2\sin^2(\pi rf\tau)}\right]df \qquad (A.17)$$

where N is the number of samples, τ is the measurement time and r is the ratio $(\tau+\tau_d)/\tau$, where τ_d is the dead time between measurements. If there is no dead time r = 1.

If $N = \infty$ and $\tau = 0$, we have

$$\left\langle \sigma_y^2(\infty,r,0)\right\rangle = \int_0^\infty S_y(f)df \qquad (A.18)$$

independent of r, which is the standard definition of the variance of a continuous variable.

If $N = \infty$ and τ is finite, we have the standard variance of the averaged frequency fluctuations

$$\left\langle \sigma_y^2(\infty,r,\tau)\right\rangle = \int_0^\infty S_y(f)\frac{\sin^2(\pi f\tau)}{(\pi f\tau)^2}df \qquad (A.19)$$

which is the total power output from a filter with $S_y(f)$ as input and transfer function $H_1(f,\tau)$ where

$$H_1(f,\tau) = \frac{\sin^2(\pi f\tau)}{(\pi f\tau)^2}. \qquad (A.20)$$

On a linear plot, the response of this filter is shown in Figure A.7. The total area under the curve is $1/(2\tau)$; therefore, a measurement of τ seconds looks like a bandwidth of $1/(2\tau)$ (in H).

A problem arises in noise processes like flicker where the noise power increases as the frequency decreases like $1/f^n$. In that case the integral depends on N and may not converge.

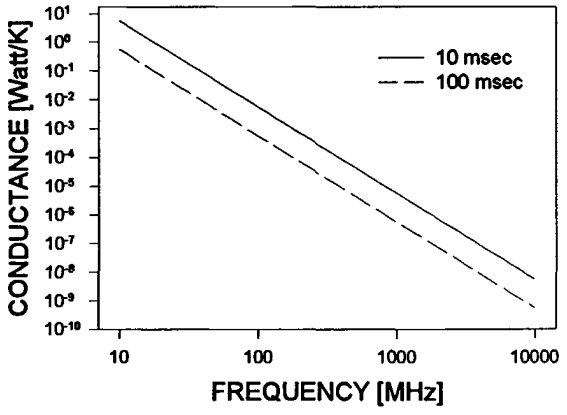


Figure A.5. Thermal conductance for response time of 10 msec and 100 msec as a function of resonator frequency.

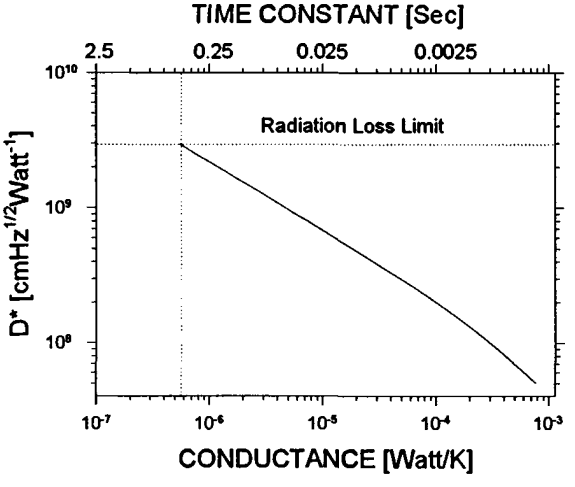


Figure A.6. D* vs G (and τ_T) at $f=1/(2\pi\tau_T)$ for 600 MHz.

If N and τ are finite and there is no dead time the filter looks like a bandpass filter whose lower band edge varies as $1/(N\tau)$.

If N = 2 and there is no dead time, we have the definition of the Allan variance which looks like a filter with transfer function $H_2(f,\tau)$

$$H_2(f,\tau) = 2\frac{\sin^2(\pi f\tau)}{(\pi f\tau)^2}\left[1-\frac{\sin^2(\pi f 2\tau)}{2^2\sin^2(\pi f\tau)}\right] = 2\frac{\sin^4(\pi f\tau)}{(\pi f\tau)^2}. \quad (A.21)$$

This filter looks like Figure A.8. Both the bandwidth and center frequency of the bandpass filter move with averaging time. The decrease of the bandwidth as the frequency decreases is by exactly the right amount to cancel the 1/f dependence of the flicker noise and is why flicker of frequency is the only noise process whose Allan variance is independent of averaging time. This integral also converges for random walk ($1/f^2$) frequency noise.

Physically, the Allan variance is a measure of the fluctuations from measurement to measurement when the measurement intervals are contiguous.

The filter function of the 2-sample variance when the measurements are not contiguous depends on T, the frame interval, and τ, the measurement interval. The function is

$$H_3(f,T,\tau) = 2\frac{\sin^2(\pi f\tau)}{(\pi f\tau)^2}\left[1-\frac{\sin^2(\pi fNT)}{N^2\sin^2(\pi fT)}\right]. \quad (A.22)$$

Figure A. 9 shows the filter functions for a T/τ ratio of 1 and a ratio of 10 for τ = 30 μsec. The effect of the dead time is to increase the contribution of low frequencies.

The 2-sample variance of the spectrum of thermal fluctuations is given by [72]

$$\sigma_{\Delta T}^2(T,\tau) = \int_0^\infty (4\kappa_B T^2/G)\frac{1}{1+(f/f_{c0})^2}|H_3(f,T,\tau)|^2 df. \quad (A.23)$$

The 2-sample deviation with no dead time for the noise discussed above is shown in figure A.10 for the resonator frequency of 500 MHz. The effect of increasing the dead time is shown in figure A.11. A framerate of 30 Hz is equivalent to a repetition ratio of 1100 for τ = 30 μsec. Both thermal and flicker noise increase significantly when dead time is introduced.

The total frequency noise, σ_N, is given by a root mean square of the temperature fluctuations times the temperature coefficient and the oscillator noise:

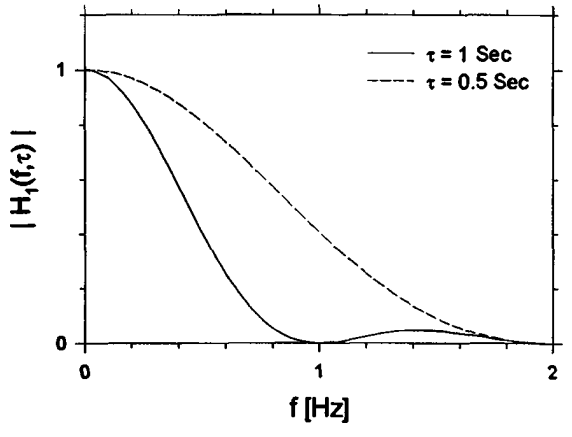


Figure A.7. Filter shape for the ∞-sample variance for τ = 1 sec and τ = 0.5 sec.

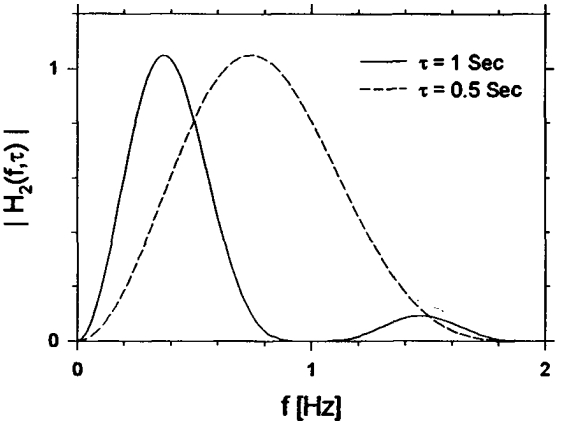


Figure A.8. Filter shape for the two-sample variance for τ = 1 sec and τ = 0.5 sec.

$$\sigma_N = \sqrt{(\sigma_{\Delta T}\alpha_T)^2 + (\sigma_y)^2}\,. \qquad \text{(A.24)}$$

The noise equivalent temperature difference, NETD, is defined as [73]

$$\text{NETD} = \frac{(4F^2+1)\sigma_N}{\tau_0 A(\Delta P/\Delta T)_{\lambda_1-\lambda_2} R} \qquad \text{(A.25)}$$

where F is the numerical aperture of the optics, τ_0 is the transmittance of the atmosphere between the target and the sensor, and $(\Delta P / \Delta T)_{\lambda_1-\lambda_2}$ is the temperature dependence of the blackbody function over the wavelength interval from λ_1 to λ_2, respectively. For thermal imaging systems operating in 8-14 μm atmospheric window, the value of $(\Delta P / \Delta T)_{\lambda_1-\lambda_2}$ is 2.62 W/m^2·K [70].

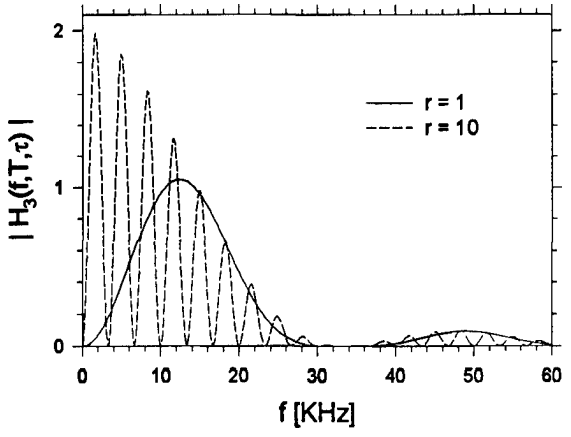


Figure A.9. Filter shape for the two-sample variance for r=1 and r=10.

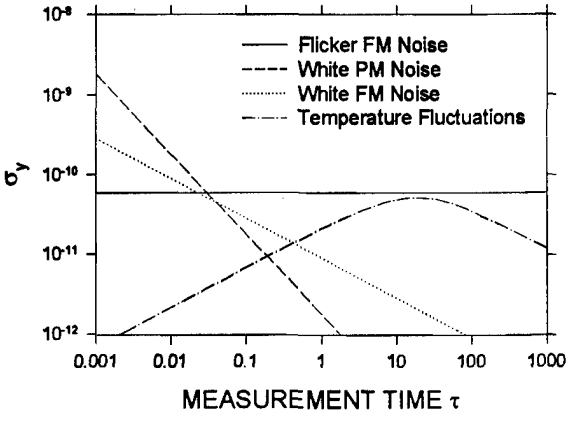


Figure A.10. Two-sample variance of noise processes as a function of measurement time at r = 1 for the resonator frequency of 500 MHz.

Appendix B - Chemical and Biological Sensor Calculations

The total thermal energy released by adsorption, desorption or chemical reaction occurring on the surface of a resonator with area A can be expressed as nAe_R where n is the number of reacting species per unit area, e_R is the energy absorbed or released per molecule interacting with the surface. The value of e_R is equal to E_R/N_0 where E_R is the activation energy per mole and N_0 is Avogadro's number. If the fraction of available sites is μ and σ_A is the atomic surface density then n is μ·A and the available energy is $\mu\sigma_A Ae_R$. For simplicity, it is assumed that the energy is completely absorbed in the resonator, the measurement time is short compared to the thermal response time, and there is no change in mass loading. Thus, this temperature increase is

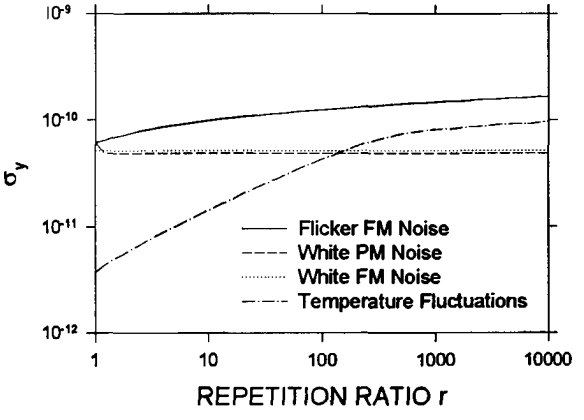


Figure A.11. Two-sample variance of noise processes as a function of repetition ratio r.

$$\Delta T = \frac{\mu \sigma_A A e_R}{c_V V}. \qquad (B.1)$$

The limit of detectability is defined as the fraction of a monolayer which causes a temperature change which induces a frequency change equal to the noise:

$$\Delta T_{noise} = \sigma_N \frac{1}{f_0}\frac{df}{dT} = \frac{\mu_{limit}\sigma_A e_R}{c_V t} \qquad (B.2)$$

where t is the thickness of the resonator. The typical values of σ_A and E_R are assumed here to be 10^{13} cm^{-2} and 25 Kcal/mol, respectively.

The delectability limit for mass loading, $\Delta m/m$, is determined by the condition that the frequency change due to the mass loading is equal to the noise, *i.e.*,

$$\sigma_N = \frac{df_{limit}}{f_0} = \frac{\Delta m_{limit}}{m} = \frac{\mu_{limit} A \sigma_{quartz}}{V \rho_{quartz}} \qquad (B.3)$$

where $\sigma_{quartz} = 9.9668\times10^{-10}$ g/cm^2 and $\rho_{quartz} = 2.65$ g/cm^3 are the surface density and the mass density of quartz, respectively.

References

[1] J. R. Vig, et al., "Chemically Polished Quartz," Proc. 31st Ann. Symp. on Frequency Control, pp. 131-143, 1977.

[2] J. R. Hunt and R. C. Smythe, "Chemically Milled VHF and UHF AT-cut Resonators," Proc. 39th Ann. Frequency Control Symposium, pp. 292-300, 1985.

[3] R. C. Smythe and R. B. Angove, "Chemically-Milled UHF SC-cut Resonators," Proc. 42nd Ann. Frequency Control Symposium, pp. 73-77, 1988.

[4] W. P. Hanson, "Chemically Polished High Frequency Resonators," Proc. 37th Ann. Frequency Control Symposium, pp. 261-271, 1983.

[5] P. W. Kruse, L. D. McGlauchlin and R. B. McQuistan, *Elements of Infrared Technology*, John Wiley and Sons, Inc., New York (1962).

[6] R. D. Hudson, *Infrared System Engineering*, John Wiley and Sons, Inc., New York (1969).

[7] P. W. Kruse, "A Comparison of the Limits to the Performance of Thermal and Photon Detector Imaging Arrays," Infrared Physics and Technology, 1995, to be published.

[8] H. C. Wright, *Infrared Techniques*, Clarendon Press, Oxford, 1973.

[9] R. J. Keyes, editor, *Optical and Infrared Detectors*, Vol. 19 of Topics in Applied Physics, especially Chapter 3, "Thermal Detectors," by E. H. Putley, Springer Verlag, 1980.

[10] R. A. Heising, *Quartz Crystals for Electrical Circuits*, pp. 26-27, 32, D. Van Nostrand Co., 1946.

[11] W. H. Wade and L. J. Slutsky, "Quartz Crystal Thermometer," Rev. Sci. Instr., vol. 33, pp. 212-213, 1962.

[12] I. Gorini and S. Sartori, "Quartz Thermometer," Rev. Sci. Instr., vol. 33, pp. 883-884, 1962.

[13] W. L. Smith and W. J. Spencer, "Quartz Crystal Thermometer for Measuring Temperature Deviations in the 10^{-3} to 10^{-6} °C Range," Rev. Sci. Instr. vol. 34, pp. 268-70, 1963.

[14] D. L. Hammond, C. A. Adams and P. Schmidt, "A Linear, Quartz Crystal, Temperature Sensing Element," ISA Transactions, vol. 4, pp. 349-354, 1965.

[15] H. Ziegler and J. Tiesmeyer, "Digital Sensor for IR Radiation," Sensors and Actuators, vol. 4, pp. 363-367, 1983.

[16] H. Ziegler, "A Low-Cost Digital Temperature Sensor System," Sensors and Actuators, vol. 5, pp. 169-178, 1983.

[17] J-J Gagnepain, et al., "Temperature Probe Using a Plate of Quartz," U.S. Pat. No. 4,398,115, Aug. 9, 1983.

[18] J. E. Ralph, R. C. King, J. E. Curran and J. S. Page, "Miniature Quartz Resonator Thermal Detector," Proc. 1985 Ultrasonics Symp., pp. 362-364, 1985.

[19] E. P. EerNisse, R. W. Ward, and R. B. Wiggins, "Survey of Quartz Bulk Resonator Sensor Technologies," IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control, vol. 35, pp. 323-330, 1988.

[20] L. Spassov, "Piezoelectric Quartz Resonators as Highly Sensitive Temperature Sensors," Sensors and Actuators A, vol. 30, pp. 67-72, 1992.

[21] M. R. Hamrour and S. Galliou, "Analysis of the Infrared Sensitivity of a Quartz Resonator - Application as a Thermal Sensor," Proc. 1994 Ultrasonics Symp., pp. 513-516, 1994.

[22] J. A. Kusters, M. C. Fisher, and J. G. Leach, "Dual Mode Operation of Temperature and Stress Compensated Crystals," Proc. 32nd Ann. Symp. on Frequency Control, pp. 389-397, 1978.

[23] S. S. Schodowski, "Resonator Self-Temperature Sensing Using a Dual-Harmonic-Mode Crystal Oscillator," Proc. 43rd Ann. Symp. on Frequency Control, pp. 2-7, 1989.

[24] R. L. Filler and J. R. Vig, "Resonators for the Microcomputer Compensated Crystal Oscillator," Proc. 43rd Ann. Symp. on Frequency Control, pp. 8-15, 1989.

[25] M. Nakazawa, H. Yamaguchi, A.D. Ballato, and T.J. Lukaszek, "Stress-Compensated Quartz Resonators Having Ultralinear Frequency-Temperature Responses," Proc. 38th Ann. Frequency Control Symp., pp. 240-244, 1984.
[26] B. Sinha, "Stress Compensated Orientation for Thickness Shear Quartz Resonators," Proc. 35th Ann. Frequency Control Symp., pp. 213-221, 1981.
[27] T. E. Parker, "1/f Frequency Fluctuations in Quartz Acoustic Resonators," Appl. Phys. Lett., vol. 46, pp. 246-248, 1985.
[28] T. E. Parker, "Characteristics and Sources of Noise in Stable Oscillators," Proc. 41st Ann. Symp. on Frequency Control, pp. 99-110, 1987.
[29] J. R. Vig and F. L. Walls, "Fundamental Limits on the Frequency Instabilities of Quartz Crystal Resonators," Proc. 1994 IEEE Int'l Frequency Control Symp., pp. 506-523, 1994.
[30] G. K. Montress and T. E. Parker, private communication, 1995; and "Design and Performance of an Extremely Low Noise Surface Acoustic Wave Oscillator," Proc. 1994 IEEE Frequency Control Symp., pp. 365-373, 1994.
[31] M. M. Driscoll and W. P. Hanson, "Measured vs. Volume Model-Predicted Flicker-of-Frequency Instability in VHF Quartz Crystal Resonators," Proc. 1993 IEEE Int'l Frequency Control Symp., pp. 186-192, 1993.
[32] F. Walls, "A New Model of 1/f Noise in BAW Quartz Resonators," Proc. 1992 IEEE Int'l Frequency Control Symp., pp. 327-333, 1992.
[33] T. E. Parker and D. Andres, "Dependence of SAW Resonator 1/f Noise on Device Size," Proc. 1993 IEEE Int'l Frequency Control Symp., pp. 178-185, 1993.
[34] Y. K. Yong and J. R. Vig, "Resonator Surface Contamination - A Cause for Frequency Fluctuations?," Proc. 42nd Ann. Frequency Control Symp., pp. 397-403, 1988.
[35] T. E. Parker, NIST, *private communication*, 1994.
[36] F. L. Walls, NIST, *private communication*, 1994.
[37] S. Kumar, D. Cho and W. N. Carr, "Experimental Study of Electric Suspension for Microbearings," J. of Microelectromechanical Systems, vol. 1, pp. 23-30, 1992.
[38] S. Kumar and D. Cho, "Electrostatically Levitated Microactuators," J. Micromech. Microeng. vol. 2, pp. 96-103, 1992.
[39] U.S. Patent No. 5,015,906, "Electrostatic Levitation Control System for Micromechanical Devices," D. Cho, S. Kumar and W. Carr.
[40] R. E. Pelrine, "Room Temperature, Open-Loop Levitation of Microdevices Using Diamagnetic Materials," Proc. IEEE MicroElectroMechanical Systems, IEEE Cat. No. 90CH2832-4, pp. 35-37, 1990.
[41] K. C. Lee, "The Fabrication of Thin, Freestanding, Single-Crystal, Semiconductor Membranes," J. Electrochem. Soc., vol. 137, pp. 2556-2574, 1990.
[42] W. G. Spitzer and D. A. Kleiman, "Infrared Lattice Bands of Quartz," Physical Review, vol. 121, pp.1324-1335, 1961.
[43] E. D. Palik, *Handbook of Optical Constants of Solids*, vol. 1. Orlando, Florida: Academic Press, 1985.
[44] L. N. Hadley and D. M. Dennison, "Reflection and Transmission Interference Filters," J. Opt. Soc. Am., vol. 37, pp. 451-465, 1947.
[45] W. Lang, K. Kühl and H. Sandmaier, "Absorbing Layers for Thermal Infrared Detectors," Transducers '91. Int'l Conf. on Solid-State Sensors and Actuators, Digest of Technical Papers, IEEE Cat. No. 91CH2817-5, pp. 635-638, 1991, also in Sensors and Actuators A, vol. 34, pp. 243-248, 1992.
[46] K. C. Liddiard, "Application of Interferometric Enhancement to Self-absorbing Thin Film Thermal IR Detectors," Infrared Phys., vol. 34, pp. 379-387, 1993.
[47] A. D. Parsons and D. J. Pedder, "Thin-film Infrared Absorber Structures for Advanced Thermal Detectors," J. Vac. Sci. & Technol., vol. A 6 (3), pp. 1686-1689, 1988.
[48] O. S. Havens, *Optical Properties of Thin Solid Films*, Butterworth Sci. Publ., chapters 4 and 6, 1955.
[49] D. J. Advena, V. T. Bly & T. J. Cox, " Deposition and Characterization of Far-infrared Absorbing Gold Black Films," Appl. Optics, vol. 32, pp. 1136-1144, 1993.
[50] D. E. McCarthy, "The Reflection and Transmission of Infrared Materials: I, Spectra from 2 to 50 Microns," Appl. Optics, Vol. 2, pp. 591-595, 1963.
[51] H. F. Tiersten, *private communication*, 1995.
[52] R. C. Smythe, *private communication*, 1995.
[53] R. B. Belser and W. H. Hicklin, "AT-cut Resonators with Annular Electrodes," Proc. 21st Ann. Symp. on Frequency Control, pp. 211-223, 1967.
[54] K. Nakamura, et al., "Trapped-Energy Resonators with Elliptical Ring Electrodes," Proc. 44th Ann. Symp. on Frequency Control, pp. 372-377, 1990.
[55] A.D. Ballato, et al., "Simple Thickness Plate Modes Driven by Lateral Fields," Proc. 39th Ann. Symp. on Frequency Control, pp. 462-472, 1985
[56] A.W. Warner, Jr., and B. Goldfrank, "Lateral Field Resonators," Proc. 39th Ann. Symp. on Frequency Control, pp. 473-474, 1985
[57] R. C. Smythe and H. F. Tiersten, "An Approximate Expression for the Motional Capacitance of a Lateral Field Resonator," Proc. 41st Ann. Symp. on Frequency Control, pp. 311-313, 1987.
[58] D. L. Hammond and L. S. Cutler, "Crystal Resonators," U. S. Pat. No. 3,339,091, Aug. 29, 1967, and Reissue 26,707, Nov. 4, 1969.

[59] R. J. Besson, "A New 'Electrodeless' Resonator Design," Proc. 31st Ann. Symp. on Frequency Control, pp. 147-152, 1977.
[60] R. A. Wood, "Uncooled Thermal Imaging with Monolithic Silicon Focal Plane Arrays," Proc. SPIE 2020, Infrared Technology XIX, pp. 322-329 (1993).
[61] C. M. Hanson, "Uncooled Ferroelectric Thermal Imaging," Proc. SPIE 2020, Infrared Technology XIX, pp. 330-339 (1993).
[62] R. A. Wood and N. A. Foss, "Micromachined Bolometers Achieve Low-cost Imaging," *Laser Focus World*, pp. 101-106, June 1993.
[63] R. Demeis, "Select the appropriate camera for your application," *Laser Focus World*, pp. 105-112, July 1995.
[64] Applications of Piezoelectric Quartz Crystal Microbalances, C. Lu and A. W. Czanderna, editors, Elsevier Science Publishers, 1984.
[65] M. D. Ward and D. A. Buttry, "In Situ Interfacial Mass Detection with Piezoelectric Transducers," Science, vol. 249, pp. 1000-1007, 31 Aug 1990.
[66] J. K. Gimzewski, et al., "Observation of a Chemical Reaction Using a Micromechanical Sensor," Chemical Physics Letters, vol. 217, pp. 589-594, 1994.
[67] Danielsson, "Calorimetric Biosensors," J. of Biotechnology, vol. 15, pp. 187-200, 1990.
[68] Private communication with Thomas Payne, Avance Technology Inc., and Juergen Staudte, XECO, Inc., Cedar City UT, June 1995.
[69] H. F. Tiersten, "The Influence of an Asymmetric Air Gap on the Vibration Characteristics of Quartz Trapped Energy Resonators," elsewhere in this Proceedings volume.
[70] P. W. Kruse, *Private communication*. 1994.
[71] E. H. Putley, "Thermal Detectors," Chap. 3 of *Optical and Infrared Detectors*, edited by R. J. Keyes, Springer Verlag, New York, NY, 1980.
[72] P. Lesage and C. Audoin, "Characterization and measurement of time and frequency stability," Radio Science,Vol. 14, pp.521-539, 1979.
[73] B. R. Johnson and P. W. Kruse, "Silicon microstructure superconducting microbolometer infrared arrays," *SPIE* Vol. 2020 Infrared Technology XIX, pp.2-11, 1993.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DIRECT SIMULATION OF NEXT GENERATION DEVICES BY ATOMISTIC SIMULATION

Christopher Meli and Jeremy Broughton,
Naval Research Laboratory,
Washington, DC 20375
and
Priya Vashishta and Rajiv Kalia,
Physics Department,
Louisiana State University,
Baton Rouge, LA 70803

I. Introduction

Present devices which rely upon mechanical principles can be built in the microscale regime using technology developed for electronic device fabrication. Mechanical devices have been made from silicon and transition metals - using, for example, lithographic methods. We have all seen those images which do so well at capturing our imagination on the cover of Science, and related journals, depicting simple gears of width less than a human hair. Such devices are entering the size regime where continuum mechanics may not be the only way of describing device performance.

A case in point is afforded by quartz crystal oscillators (QCO's). As we shall see, these devices rely upon very simple mechanical principles for their operation. Present QCO's can be manufactured in the 100 micron length scale regime; they involve megahertz frequencies. Next generation oscillators will hit the sub-micron regime where they will afford gigahertz frequencies and much higher sensitivities (e.g., force,mass, etc). It is this size regime that may be directly accessed by modern atomistic simulation; that is the multi million atom system size.

Modern parallel processors such as the ones at Louisiana State University or those accessible at NSF, DoE and DoD sites are capable, now, of handling 1-10 million (and more) atom systems for elapsed times of nanoseconds - precisely the time regime of interest. Current computer technology doubles performance every year - enabling easier and easier access to direct device modelling as time progresses.

II. System

We chose to study the system depicted in Figure 1, which shows a flexural plate wave (sometimes called vibrating beam) oscillator. Such systems are typically constructed from piezoelectric crystals, most often alpha-quartz, SiO_2. By applying an electric field in the appropriate direction, these devices can be made to flex normal to, in this case, the (1210) face. (Note, however, that fields applied in other ways can excite surface acoustic waves and shear modes - leading to devices with other characteristics and advantages). Such devices vary from the scale of centimeters down to microns, the sensitivity increasing as the scale is decreased. Since quartz is an anisotropic material, attention is paid to the cut of the exposed faces of the oscillator. The particular one shown in Figure 1 has the property that the resonant frequency of this device has a stationary point around room temperature - that is, the effects of temperature on the operating characteristics of the device vary slowly under ambient conditions; clearly a desirable feature. The crucial equation governing these oscillators - when excited in flexural mode - is given, to lowest order, by elasticity theory as:

$$\omega \propto \frac{d}{L^2}\sqrt{\frac{K}{\rho}} \qquad (1)$$

ω = Frequency of Device
d = Thickness of Device
L = Length of Device
K = Elastic Constant
ρ = Density of Device

0-7803-2500-1/95 $4.00 © 1995 IEEE

The elastic constant here, Young's modulus, is that due to elongation/compression along the longitudinal axis of the device.

This equation illustrates why such devices are useful. If the device is compressed or elongated, the modulus changes and the operating frequency has to do likewise. Thus QCO's make good accelerometers; they might be used in guidance systems for example. They also can be used as gravitometers - state-of-the art devices detect micro-g's. This is very impressive sensitivity, and yet ARPA (for example) would like to build devices with nano-g sensitivity. This sensitivity requires thousand Angstrom devices; i.e., those envisaged for simulation here. This sensitivity, however, comes at a price - that of increased noise. The effects of impingement of gas phase atoms and even photons upon the surface of these devices contribute to their operating properties - the consequences of such may be explored also by direct simulation.

If anything alters the surface of the device - such as mass adsorbed, or surface reconstruction, or surface rebonding - the above equation shows that the frequency must change. Thus QCO's are used as mass detectors, viscometers and densitometers. Assuming that mass deposition does not affect elastic constants in the device, the basic equation controlling frequency change, again from elasticity theory, is:

$$\delta\omega = \frac{1}{2M}\,\omega\,\delta M \qquad (2)$$

$$M = \textit{Mass of Device}$$

Such assumptions can, of course, be explored by direct simulation.

Returning to equation (1), substituting the known modulus for alpha quartz, we arrive at the observed operating frequency for a 100 micron device of several megahertz. Next generation devices, aiming for higher sensitivity, are aiming for the fraction of a micron regime. For example, if the length of the device, as shown in Figure 1 is 0.2 microns, i.e., 2000 Angstroms, then we obtain frequencies in the gigahertz range. This is precisely the length and time scale for which direct simulation, on the most powerful National machines, is possible. The device as depicted in the figure has approximately 2 million atoms. Frequencies in the gigahertz range imply that several cycles of oscillation should be attainable within a nanosecond elapsed-time simulation.

Operationally, it is important to use very high quality crystalline quartz - otherwise internal relaxation phenomena (e.g., motion of defects) cause so-called "internal friction" leading to undesirable device characteristics such as resonant frequencies which change with time under constant external conditions. It is here that simulation may be of significant benefit since the effect of "culprit" defects may be studied directly by initializing a simulation with the defect already included.

III. Method

The methodology here is that of molecular dynamics (MD). Given an interatomic Hamiltonian for the system, Newton's laws of motion may be solved for each interacting atom in the system by splitting time into very small increments (say, 10^{-15} second). Every particle's trajectory is then found in lock-step as the system evolves in time. Properties are then ensemble averages of the system behavior. Temperature, of course, is simply related (these are classical mechanical systems) to the mean kinetic energy of the particles. By increasing the kinetic energy of the atoms, the temperature of the system is made to increase. Pressure may be changed by varying the system size (or length) of the system.

We have adopted a two pronged approach to the simulation process. The first is the brute-force simulation of the device. Non-linearities may be investigated this way - such as the effect of local and/or extended defects and the way their internal motion couples to the device vibrational modes. The second is medium scale (100,000 atom) modelling of the material comprising the device in order to determine such properties as elastic constants, defect concentrations, thermal expansivities and phase transitions. The latter may be run in bulk-3D mode or in bar mode (i.e., 1D) to determine the effects of edge/surface/volume ratios (i.e., device size); and of surface mass deposition and of surface reconstruction on device frequency. Our simulations probe these properties as a function of temperature and applied stress.

These simulations, of course, require a trustworthy interatomic potential. One such is a modification of that due to Vashishta and Kalia which is described in reference (1) where it was used to study atomic correlation functions in molten, crystalline and amorphous states of SiO_2. The modification involves replacing the long-range Coulomb interaction between atoms by a screened (Yukawa) Coulomb form. The potential is of the

Stillinger-Weber[2] type; that is, there is a two-body part that couples each pair of atoms and a three-body part that involves only O-Si-O and Si-O-Si bonds. This potential has the right lattice parameter and elastic constants relative to experiment. It also gets the alpha-beta transition correctly in quartz. It does a reasonable job for the phonons. From the point of view of elucidating general principles, even if the potential has deficiencies in accuracy, much can be learnt about the effect of atomistic defects on device operating behavior - because we expect such behavior to be dominated by gross features of the potential.

Long timescale events - such as defect migration - may also be accessed in the medium scale simulations by smart use of coupling absolute rate theory and statistical mechanics. Such methodology can be used to look at the timescales of suspected "culprit" (i.e., undesirable) defects. Also defect free energies can be accessed by defining inventive reversible paths which may be integrated during the simulation procedure. Such approaches are on our agenda.

IV. Initial Results

We have studied the effects of system size on the Young's modulus of the QCO. Figure 2 shows this at 10 and 300K for a perfect bar and at 300K for a system with 1% vacancies (our present approximation to a system with typical as-grown defects). The aspect ratio is 1:2 for the x:z dimensions. (The y-direction is the longitudinal axis of the rod). The asymptotic limits for infinite system size are obtained from 3D simulations using periodic boundary conditions. Note that the Young's modulus obtained using our interatomic potential is a little smaller than that found experimentally - we are working upon improving our potential. Nevertheless, the main conclusions (i.e., trends) reached here are unlikely to change upon using a new potential. The implications of Figure 2 are that for a device aspect ratio of 1:14:2, the Young's modulus of the bar reaches 99% of the bulk value for a device length of 0.13μ.

We have also directly monitored the center-of-gravity (in the x-direction) versus time of a bar of length 170Å (comprising 20,000 atoms) initially "plucked" by 1Å at its center. Figures 3 and 4 show the behavior of this quantity for 10 and 300K. Notice the very different behavior. Upon Fourier transforming (Figures 5 and 6), the latter system exhibits a frequency at twice the fundamental showing that anharmonic terms in the equation of motion of the bar's flexure are very important at the higher temperature.

Knowing the Young's modulus for this specific sized bar at both 10 and 300K, it is possible to use Equation 1 to predict what continuum mechanics would give for a frequency and compare it with the direct value as given by MD simulation. Continuum mechanics predicts 3.3 and $3.0\text{x}10^{10}$ for 10 and 300K respectively. MD gives 3.05 and $2.47\text{x}10^{10}$ respectively. Thus, for these system sizes, continuum mechanics predicts the correct order of magnitude for these quantities but is in error at the 10-20% level.

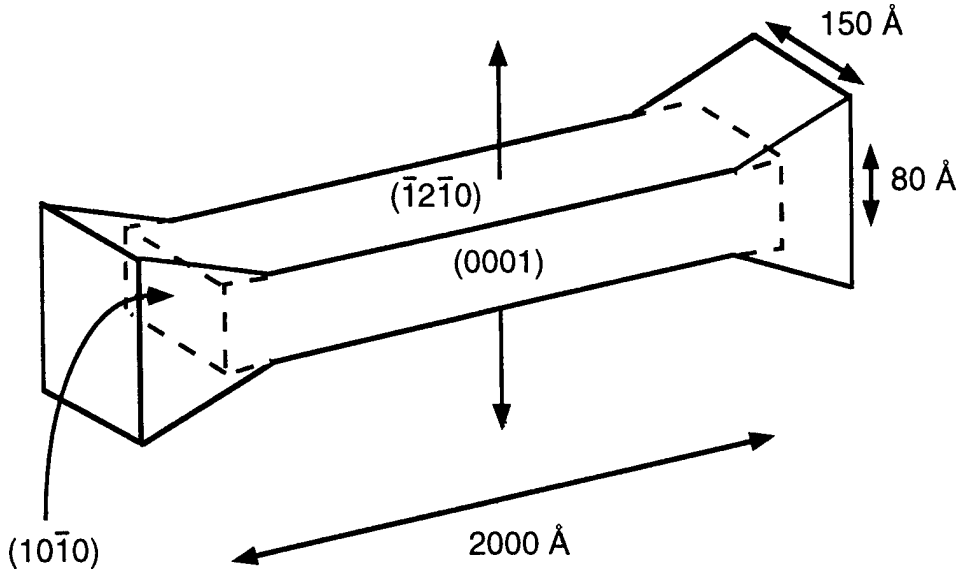
V. Summary

The bottom line is that such simulations can answer many device-related questions - some quantitatively and others qualitatively. Some nagging device flaw features may not be accessible - but only by our lack of imagination at guessing the nature of the "culprit" defects. Once such defects are defined, simulation machinery exists to investigate their ramifications.

The system sizes amenable to direct simulation are in the right regime (and of the right timescale) to be of relevance to next generation devices. It is in the 0.1 micron regime in which deviations from bulk continuum behavior manifest themselves.

REFERENCES:

1) P.Vashishta, R.K.Kalia, J.P.Rino and I.Ebbsjo, Phys.Rev. B41,(1990),12197.
2) F.H.Stillinger and T.A.Weber, Phys.Rev. B31,(1985),5262.



Two Million Particle Device

Figure 1.

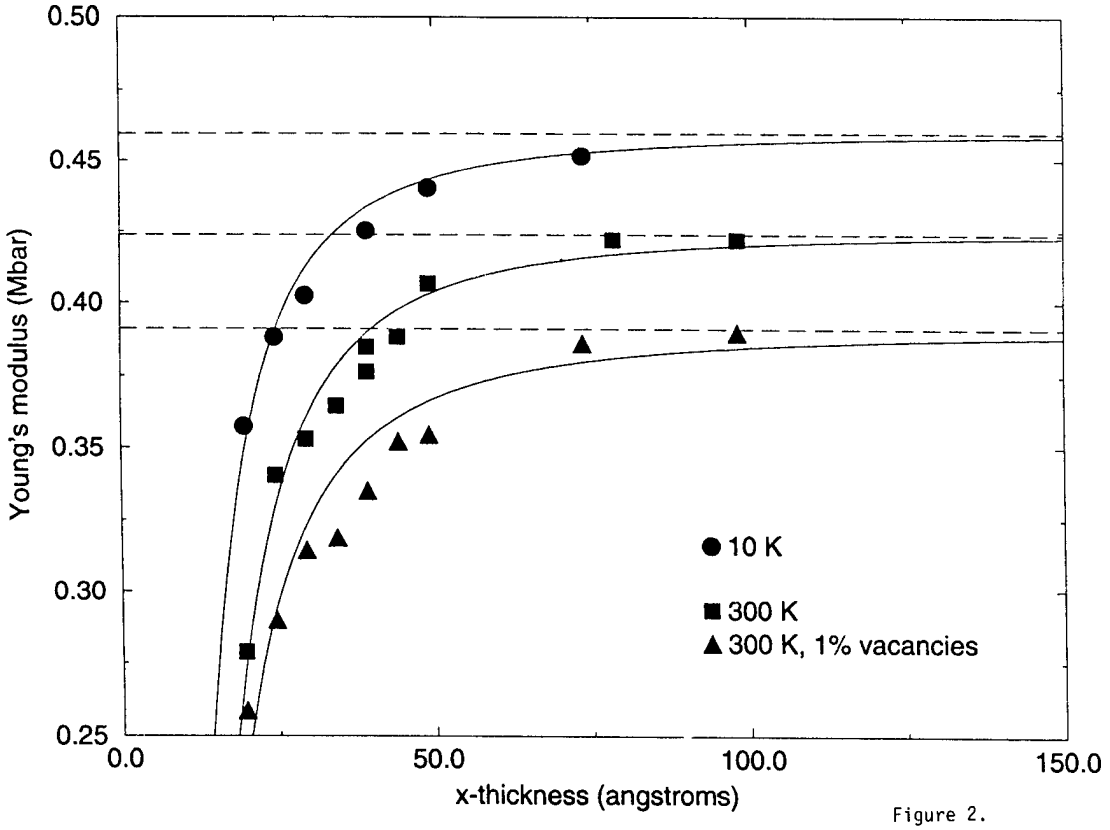


Figure 2.

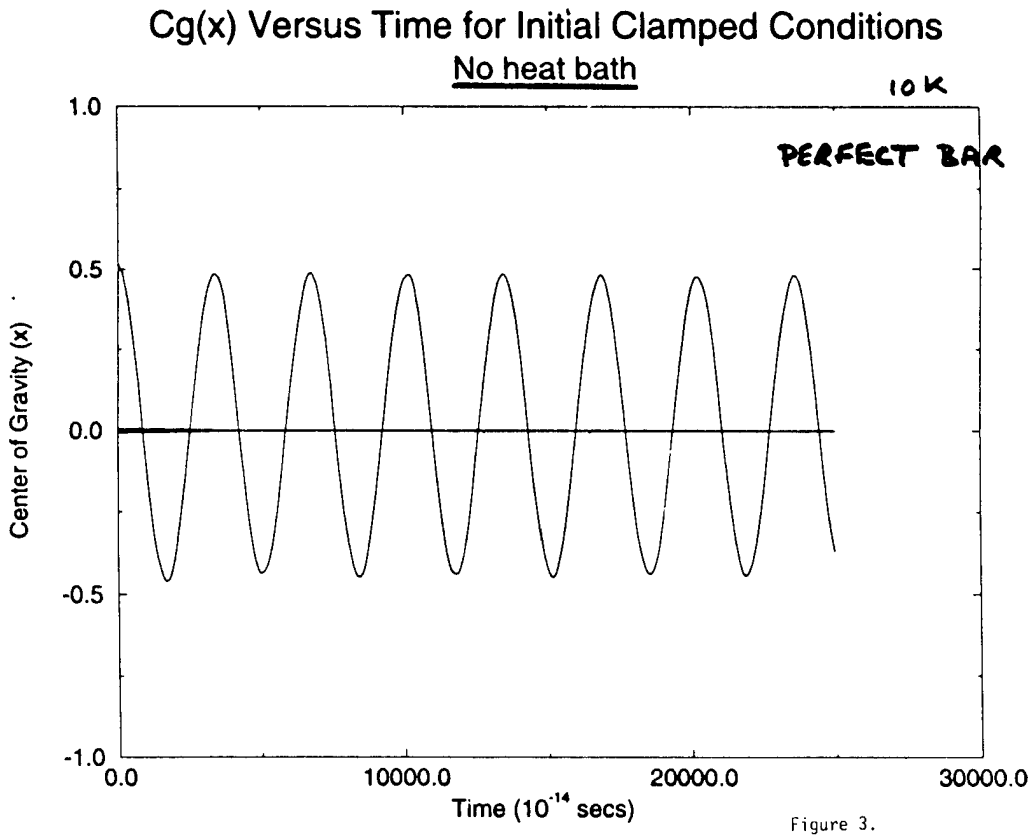


Figure 3.

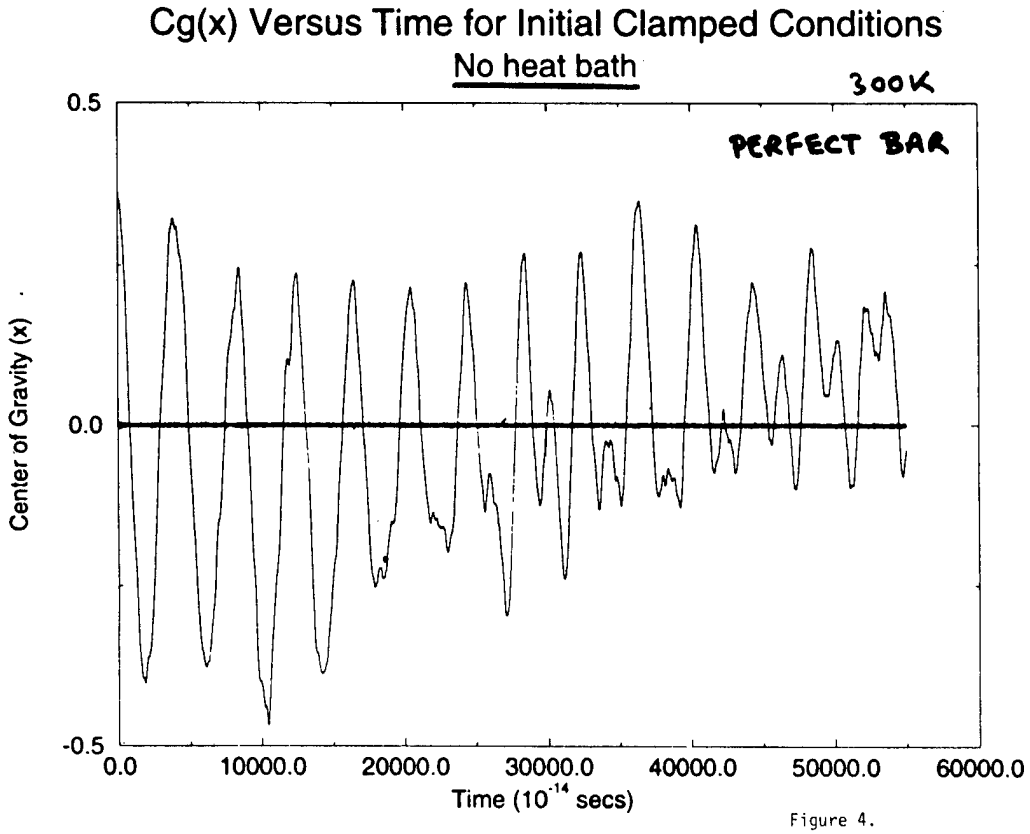


Figure 4.

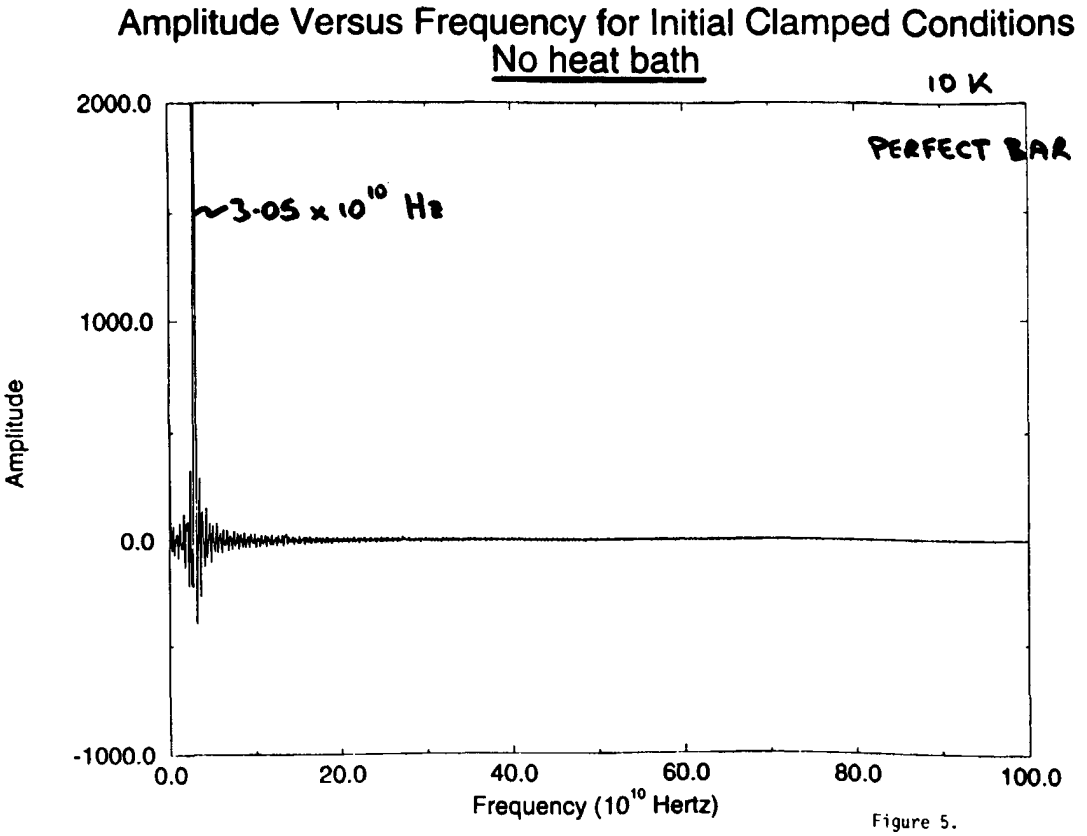


Figure 5.

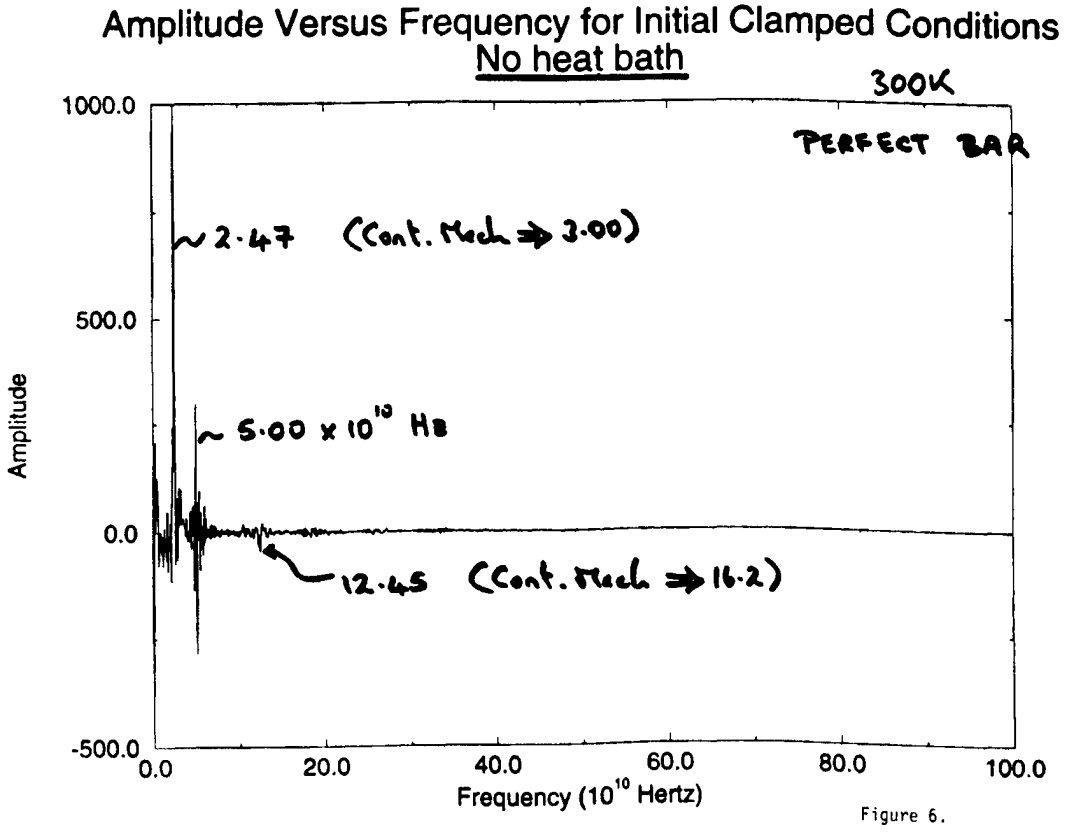


Figure 6.

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A Smart H_2 Sensor Array of QCMs coated with PdNi Alloy of Various thicknesses

G. Liu and M. D. Porter
Ames Laboratory - USDOE,
Microanalytical Instrumentation Center,
And Department of Chemistry
Iowa State University, Ames, IA 50011

G. J. Bastiaans and S. Burns
Microelectronics Research Center,
and Department of Electrical and Computer Engineering
Iowa State University, Ames, IA 50011

Abstract

This paper describes some of our results in developing high performance hydrogen gas sensors using piezoelectric devices (i.e., Quartz Crystal Microbalances (QCMs) and Thin Film Resonators (TFRs)) coated with a PdNi (90/10) alloy. A sensor array has been built by coating quartz crystals with various PdNi thicknesses and tested via expose to H_2 in N_2 in the concentration range from 20 to 500 ppm. The performance of the sensor array has been also characterized in the presence of various potential interference gases and vapors, such as carbon dioxide, nitrous oxide, water, toluene, acetone, acetonitrile, hexane, methyl isobutylketone, methylene chloride, and ethanol. Using the unique chemistry of Pd-H_2 system, we have developed an relatively straightforward method to compensate for contributions from the mentioned potential interference gases

Introduction

The ability to monitor hydrogen gas using small, field deployable sensors is of critical importance to ongoing efforts in nuclear energy and space exploration programs. To this end, we have been developing high performance hydrogen gas sensors using piezoelectric devices (i.e., Quartz Crystal Microbalances (QCMs) and Thin Film Resonators (TFRs)) coated with a PdNi (90/10) alloy. In this paper, we will briefly discuss some of our results aimed at characterizing the performance of these sensors.

Recently, palladium has been widely used as a sensing coating for hydrogen detection[1,2]. The basis of using Pd-based material for this purpose derives from its adsorptive dissociation and dissolution of hydrogen into the bulk of the metal[3] The process is given in reaction (1),

$$H_2 + Pd \underset{k_d}{\overset{k_a}{\rightleftharpoons}} 2H_{ad}(Pd) \overset{K_2}{\rightleftharpoons} 2H_{bulk}(Pd) \qquad (1)$$

where k_a and k_d are the adsorption and desorption rate constants, respectively, and K_2 is the equilibrium constant for the dissolution of hydrogen atoms into the bulk of palladium material. Since the dissolution process is usually fast, the rate of hydrogen uptake into palladium material is determined by the adsoprtion and desorption step, which is described by the equilibrium constant, K_1

$$K_1 = k_a / k_d = [H_{ad}(Pd)]^2 / p(H_2) \qquad (2)$$

where $[H_{ad}(Pd)]$ represents the amount of hydrogen atoms adsorbed on the surface, and $p(H_2)$ is the partial pressure of hydrogen gas in the gas phase. Rearranging equation (2) and using the relation, $K_2 = H_{bulk}(Pd)] / [H_{ad}(Pd)$ yields

$$[H_{bulk}(Pd)] = K' [p(H_2)]^{1/2} \qquad (3)$$

where $K' = K_1^{1/2} K_2$. Equation (3) indicates that the uptake of H_2 gas by the Pd coating is directly proportional to the square root of hydrogen gas partial pressure. Importantly, the absorption of H_2 in Pd is a bulk phenomenon. The uptake of the target gas depends on the PdNi coating thickness. This dependence will be used to construct an interference correction scheme for the hydrogen sensor based on an array concept.

0-7803-2500-1/95 $4.00 © 1995 IEEE

Results and Discussion

I. Responses of the Sensors to Hydrogen: Figure 1 presents the observed frequency responses, given as the change in the resonance frequency, Δf_s , as a function of hydrogen partial pressure and PdNi coating thickness. The error bars are determined from duplicate measurements of three to four PdNi coatings of the same thickness. The straight lines are obtained by a least squares analysis. As expected, the response at each coating thickness increases linearly with $[p(H_2)]^{1/2}$. More importantly, the response at each H_2 concentration increases as the thickness of the coating increases. This dependence confirms that uptake of H_2 into PdNi coatings is a bulk phenomenon. These data also show that the detection limit is ~3ppm, based on a relative noise level in the frequency measurement of 0.1 ppm.

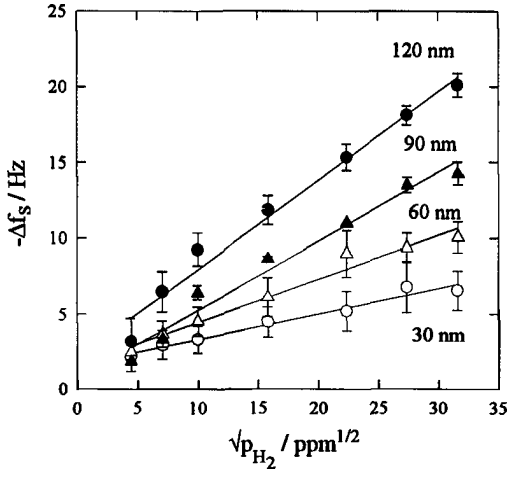


Figure 1. Responses of QCMs / PdNi to hydrogen gas as a function of the coating thickness. The straight lines are obtained from fitting the data using a least squares analysis.

Table 1 Sensitivity parameter of the sensor array

t / nm	S / Hz $ppm^{-1/2}$	S_n / Hz nm^{-1} $ppm^{-1/2}$	S_n^a
30	0.170	5.7×10^{-3}	
60	0.287	4.8×10^{-3}	5.13 ± 0.4
90	0.454	5.1×10^{-3}	
120	0.585	4.9×10^{-3}	

Table 1 lists the thickness, t, of four devices, slope values, S, of each plot, the thickness - normalized slope, S_n, and the sensitivity parameter, S_n^a, which is defined as the average of the normalized slope values. This parameter will be used as a criterion to determine the "cross reactivity" between H_2 and interference vapors for validating an interference compensation scheme.

II. Interference Tests: The PdNi coating has been tested for interference complications by exposure to potential interference gases. As an example, Figure 2 shows the resonance frequency
changes as a function the concentrations of water and acetonitrile vapors. In both cases, the responses decrease linearly with concentration but are not dependent of the thickness of the PdNi films. This behavior indicates that the response to interferences arises from only a surface adsorption process and not from uptake into the bulk of the PdNi coating.

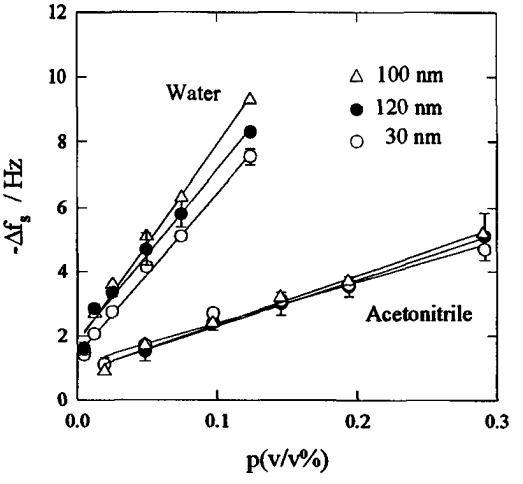


Figure 2. The responses of QCMs/PdNi to interference vapors.

III. Interference Compensation Scheme: Using the unique chemistry of Pd-H_2 system, a smart hydrogen gas sensor can be designed by fabricating a sensor array where each sensor is coated with a PdNi coating of a thickness t. The basis of this approach is as follows. Since the response of each sensor in the array, $\Delta f_j(t)$, where the subscript j denotes an individual sensor, is the sum of the response to hydrogen, $\Delta f_j^H(t)$, and to interference vapors, Δf_j^I, we can write

$$\Delta f_j(t) = \Delta f_j^I + \Delta f_j^H(t). \qquad (4)$$

Based on our results, we then expect a plot of $\Delta f_j(t)$ as a function of t to be linear. From an extrapolation to zero thickness, the response from interferences (Δf_j^I) can then be found and applied to the determination of hydrogen contents in the gas phase.

In summary, we have developed a new approach to the creation of high performance hydrogen gas sensors using quartz substrates. Future studies will explore the viability of our interference compensation strategy, transfer the findings to TFRs, and manufacture field deployable instruments.

References

[1] C. Christofides and A. Mandelis, "Solid - state sensors for trace hydrogen gas detection", J. Appl. Phys., vol. 68, pp R1-30, June, 1990

[2]. R. C. Hughes and W. K. Schubert, "Thin films of Pd/Ni alloys for high hydrogen concentrations", J. Appl. Phys. vol. 71, pp 542-544, January, 1992

[3] F. A. Lewis, The Palladium Hydrogen System, London, New York, Academic press, 1967, ch 2, pp 13-41

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A Micromachined Sensor Array Using Thin Film Resonators

J. Xia , S. Burns, M. Porter, T. Xue, G. Liu, R. Wyse, C. Thielen

Microelectronics Research Center, Ames Laboratory, Iowa State University,
1925 Scholl Road, Ames, IA 50011

Abstract

This paper describes the design, modeling and implementation of a microsensor array using aluminum nitride thin film resonators. A finite element method, formulated to accommodate the anisotropic and piezoelectric properties of aluminum nitride, is used to model acoustic wave coupling between resonators and to define the mask design. Two arrays on the same aluminum nitride membrane were fabricated using standard semiconductor processing. The measurements were in agreement with the modeling. A multilayer Mason model was applied to investigate the mass loading and viscoelastic effects via hydrogen absorption on the PdNi coated TFR. The preliminary H_2 sensitivity tests on a TFR hydrogen sensor are also presented.

Introduction

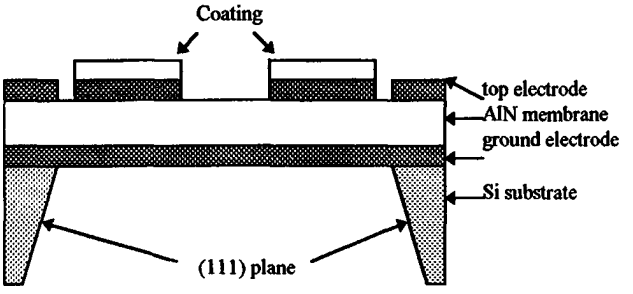
Aluminum nitride (AlN) thin film resonators (TFRs), often associated with communication system applications [1], have been developed as chemical sensors [2]. The detection at trace levels of volatile organic hydrocarbons (VOCs), hydrogen, nitrous oxide, ammonia and other materials can be obtained with specificity achieved via the affinity of an analyte to different coating deposited on TFRs. To minimize temperature and other environment-induced drift, and to investigate surface water or other vapor interference effects, a differential approach has also being developed. The outputs from several identical TFR-controlled oscillators with one passivated as a reference sensor and others as analyte selective sensors are mixed to produce a low frequency baseband output proportional to the gas phase measured [3]. Although individually diced TFRs can be used in this configuration, smaller sensors and better matching are obtained by patterning all the TFRs on the same AlN membrane. The main advantage of this design, in addition to smaller size, is that no AlN etching and patterning are required. Another way to construct an array is to pattern the AlN film to have individual TFRs on different AlN membranes. The advantages of this latter design are that more energy can be trapped in the device region and TFRs can be cointegrated with other on-chip active devices. This paper will concentrate on the first design since it is relatively easy to be fabricated and more compatible to the present TFR processing. The major challenge in this sensor array design is to evaluate the acoustic coupling between individual TFRs. These lateral waves, propagating in AlN membrane, consist of quasi-Rayleigh and acoustic plate modes (APM) that are mainly excited by electrode edge discontinuity. To optimize the array design with less trial and error, thus without time-consuming and cost, a two-dimensional transient finite element method was developed to model wave propagation between adjacent TFRs and to assist actual mask layout. After solving the coupling problem, individual TFRs can then be treated as one-dimensional devices since large width to thickness ratios (> 50) are used. The Mason model can be used to model wave propagation in the TFR composite structure. The array structures, modeling, processing and experimental data along with characterization of the TFR hydrogen sensor are presented in this paper.

Array Design and Modeling

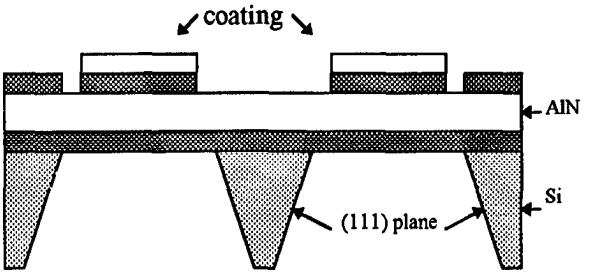
Two TFR array structures on the same AlN membrane are shown in Figure 1, with 1) TFRs on the same backside-window membrane, defined as the SM array, and 2) TFRs on different backside-window membrane, defined as the DM array. Individual TFRs have the edge-supported structure. c-axis normal, aluminum nitride is deposited by a DC planar magnetron sputter and is used as the piezoelectric membrane. The Si backside etch process is implemented so that the TFR can work at its fundamental

0-7803-2500-1/95 $4.00 © 1995 IEEE

thickness mode with ~ 1 GHz resonant frequency. Membranes with large width to thickness ratios are designed to have TFRs work into 50 Ω circuitry. Also the flexural spurious waves may be averaged to a small response by a large electrode area. The separation distance between individual TFRs depends on the anisotropic backside etch topology, alignment tolerance and, most importantly, acoustic wave coupling and electromagnetic crosstalks.



1) Array on the same backside-window membrane (SM)



2) Array on different backside-window membrane (DM)

Figure 1. TFR array on the same AlN membrane

The acoustic wave coupling between TFRs derives from the wave propagation laterally through the AlN film and can be measured as insertion loss. Bulk shear waves are attenuated exponentially and thus should not play an important role in the array design. Those surface and plate waves propagating in a complex dependence of amplitude of motion along lateral direction [4], may play a primary role for wave coupling. Since anisotropic and complex boundaries were encountered in the TFR array geometry, numerical solutions are usually required. A two-dimensional transient finite element method has been developed to evaluate the wave coupling. The calculation geometry is shown in Figure 2. The modeling scheme is similar to the method used in Nondestructive Evaluation (NDE) and Acoustic Microscopy. Mechanically free boundary conditions in all the surfaces and material uniformity are assumed.

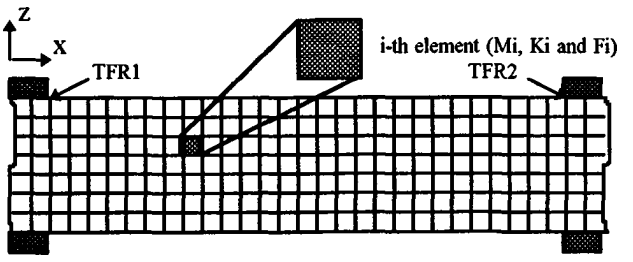


Figure 2. FEM mesh geometry

The displacement of the x-component is much smaller than that of the z-component because of normal electrical field and c-axis AlN membrane used [5]. A z-component displacement field of 1 GHz central frequency is assumed to be excited at the TFR1. A small amplitude is also assumed so that Newton's law and Hooke's law are applicable. Upon discretization of the problem domain to quadrilateral elements, the following matrix equation is created using either the weighted residual or energy functional approach [6]

$$M\ddot{U} + KU = F \qquad (1)$$

where K is the global piezoelectric stiffness matrix, M is the global mass matrix, F is the traction boundary condition, and U is the displacement vector. The electric potential, under quasistatic approximation, can then be expressed as [7]

$$\phi(x_1, x_3, t) = \frac{(l_i e_{iL} l_{Lj})}{l_i \varepsilon_{ij}^S l_j} \dot{U}_j \qquad (2)$$

The coupling level can then be approximately calculated as the magnitude damping in lateral direction and can be measured as insertion loss. The usual approach to solve equation (1) is by directly integrating through the explicit schemes of the center difference method which can be derived as

$$MU_{t+\Delta t} = F_t \Delta t^2 - (K\Delta t^2 - 2M)U_t - MU_{t-\Delta t} \qquad (3)$$

The left side has only the mass matrix which means that it can be lumped to a diagonal matrix so that the matrix inversion process is avoided. This technique may greatly improve the calculation efficiency, especially for very

short wavelength devices, such as TFRs. The iteration equation can then be written as

$$U_{t+\Delta t} = (M^D)^{-1}\Delta t^2 F_t - (M^D)^{-1} K\Delta t^2 U_t + 2U_t - U_{t-\Delta t} \quad (4)$$

M^D is the lumped mass matrix. Figure 3 shows the modeled magnitude variation along the lateral direction (x). The ripple data is due to the transverse resonance of plate waves. This simulated data was then used to guide the mask design and layout.

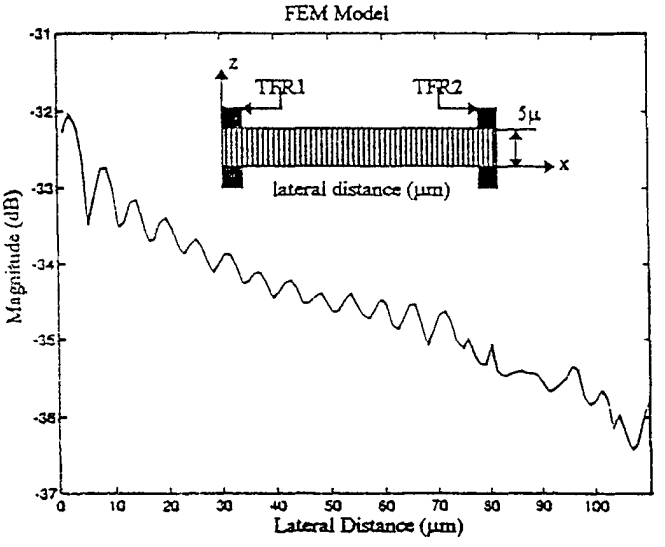


Figure 3. Magnitude vs. lateral direction

Processing and Measurements

Figure 4 is a representative of the array processing flow and for simplicity, only the major steps on the SM array are shown in the cross-sections. The bottom Cr electrode is used as the etch-stop for the backside window etch and as the adhesion layer. The c-axis AlN film with 5 μm nominal thickness is dc sputtered by a nitrogen plasma. Lift-off processing is used for the top electrode metallization. Selective chemical films are then spin-coated on top of electrodes. It has been found that the top metallization and Si backside anisotropic etching have significant effects on resonator performance. A rigid membrane and planar surfaces may be achieved by incorporating a thin CVD SiO_2 glass layer or Si_3N_4 layer in a TFR composite structure. This membrane-supported layer can not only protect the ground electrode, specifically aluminum, from the backside etch but also suppress the diffraction spurious responses due to rough surfaces.

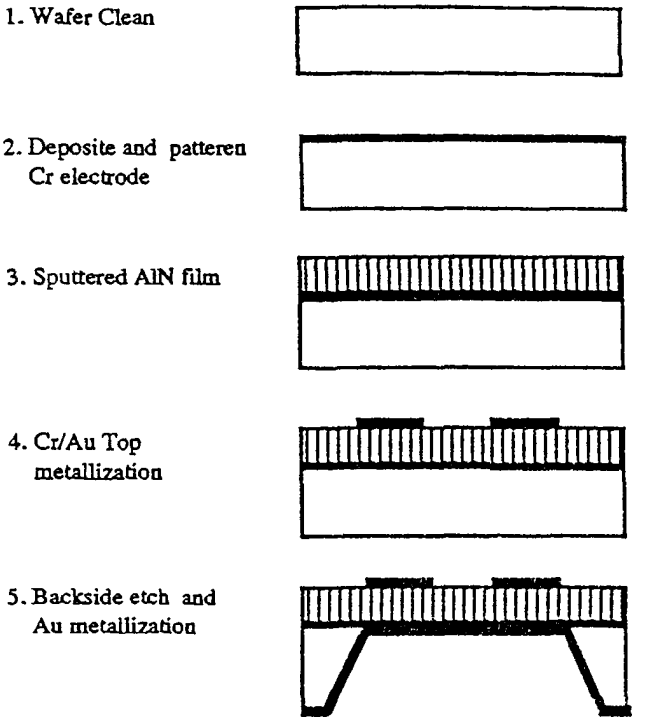


Figure 4. Array process flow

After the process is completed, an on-wafer measurement is conducted by using a Cascade Microtech microprobe station. One-port return loss and two-port insertion loss were then measured using a HP 8753C network analyzer. The coupling level near resonance of the SM array shown in Figure 5a is below 40 dB if the TFR edge-to-edge separation distance is above 400 μm. The rippled data are due to the AlN thickness variation across the wafer, which produces small resonant changes among individual TFRs. Figure 6 shows that the DM array has larger attenuation than the SM array because waves also propagate and attenuate in the Si substrate between adjoining TFRs.

TFR Hydrogen Sensor Characterization

TFRs have being designed and tested for a variety of sensor applications. Among them, H_2 mass sensors are discussed here. A thin PdNi alloy layer is deposited and patterned as sensing coating on top of TFR substrate. Absorbed hydrogen is first dissociated to H-H at the PdNi surface which then forms an irreversible H-Pd bonding network. Incorporated hydrogen adds mass to TFR which

then changes key parameters, such as series resonant frequency, series resistance and Q. Larger width to thickness ratios are used for matching TFRs to the 50 Ω measurement system. An extended one-dimensional Mason model [8] can be used to model wave propagation in TFR/PdNi composite. The difference between low frequency quartz sensors and high frequency TFR sensors, except for the different mode used, is that TFRs work at microwave frequencies so that all the metal layers are thus required to be included in the modeling. Figure 7, 8 show the modeled variations of resonant frequency, Q and series resistance of the TFR/PdNi sensor due to the mass and viscoelastic changes of PdNi layer via hydrogen absorption. It is found that viscosity does not change resonant frequency but vary circuit elements, such as motional resistance and Q value. The sensor characterization follows the EIA 512 standard [9]. The preliminary hydrogen sensitivity test on a PdNi coated TFR is shown in Figure 9. The PdNi thickness is 200 nm. Low detection limit of less than 5 ppm hydrogen was observed in a device. Also linear dependence of frequency vs. square root of hydrogen concentration has been found. Hydrogen irreversible absorption and the linear relation between frequency and hydrogen concentration suggest that mass loading may dominate the responses.

Conclusions

Two sets of aluminum nitride TFR arrays have been fabricated and tested. The coupling level between TFRs around 1 GHz is below 40 dB if the TFR edge-to-edge separation distance is kept beyond 400 μm. A finite element modeling has been shown in agreement with the measurements. It has also been found that the AlN membrane is very rigid; 500:1 or larger width to thickness ratios can be fabricated. The multilayer Mason model can be used to investigate the mass loading and viscoelastic effects of TFR/PdNi hydrogen sensor. Hydrogen sensitivity tests have shown that low ppm detection limit can be achieved. Through this study, we find that these array structures can be implemented as a differential chemical sensor array. They also open the possibility of a new and significant level of circuit cointegration. Work continues on incorporating TFRs in an array system where TFR is the frequency control element. TFR responses to hydrogen and other analytes are being studied.

Acknowledgment

This work was supported by the ETD program of the U. S. Dept. of Energy at the Ames Laboratory. The array fabrication was done at the Microelectronics Research Center of Iowa State University.

References

[1] K. Lakin, G. Kline, R. S. Ketcham, A. R. Landin, W. A. Burkland, K. T. Mccarron, S. D. Breymen, and S. G. Burns, *IEEE 41th Annual Symp. Frequency Control*, 29-31, May 1987

[2] R. O'Toole, S. Burns, G. Bastiaans, M. Porter, *Analytical Chemistry*, 64, 1289-1294, June 1992

[3] S. Burns and R. Weber, Invited paper, *IEEE 34th Symp. on Circuits and Systems*, 14-17, 1992

[4] R. M. White, P. J. Wicher, S. W. Wenzel, E. T. Zeller, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 1987, UFFC-34, 162-71

[5] K. Lakin, *IEEE 47th Annual Symp. Frequency Control*, 502-508, 1993

[6] T. Xue and W. Lord, *Review of Progress in QNDE*, vol. 14, 1995

[7] B. A. Auld, "Acoustic Field and Waves in Solid", I & II, John Wiley & Sons, 1970, 1990

[8] A. D. Ballato, H. L. Bertoni, and T. Timir, *IEEE Transactions on Microwave Theory and Techniques*, MTT-22, No. 1 (Jan. 1974), 14-25

[9] E. I. A., " Standard Methods for Measurement of the Equivalent Electrical Parameters of Quartz Crystal Units, 1 kHz to 1 GHz." ANSI/EIA-512-1-1990

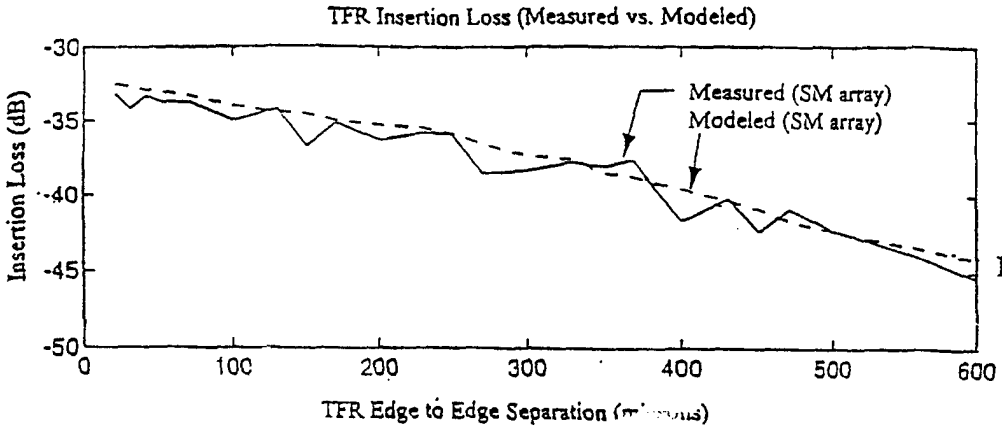


Figure 5. Modeled and measured coupling

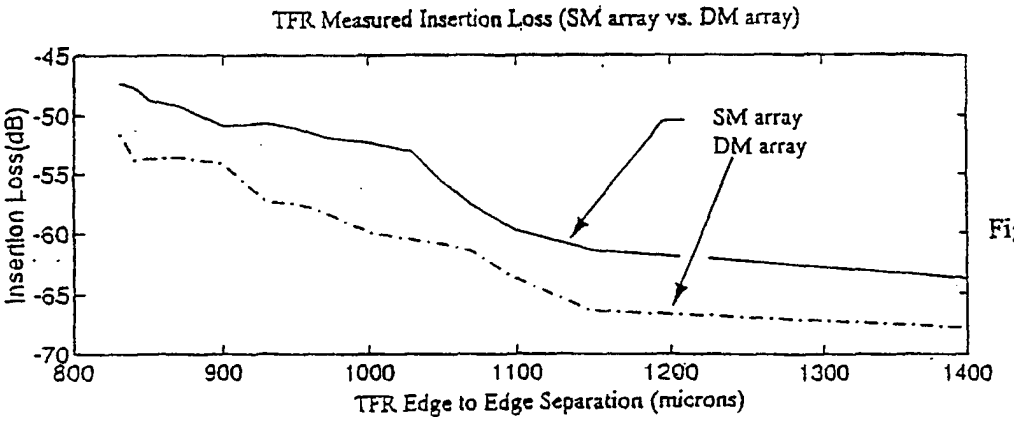


Figure 6. Measured insertion loss on the SM and DM array

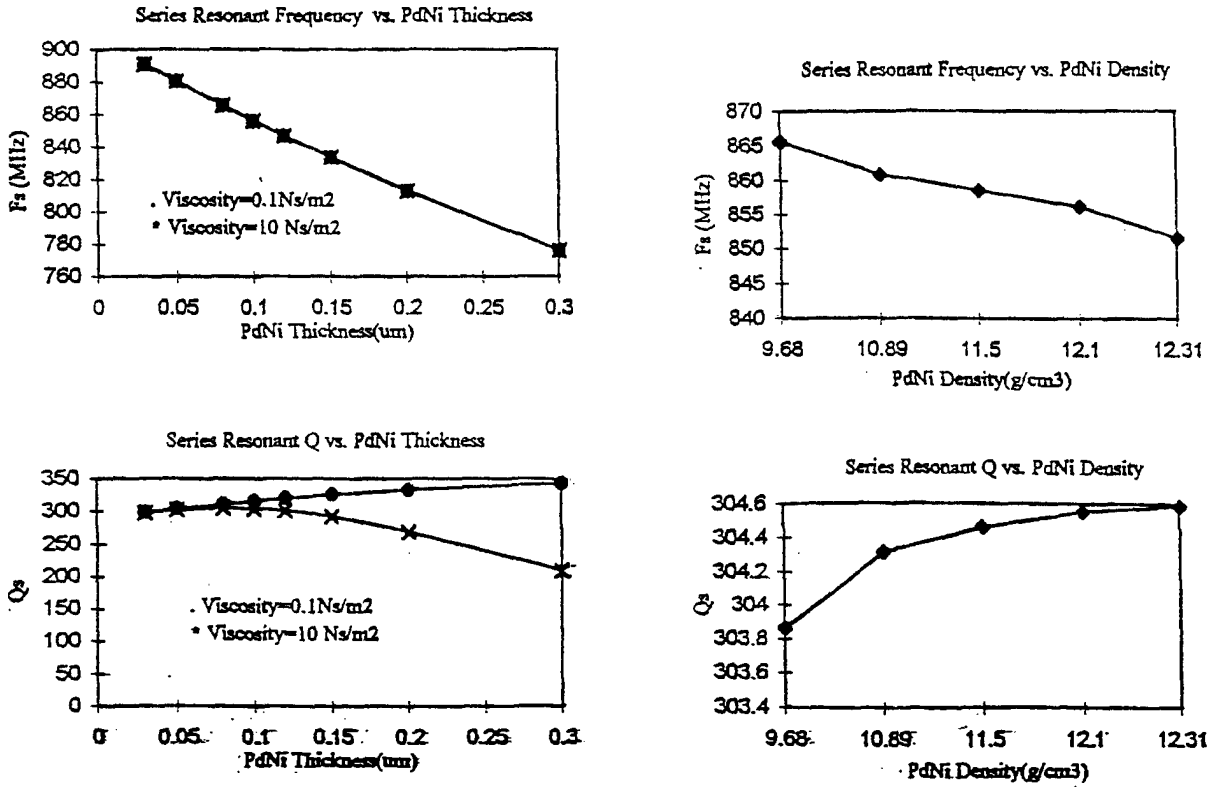
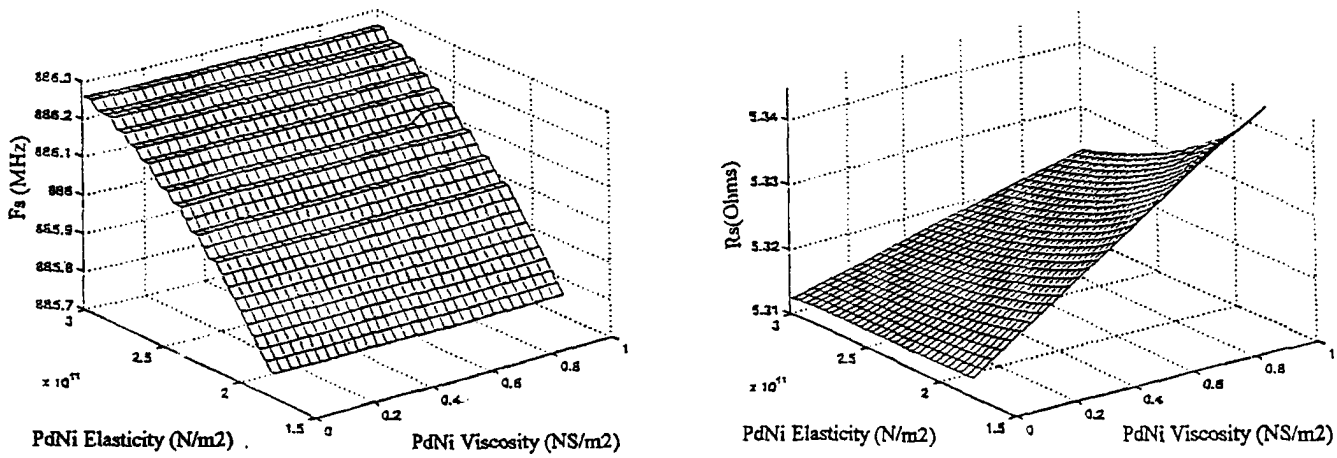


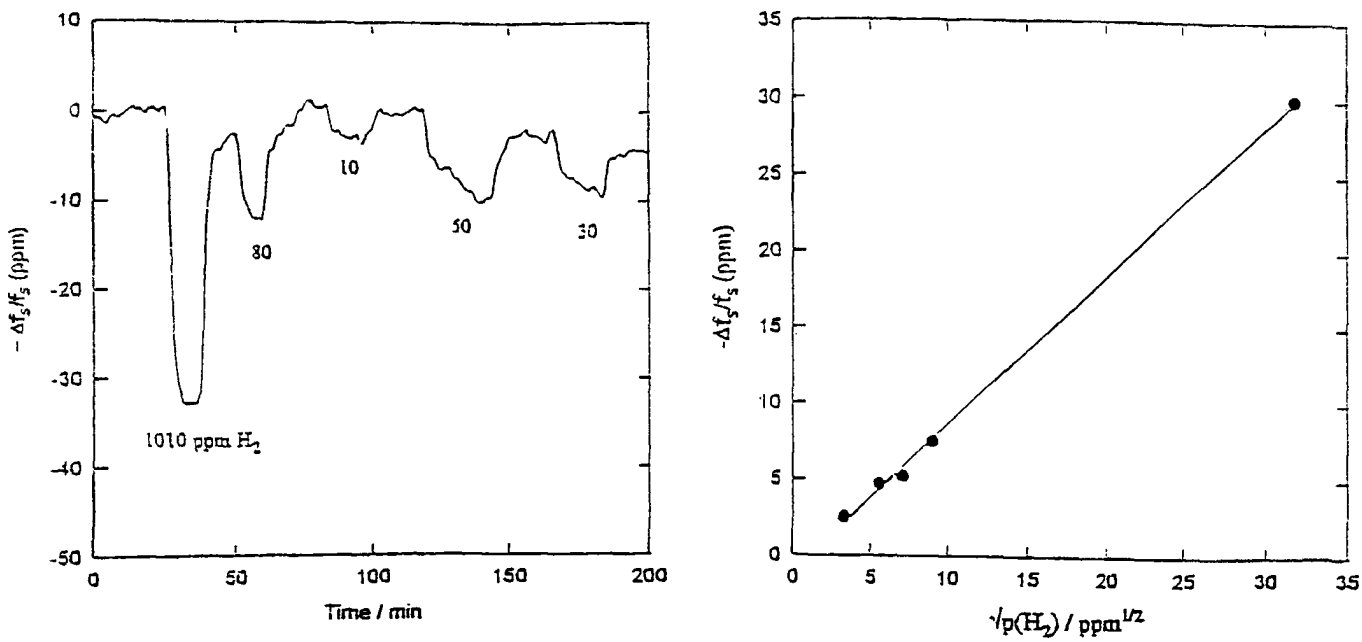
Figure 7. Modeled resonant frequency (Fs) and Qs vs. PdNi mass loading



a) Fs vs. PdNi viscoelasticity

b) Rs vs. PdNi viscoelasticity

Figure 8. Modeled Resonant Frequency (Fs) and Rs vs. PdNi viscoelasticity



a) Responses of TFR/PdNi to hydrogen

b) Frequency vs. hydrogen concentration

Figure 9. TFR/PdNi hydrogen sensitivity

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SPATIAL FREQUENCY-BASED MICROBENDING FIBER OPTIC SENSOR

Jill A. Brosig
Motorola Quartz Products Division
1301 E. Algonquin Rd.
Schaumburg, IL 60196

ABSTRACT: A new approach to fiber optic microbending sensors is demonstrated, where the spatial frequency of the bends, not the amplitude, is varied. This paper shows how to design this sensor, including the effect on sensitivity due to number of bends, bend amplitude, and deformation shape. By prebending the optical fiber into a series of bends, and then embedding it into a compliant material, we created a sensor whose spatial frequency can be changed by compressing the material. Results from our prototype exhibited the predicted transmission loss by alternating bend spacing. Due to the resonant nature of the mode-coupling effect, this modified sensor has the capability of equating to sensitivity levels found in the human finger tip. The application for the quartz industry is to use this sensor to automate those crystal processes which are currently restricted solely to manual means.

Introduction

Traditional microbending sensors have found a wide variety of applications. They can be used to measure pressure, acoustic waves, displacement, temperature, magnetic field, essentially any measurand that can be made to produce a change in the amplitude of the microbends. Taking the example of force, and using robotics as an application, a conventional microbend sensor can be used to detect forces in the robot's "finger", such as when it is grasping an object. The forces measured are by necessity normal to the gripping surface. There is a strong need in the robotics industry, however, for a sensor that can detect the component of force ***tangential*** to the gripping surface. For example, when grasping and lifting an object, a human can judge both the strength of the grip (normal force) and the weight of the object (tangential force). While current robots can perform gripping (normal) functions well, robotic technology lacks a good, compact tangential force sensor by which the robot can assess weight. The spatial frequency-varying microbending sensor proposed here is a possible solution to the tangential force problem. Furthermore, the sensor could also be designed to measure normal force, and a variety of other physical quantities including, but not limited to, pressure, force, displacement, fluid level, vibration, acceleration, electric or magnetic fields, and temperature.

The quartz industry has found that with larger, yet thinner quartz wafers, breakage becomes an issue. Due to their fragility, these pieces can currently only be handled by human hands. This consequently presents a barrier for cost reduction measures such as automation. With the formation of a "human" hand from this new type of microbending sensor, however, cost saving can be realized by using automation to handle quartz in an efficient and repeatable fashion.

Microbending Fiber Optic Sensor

A microbending sensor exploits mode coupling in a fiber that is deformed into a series of small bends. The bending results in coupling of energy between various guided modes and from guided to radiation modes, the latter causing transmission loss. The amount of loss depends on both the amplitude of the bends and the frequency of the bends. A conventional microbending sensor consists of a fiber that is passed between two corrugated plates as shown

0-7803-2500-1/95 $4.00 © 1995 IEEE

in Figure 1(a). As pressure is applied to the top plate, the fiber is compressed between the plates and bends are introduced. The greater the pressure, the deeper the bends, the larger the coupling of radiation from guided to radiated modes, and the lower the total throughput of the fiber.

In our approach, the fiber starts out in an already bent state, and is embedded into a compliant material. When a force is applied in the direction shown in Figure 1(b), the amplitude remains relatively constant, but the spatial frequency of the bends is changed.

Theory: To understand the mechanisms behind microbending loss, and the factors which influence sensor design, first consider an optical fiber of core index n_1 and cladding index n_2. For guiding to occur, n_1 must be large than n_2. The fiber supports some number N of guided modes, which depends on the V parameter. For a large number of modes in a step index fiber, N can be approximated by 161 (other similar but not identical approximations can be found in [7, 8])

$$N \approx V^2/2 \qquad \text{(step-index)} \qquad \{1\}$$

where the V parameter is given by

$$V = (2\pi a/\lambda)(n_1^2 - n_2^2)^{1/2} \qquad \{2\}$$

Each mode has a propagation constant $\beta_m = n_1 k \cos\theta_m$, in which θ_m represents the angle the mode's equivalent ray makes with the fiber axis. The fundamental mode, which propagates along the fiber axis, has $\beta_o = n_1 k$. The highest order mode that is still bound propagates at the critical angle for the core/cladding interface, $\theta_c = \cos^{-1}(n_2/n_1)$. Thus all guided modes have a propagation constant between $n_1 k$ and $n_2 k$, inclusive. The position of the highest order mode may be at $n_2 k$, or may be somewhere near it, depending on the total size of the β–space and the mode spacings.

All energy contained in the guided modes propagates down the fiber. Energy that is coupled into a radiation mode ($\beta < n_2 k$) is lost. Bending of the fiber causes power loss, and can be divided into two cases: macrobending, in which the radius of the bend is large compared to the fiber radius (bend radius large than -lcm), and microbending, which we discuss here.

We consider the case when a fiber is bent into a series of small radius bends which are evenly spaced. Energy is efficiently coupled from the mth mode to the nth (and back) when the spatial frequency A of the perturbation satisfies the condition that [10]

$$\lambda = 2\pi / (\beta_m - \beta_n) \qquad \{3\}$$

We now consider the distribution of the modes in β-space. The β's of the guided modes are quantized. There can be no mode with , $\beta < n_1 k$, and there is a continuum of radiation modes of $\beta > n_2 k$. Furthermore, there exist many degenerate modes; the modes are therefore categorized according to the principal quantum number m; each m labels a group of modes with nearly identical propagation constants. [11] The quantity M represents the total number of such mode groups, which is not the same as the total number of guided modes.

The distance in β-space between adjacent guided modes in a step-index fiber is given by [12]

$$\beta_{m+1} - \beta_m = (2(\Delta)^{1/2} / a)\,(m/M). \qquad \{4\}$$

In the case of a graded index fiber the spacing between modal groups in β space is [11]

$$\beta_{m+1} - \beta_m = (2(\Delta)^{1/2} / a), \qquad \{5\}$$

and does not depend on the modal group label m. Thus in graded index fiber, the modes are spaced evenly in β-space, meaning that one fixed spatial frequency of the microbends will cause

power to be transferred back and forth between all adjacent modes. For the step index case, however, the spacing between modes is dependent on the mode index. A particular λ will couple light between one specific pair of modes. Figure 2 shows how the modes are spaced for each type of fiber, where each vertical line indicates the location of β_m as a function of m. Note that for the same fiber parameters (that is, a, n_1, n_2, k), the step index fiber carries twice as many modes as the graded index fiber. These figures were generated for hypothetical fibers of the same radii, Δ's, and refractive indices ($n_1$step = $n_1(0)$graded). The total number of modal groups M is determined by how many modes will fit into the allowed β-space ($n_2k<p<n_1k$). The number of modal groups is related to the total number of modes by [11]

$$N = M^2 \text{ (step or graded index)} \qquad \{6\}$$

In Figure 2, the fiber parameters were chosen such that a relatively small number of modes would be carried, for illustration only. Figure 2a (graded) shows that the spacing between modes is constant regardless of mode number. Figure 2b (step) indicates that the lower-order modes are closer together than the higher modes. Bends of large λ in a step index fiber will therefore cause effective mode coupling among the lower order modes, but will not among higher order modes. A large λ will also not be conducive to coupling energy from lower into higher order modes, and the transmission will not be affected. A smaller λ, on the other hand, will cause intermodal coupling in the lower β (higher mode number) region, and energy can be coupled to radiation modes and lost.

Although in theory energy will be transferred back and forth in one specific mode pair in the step-index case, Figure 2b shows that the modes within a small region of β-space are virtually evenly spaced. In practice, therefore, mode coupling is induced among the modes in a neighborhood of β-space, and this region will be somewhat wider in the higher-order mode regime than among the lower order modes, whose spacing varies more rapidly.

In both the step and graded cases, coupling from higher order modes to radiation modes is desired for effective sensing. The are two important points to be noticed:

1) Between guided modes, it is just as likely for energy to go from lower to higher modes as to go from higher to lower. Once light is coupled from a guided mode to a radiation mode, it cannot return, so the highest order mode is being depleted of energy. Light will continue, however, to come into the highest mode from lower modes, so the net effect is to replenish the highest mode, and transmission loss continues to be observed.

2) In a graded index fiber, one spatial frequency will cause adjacent-mode coupling over the entire range of allowed modes, so that energy being supplied to the highest mode (and subsequently lost to radiation) can have originated from very low order modes. In a step index fiber, however, a particular λ will cause mode coupling only in a neighborhood in β-space, and the highest modes can eventually be depleted while still maintaining significant power in the fiber.

As the number of modes in a step index fiber becomes larger, the mode spacing for the higher order modes varies much more slowly from mode to mode than for a step fiber supporting a small number of modes. For a many-mode step index fiber, one can see that the highest order modes are spaced at virtually identical intervals. This implies that for a reasonable multimode fiber (N in the thousands), there will be little difference in microbend sensor sensitivity between step and graded fibers.

For a case in which a fiber is bent into a perfect sine wave at a fixed spatial frequency, the frequency spectrum of β is a δ function, and energy is coupled only between modes of the appropriate mode spacing $\beta_{m+1} - \beta_m$. Since the shape of the actual bends is generally not an exact sinusoid, but usually small curves around the deformer points separated by straight lines [13], we expect the frequency spectrum $F(\lambda)$ to have some spread. Mode coupling can therefore be expected over some range of mode spacings. Nevertheless, this spatial frequency spread is narrow, and the mode coupling, an essentially resonant effect, is significant only for mode spacings very close to $2\pi/\lambda$. Actual measurements of the transmission loss as a function of spatial frequency have been reported [11, 12] and the transmission loss shows a very sharp peak at the critical wavelength λ_c.

Sensor Design:Consider the transmission loss versus spatial frequency curve reported [12] for a graded index fiber, a sketch of which is shown in Figure 3. We exploit the sharpness of the loss versus spatial frequency curve to create a sensitive transducer. Instead of fixing λ and varying the amplitude of the microbends, as in conventional microbending sensors, we fix the amplitude and vary λ. If the sensor is prebent at some λ not at the peak, but rather on one of the steep sides of the curve, such as indicated by the dots shown in the figure, then a small change in λ should produce a large change in transmission. To vary the spatial frequency of the bends, we embed the sensing fiber in a compliant material. When the material is compressed or stretched, the fiber bends are compressed or stretched along with it, and λ changes. Figure 4 shows a detailed diagram of the sensor prototype.

To prebend and embed the sensing fibers, two molds were made from a section of a gear rack, of which the tooth spacing was 5mm. This spatial frequency is somewhat above the critical λc's needed for peak loss (2.22 and 2.64 mm for the .117NA and .146NA fibers respectively). Each mold contained nine complete periods. The fiber was tied loosely into each groove with fine wire to create the bends, which is estimated to be 0.5 mm peak-to-peak. Each mold was then filled with the gel and heated to 70°C. A second gel piece was similarly constructed for each sensor, but without a fiber. The two pieces of gel were then put together such that their teeth meshed, composing an essentially solid slab of gel with the prebent fiber embedded.

It should be noted that macrobends (those of large radius) can also cause mode coupling and transmission loss, and therefore are a source of noise. To eliminate the effect of macrobends, lead-in and lead-out fibers were attached to each end of the sensing fiber. The sensing fiber, having a lower NA than the lead-in fiber, cannot accept power from the higher order modes. Therefore, if the power in the highest mode of the lead-in fluctuates, it is not seen at the detector. This relationship of NA's also guarantees the sensing fiber will have all of its modes excited. The lead-out fiber is also chosen to have large NA, so that under ideal coupling conditions only its lower modes are excited, and thus more macrobending noise is avoided.

In addition, it is important that the lead and sensing fibers have the same core diameter, to avoid further losses and control mode excitation. Having similar cladding diameters also simplifies splicing. The lead fibers were chosen according to availability; both had NA's of 0.29, with core/cladding diameters of 100/140μm and 50/125μm, respectively. The arrangements are shown in Figure 5. The splices used were reusable mechanical insertion type splices, which were quite lossy.

The light source should be incoherent, to avoid speckle noise. An LED is the preferred source but we used a halogen bulb, due to the exceedingly high losses in the splices. The detector was a silicon photodiode.

Experimental Results

The prototype was tested as both an amplitude and frequency varying sensor. In each test the block was placed on a metal surface, and a metal plate was advanced in the appropriate direction to induce deformation. Because the gel maintains a constant volume, the sensor had to be constrained in one of the planes parallel to the direction of compression to prevent bowing of the gel and distortion of the bends. Also, the gel, which has a tacky surface, was dusted liberally with talcum powder to prevent adhesion and uneven compression. Talcum powder proved superior to oil or silicone grease.

Amplitude-varying test: The AV test was merely intended to confirm that our prototype worked as a traditional microbending sensor. The sensitivity was quite low, as was to be expected since the fiber was prebent and the soft consistency of the gel prevented much change in the bend height. A typical result is shown in Figure 6 for the 0.117 NA fiber. The abscissa is the overall height of the gel block.

AV sensors can exhibit two or three different slopes in the transmission curves. [3] The first region is generally considered to be due to the deformation of the fiber jacket, and the second region to represent the increased bending of the fiber core. Our experiments also yielded two different slopes. In Figure 6 both slopes are quite small compared to those reported in the literature since the spatial frequency λ had been intentionally selected to be off resonance.

Frequency-varing test: Figure 7 shows the estimated bias points and ranges for each sensor. These estimates are based on the assumption that the transmissions versus spatial frequency curves for our step index fibers are similar to that reported in [12] for graded index. This is reasonable, since it has been shown experimentally that there is negligible difference between the two types of fibers. [14]

To vary the spatial frequencies of the bend, the gel block containing the sensor was compressed in the longitudinal direction. The relaxed λ was 5 mm. Figure 8 shows intensity for λ from about 3.3 mm to about 4.4 mm for the 0.117 NA fiber. The most sensitive region is at the smaller λ's, consistent with data reported in the literature. [12] We appear to be operating in a region that extends close to the local minimum in Figure 3. We therefore expect higher sensitivity at lower λ's, since in this region the slope of the transmission curve is steeper.

Figure 9 shows data taken in the most sensitive region for both sensing fibers. The multiple runs show good repeatability.

Further Improvements: The prototype results establish the proof-of-concept. In Table 2 we compare our results to some other standard sensor technologies used for tactile sensors. Although our initial prototype is not competitive, with optimization this sensor can be designed to compete with the sensitivity of the human fingertip.

First, the compliant material was chosen because it has the advantage of being inexpensive, moldable, and can be made to various degrees of stiffness. It is possible, however, that more appropriate materials exist. If the same gel is used in future sensors, the tackiness of the surface needs to be addressed. The sensitivity of the sensor will depend on the stiffness of the material.

There are several ways to improve the sensitivity. The more closely the shape of the fiber approaches a true sine wave, the narrower the spatial spectral distribution, resulting in a sharper resonance. Also, increasing the number of microbends is expected to increase the sensitivity, as would increasing the amplitude of the bends. Choice of bias point will also have an effect.

The sensitivity will be traded off against dynamic range. To be able to measure large ranges of force (displacement, pressure, temperature, etc.) one wants the slope of the transmission curve to be smaller, and the resonance less pronounced.

Finally, the proximity of the highest guided mode to the edge of the radiation continuum in β-space is determined by the design of the fibers. In the examples chosen in Figure 2, the highest guided mode happened not to be the same distance from the continuum as from the next lowest mode. If the highest guided mode is very close to n_2k, then there is a large range of λ's that will effectively couple this mode to radiation. If the highest order mode is farther from the radiation modes, there is a smaller range of λ's that will cause loss. In any case, it still requires a specific λ to couple light up from lower modes.

Summary

A new type of microbending sensor has been demonstrated, in which the spatial frequency, rather than the amplitude, of the microbends is varied. By prebending the optical fiber into a series of bends, and then embedding it into a compliant material, we create a sensor whose spatial frequency can be changed by compressing the material. Our prototype exhibited the predicted transmission loss with decreasing bend spacing, albeit with very low sensitivity (0.1% change in transmission per mil of displacement versus 3.62%/mil for amplitude-varying microbending sensors). We can ultimately expect a very high sensitivity in an optimally designed frequency varying sensor, due to the resonant nature of the mode-coupling effect.

The prototype sensor was designed to operate on the steep slope of one side of the resonance curve (transmission versus spatial frequency) such that small changes in λ result in large changes in transmission. Data was obtained over small ranges of compression, which can be expanded significantly with some engineering. Furthermore the sensitivity can be increased by increasing the bend amplitudes, number of bends, decreasing the compressibility of the compliant material, and judicious choice of sensing fiber. We also project that dynamic range will ultimately be higher for graded index fiber than for step.

Although originally conceived for detection of forces tangential to a surface (as in a robotic finger, Figure 10), this sensor concept can be engineered to detect a wide variety of measurands. These include but are not limited to pressure, force, displacement, fluid level, vibration, acceleration, electric or magnetic fields, temperature, just as for traditional microbend sensors.

Acknowledgments

My profound thanks to Betty Lise Anderson of The Ohio State University whose inspiration it was to pursue this research, N. Lakagos of the Naval Research Laboratory for many helpful conversations, and to John Shen of Applied Elastomerics for supplying and shaping the compliant material for me.

References

1. T. G. Galliorenzi, J. A. Bucaro, A. Dandridge, J. G. H. Sigel, J. H. Cole, S. C. Rashleigh, and R. G. Priest, "Optical fiber sensor technology," IEEE J. Quant. Electr., QE-18, p. 626-665 (1982).

2. D. H. McMahon, A. R. Nelson, and W. B. S. Jr., "Fiber-optic transducers." IEEE Spect., December, p. 2s29 (1981).

3. D. A. Krohn, *Fiber Optic Sensors - Fundamantals and Applications*, (Instrument Society of America, Research Triangle Park, NC, 1988).

4. K. A. Murphy, A. Wang, M. K. Burford, and M. F. Gunther, "Large structures provide many applications for FO sensors." Photonics SPect., July, p. 133-136 (1992).

5. E. Udd, J. P. Theriault, A. Markus, and Y. Bar-Cohen, "Microbending fiber optic sensors for smart structures," in *SPIE Smart Structures and Skins II*, Vol. 1170, p. 478-482 (SPIE, 1989)

6. D. Gloge, "Weakly Guiding Fibers," Appl. Opt., 10, p. 2252-2258 (1971).

7. A. H. Cherin, *An Introduction to Optical Fibers,* (McGraw-Hill. Bell Telephone Laboratories. 1983).
8. M. A. KaAm, *Electro-Optical Devices and Systems,* (PWS-Kent, Boston. 1990).
9. D. Gloge and E. A. J. Marcatili. "Multimode theory of graded-core fibers," Bell Sys. Tech. J.. 52. p. 1563-1578 (1973).
10. D. Marcuse. *Theory of Dielectric Optical Waveguides,* Second ed. (Academic Press. San Diego. 1991).
11. J. N. Fields. "Attenuation of a parabolic-index fiber with periodic bends," Appl. Phys. Lett.. 36. p. 799-801 (1980).
12. N. Lagakos. J. H. Cole. and J. A. Bucaro. "Microbend Fiber-Optic Sensor," Appl Opt. 26. p. 2171-2180 (1987).
13. H. F. Taylor. "Bending effects in optical fibers." IEEE J. Lightwave Tech.. LT-2. p. 617-628 (1984).
14. N. Lakagos. "private communication." 1991).
15. "Sensors and transducers. gauging physical qualities." Machine Des.. 135-150 (1987).
16. D. M. Siegel. I. Garabieta. and J. M. Lollerbach. "An integrated tactile and thermal sensor." in *IEEE International conference on robotics and automation.* p. 1286-1291 (IEEE.1986).

Figure 1 (a) Conventional. or amplitude-varying microbend sensor. Bend height is a function of measurand. (b) New. or frequency-varying microbend sensor. Bend frequency is a function of measurand.

Figure 2. Mode group spacing for two fibers of same parameters (a=25>. n1=1.45,n2=1.4316. A=.01267. B=1Rm). (a) Graded carries N=328 modes or M=18 modal groups. (b) Step index fiber carries N=656 modes, or M=25 modal groups.

Figure 3. Hypothetical change in transmission versus bend spacing. This closely approximates actual data reported in [12]. The dots represent possible bias points.

Figure 4. Sensor prototype. The light source is a halogen bulb. Lead fibers reduce noise. Sensing fiber is prebent and embedded in a compressible gel material.

Figure 5. Arrangement of lead fibers and sensing fibers. Leads have high NA and sensing fiber has low NA, to reduce noise. Cladding diameters are matched to ease splicing; ideally coresizes should be matched for maximum coupling efficiency.

Figure 6. Transmission as a function of sensor deformation in amplitude-varying (AV) configuration. Note the presence of two distinct regions.

Figure 7. Estimated bias points and sensing ranges.

Figure 8. Intensity as a function of bend frequency for the NA=0.117 fiber. Sensor material did not compress uniformly outside the range shown; sensor could not be compressed at all beyond λ=3.3 mm. The value of λc for this fiber is 2.22mm.

Figure 9. Intensity as a function of bend spacing over most sensitive region. (a) Sensing fiber NA=0.117 (λc = 2.22mm); (b) sensing fiber NA=0.146 (AC=2.64mm). Multiple runs are shown to demonstrate repeatability.

Figure 10. Example of application: detection of forces tangential and normal to a robotic finger.

TABLES

Table 1. Sensing fiber parameters. Number of modes calculated at λ=900nm. Numerical aperture is given by NA, n_1 is the index of refraction of the core, Λ_c is the critical mechanical wavelength, at which the maximum loss occurs, N is the total number of modes, and M is the number of modal groups.

Fiber	NA	Core Size	Cladding Size	n_1	Λ_c	N	M
Step	0.146	100μm	140μm	1.46	2.64 mm	1162	34
Step	0.117	95.3μm	125μm	1.457	2.22 mm	653	25

Table 2. Comparison with other tactile sensor technologies. The units "%/mil" refer to percent change in transmission per mil of displacement. The term ΔX refers to change in length or distance.

Sensor type	Type of force measured	Responsivity or sensitivity	Sensing range
Frequency-varying microbend sensor prototype (step)	Tangential to surface	-.1%/mil	0.021"ΔX between bends
Lakagos * (Graded) [12]	Tangential	4.5%/mil	0.039"ΔX between bends
Amplitude-varying microbending [3]	Normal	3.61/mil	0.018"ΔX of deformers
Piezoelectric [15]	Stress and Strain	$10^{13}\Omega$m	10^{-8} to 10^{9} N/m
Capacitive [16]	Normal	not reported	0-200 grams

*Lakagos reported effect of spatial frequency on transmission; this device not designed as sensor.

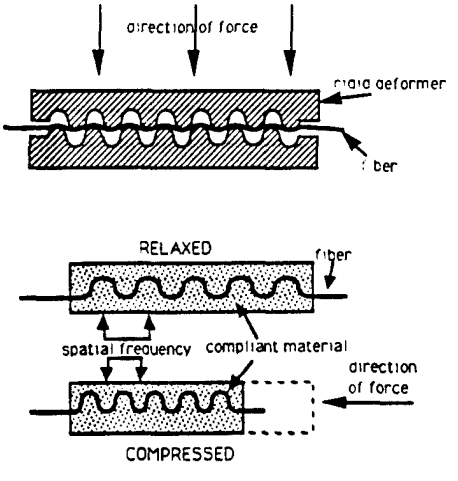


Figure 1.

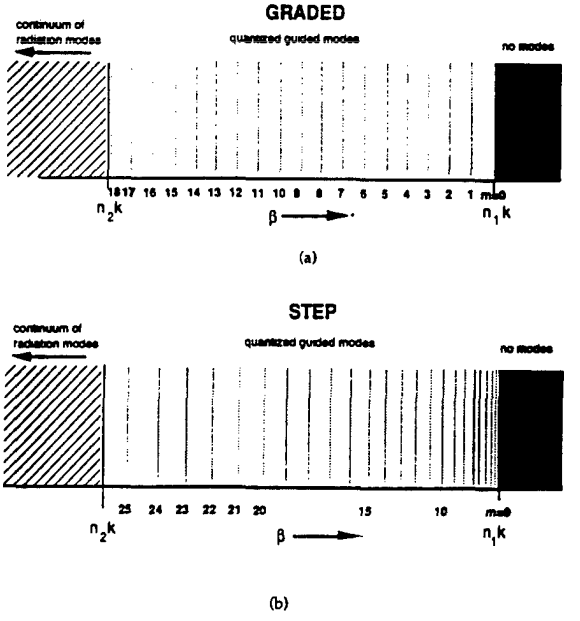


Figure 2.

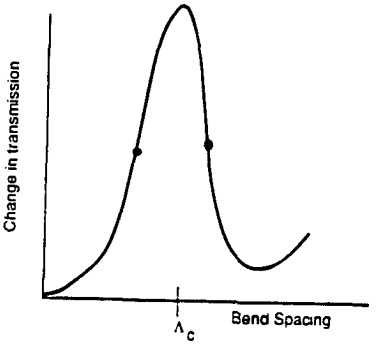


Figure 3.

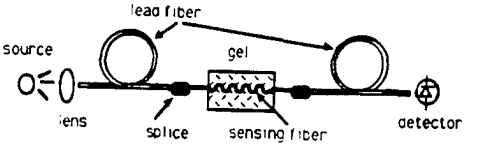


Figure 4.

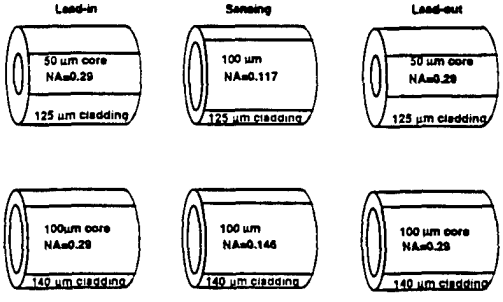


Figure 5.

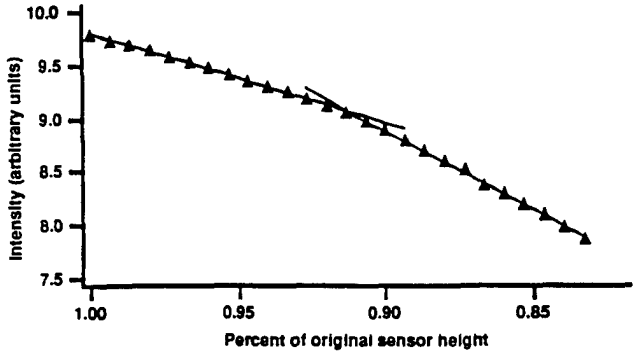


Figure 6.

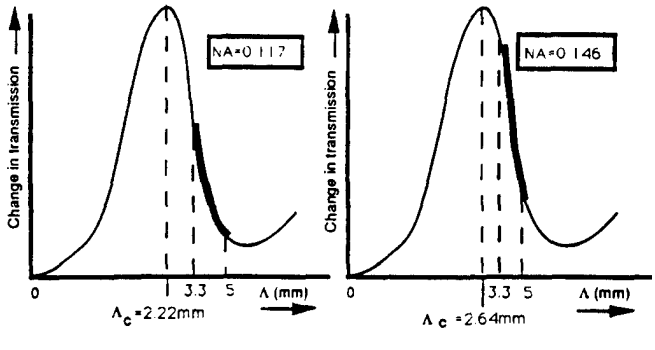


Figure 7.

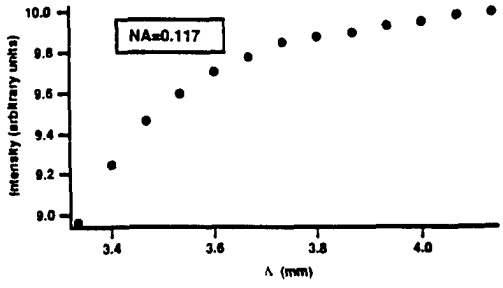
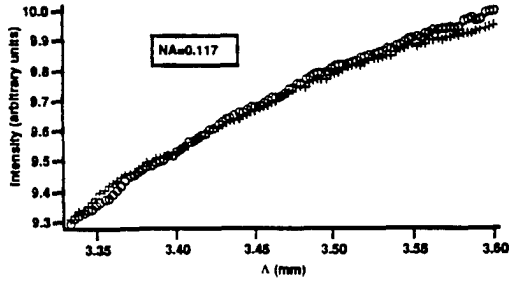
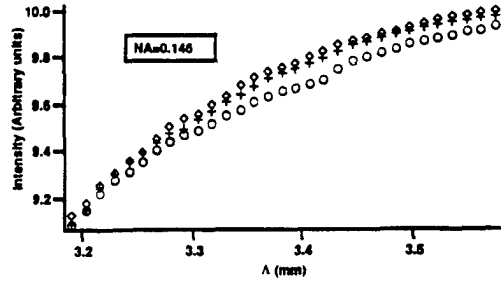


Figure 8.

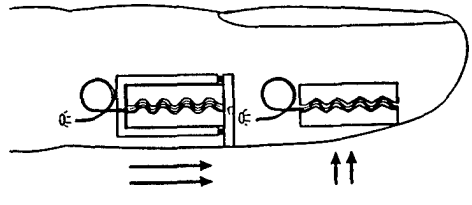


(a)



(b)

Figure 9.



1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ASPECTS OF ACTIVITY OF IEC TECHNICAL COMMITTEE NO. 49 "PIEZOELECTRIC AND DIELECTRIC DEVICES FOR FREQUENCY CONTROL AND SELECTION"

L.I. Zhourkina, Joint Stock Company "Phonon"

44, Krasnobogatyrskaya str., 105023, Moscow, Russia

ABSTRACT

This paper briefly describes the evolution attained in IEC/TC 49 during 30 years of its activity aimed at ever growing industry needs satisfaction, explains the procedure used in preparing IECQ specifications, gives certain details on IEC Standards development by using quartz crystal units as an example, shows the scientific approach used to documents preparation and gives the nearest tasks of its further activity.

INTRODUCTION

The International Electrotechnical Commission (abbreviated IEC) is an organization which was formed as a result of resolution of the delegates at the International Electrical Congress held in St. Louis in September 1904. The object of the Commission is to promote international cooperation on all questions of standardization and related matters in the fields of electrical and electronic engineering and, thus, to promote international understanding. The object is achieved by issuing publications, including recommendations in the form of international standards, which the National Committees are expected to use in their work on national standards, in so far as national conditions will permit. These standards are developed in technical committees and sub-committees. At present there are 200 such committees and sub-committees dealing with specific fields of electrotechnical and electronic engineering. Technical Committee 49 (TC 49), Piezoelectric and dielectric devices for frequency control and selection, was set up in 1961. Prior to that the activity on quartz crystal units for frequency control had commenced in SC 40-3 of TC 40, Capacitors and resistors for electronic equipment. Since that time, the scope of TC 49 has enlarged regularly to encompass new piezoelectric devices as the measure of their production volume has increased up to mass production. The duties of the international Secretariat of TC 49 are carried out by the Russian National Committee since December 1964. Prior to that time this was the responsibility of the Netherlands National Committee. The Secretariat of any technical committee is, in general terms, an executive body responsible for the completion of the work undertaken at the earliest possible time and for further progress in this Committee activity. The Secretariat acts under the guidance of the Committee Chairman and IEC Central Office seated in Geneva. Mr. D.C. Bradley from the U.S. National Committee is Chairman of TC 49 at present.

Both the Chairman and the Secretary must work in a pure international capacity. The duties of the Secretariat incorporate preparing time schedules, programs of work and the Strategic Policy Statement, organization of meetings, work at meetings, preparation of documents, and control and coordination of work delegated by technical committees.

SCOPE OF TC 49 ACTIVITY

The last scope of TC 49 approved by the IEC Committee of Action in 1990 reads as follows:

"To prepare international standards for piezoelectric and dielectric devices for frequency control and selection (excluding piezoelectric transducers dealt with by TC 39 and TC 60) and for associated materials".

Prior to that the basis of TC 49 activity was the physical effect of piezoelectricity. Now the criterion has been changed to resonant devices used for frequency control and selection.

The present scope of TC 49 activity includes preparing IEC Standards on the following devices and materials: quartz crystal units, crystal filters, crystal controlled oscillators, piezoelectric ceramic resonators and filters, surface acoustic wave (SAW) devices (resonators and filters at present), dielectric resonators and filters, piezoelectric crystal and piezoelectric ceramic materials, piezoelectric and dielectric device parameter measurement methods, terminology and outlines. These standards are developed and discussed

0-7803-2500-1/95 $4.00 © 1995 IEEE

in pertinent Working Groups consisting of international experts in the above areas of engineering. Controversial problems are discussed and resolved at the Plenary Committee meetings. At present 10 such permanent Working Groups act in TC 49 covering all the scope of the Committee.

List of Working Groups and, Past and Present WG Chairmen (Conveners)

Working Group No.	Date of est.	Designation	Working Group Chairmen
1	1995	Quartz crystal units for frequency control and selection	Messrs. C.S. Brown, P.C. Tilmouth, E.W. Kentley (United Kingdom)
	1990	Coordination of work on preparing IECQ specification in different Working Groups	E.W. Kentley (United Kingdom)
2	1966	Crystal filters	R.A. Spears (United Kingdom); Prof. M. Onoe (Japan); Mr. B. Neubig (Germany)
3	1968	Temperature control devices quartz crystal units (Disbanded in 1992 and merged with WG 7)	E.W. Kentley (United Kingdom)
4	1968	Piezoelectric ceramic resonators and filters.	Mr. D.J. Grevink (Netherlands), Mr. S. Fujishima (Japan)
5	1968	Materials for frequency control	Mr. S.A. Stevenson, Mr. C.S. Brown (United Kingdom), Messrs. R.A. Sykes, D.R. Kinloch, Dr. B. Sawyer, G. Johnson (U.S.A.)
6	1969	Measuring methods	Mr. C. Franx (Netherlands), Mr. J. Dowsett (United Kingdom), Mr. F. Fick (Germany), Mr. W.L.Smith (U.S.A.), Messrs. R. Williamson, C.S. Morris (United Kingdom)
7	1971	Quartz crystal controlled oscillators	Mr. W.L. Smith, Prof. D. Malocha (U.S.A.), Mr. Y.S. Yerna (Netherlands)
8	1980	Preparation of Chapter 561 of the International Electrotechnical Vocabulary	Messrs. G. Gilbert, Ch.Beauvy (France), Dr. D. Fuchs, Mr. F. Fick (Germany), Mr. S. Okano (Japan)
9	1983	Enclosures and related techniques	
10	1983	Surface acoustic wave devices and	Messrs. G. Tobolka, R. Veith, G. Riha (Germany)
	1990	dielectric devices	G. Riha (Germany)
11	1992	Glossary	Mr. D.R. Woodall (United Kingdom)

HISTORY AND OPERATION

Among all the devices standardized by TC 49, quartz crystal units are the most important, since they fulfill two essential functions: frequency control (quartz crystals for oscillators) or frequency selection (quartz crystals for filters). Therefore the process of TC 49 standards preparing, their structure and further evolution will be shown by using quartz crystal units as an example.

At the beginning of its work, the Committee devoted itself exclusively to preparing standards for frequency control crystals, resulting in the appearance in 1962 of Publications IEC 122-1, IEC 122-2 and IEC 122-3, Quartz crystal units for oscillators, these forming a comprehensive series.

The second edition of this series issued in 1976, 1977 and 1983, respectively, covered crystal units both for frequency control and selection.

In Addition to this, IEC Standards for piezoelectric and dielectric devices are prepared by TC 49 according to a uniform presentation, ("pattern"), covering all aspects of technical specification of a device:

Part 1. Chapter 1: General information (including terms and definitions), standard values of parameters and Chapter 2: Test conditions.

Part 2. Chapter 3: Guide to the use of....

Part 3. Chapter 4: Standard outlines and Chapter 5: Lead connections.

Part 4. Chapter 5: Article sheets

For example, the complete IEC 122 Publication series is entitled as follows:

122-1 "Quartz crystal units for frequency control and selection. Part 1: Standard values and test conditions".

122-2 "Quartz crystal units for frequency control and selection. Part 2: Guide to the use of quartz crystal units for frequency control and selection".

122-3 "Quartz crystal units for frequency control and selection. Part 3: Standard outlines and lead connections".

Part 4, Article sheets, has never been pursued because of the difficulty in assembling a collection of standardized article sheets due to the fact that they were previously custom-made devices as a rule. Therefore article sheets have been replaced by a check list of parameters to be specified in detail specifications and incorporated in the Guide to the use.

This structure is observed by TC 49 up to the present time, for those devices for which no IECQ standards are yet available. With the advent of IECQ Standards, TC 49 has received article sheets and this structure is supplemented by sectional and blank detail specifications which will become Part 4 and Part 1 will be replaced by IECQ generic specifications.

At the beginning of eighties the world-wide recognition of the importance of product quality caused establishment of IEC Quality Assessment System for Electronic Components (IECQ). The decision reflected a recognition of the importance that writing test methods and procedures and general performance requirements represented only one part of the market. The market requires, in addition to these important aspects, of standardization of product description, a means to ensure that the products actually being supplied do conform to these specifications. This "conformance-to-specification" need required two things that rarely, if ever, were included in IEC documents. The first of these were the provisions for demonstrating that the product was capable of meeting the requirements of the specification and that it continues over time to meet these requirements. The second missing element related to assuring that the demonstration provisions are being complied with, and providing such assurance in a form that the buyer in one country could accept such assurance from a manufacturer in another country.

IECQ was created to satisfy both of these needs. The specification provisions, however, are the responsibility of technical committees involved. The Advisory Committee on Electronics and Telecommunication (ACET) has prepared IEC Guide 102 for technical committees. It lays out specification structures for including quality conformance requirements in specifications. A system has been also established, where by manufacturers in a participating country would operate under the surveillance of an inspectorate in their own country in full compliance with established international rules.

Two procedures of certification under the IECQ are used: qualification approval, i.e. certification of detailed products, which are then given in the IEC Qualified Products List and capability approval, i.e. capability of a manufacturer to fabricate certain products. More detailed definition of capability approval procedure is:

"Capability approval is an approval granted to a manufacturer when it has been established that his capability for manufacturing processes and quality control methods, covering a range of components, the limits of which are defined by the manufacturer in accordance with the relevant specification, fulfills the requirements of the relevant generic and/or sectional specification. This procedure is used when it is impractical to use qualification approval procedures for some components, such as custom-built components or specialized components". Components released under the capability approval procedure have the same status within the IECQ System as those released under qualification approval procedure.

TC 49 decided to use capability approval approach for quartz crystal units, crystal filters and

oscillators.

The first TC 49 series of specifications issued under the IECQ deals with quartz crystal units and covers both qualification and capability approaches. These specifications are based on the Electronic Components Committee (ECC) specifications of the Joint European Standards Institution and are very widely used in Europe. These standards are:

1178-1 (1993) Quartz crystal units - A specification in the IEC Quality Assessment System for Electronic Components (IECQ) - Part 1 Generic specification.

1178-2-1 (1993) Part 2: - Sectional specification - Capability approval - Section 1 - Blank detail specification.

1178-3 (1993) Part 3: Sectional specification - Qualification Approval

1178-3-1 (1993) Part 3: Sectional specification - Qualification Approval - Section 1. Blank Detail Specification.

At present many IECQ specifications are already issued or discussed in TC 49. These specifications, however, have come from different sources and are different in layout and contents. Therefore, the Working Group on crystal units has been entitled to coordinate them for all devices, by taking the specification structure used for crystal units as a reference.

The extraordinary rate of development of crystal units during 30 years by over 1000 times and their applications requires continuous updating of existing standards and preparation of new ones. This applies, for example, to specifying requirements for temperature ranges, tolerances on load capacitances, drive level dependency, leak rates, circuit conditions, etc. TC 49 was the first to define the term "load resonance frequency" and to give a mathematical expression for it in response to the user demand. The same applied to "load resonance frequency offset", "fractional load resonance frequency offset", "frequency pulling range", "fractional pulling range" and "pulling sensitivity", which have been later inserted by many manufacturers in their commercial catalogues. All these up-dated requirements are contained in IEC 1178-1 Generic Specification, which has completely replaced IEC 122-1.

TC 49 always moved logically and smoothly based on the industry demands and pressure from the industry. When resonators for wrist watches became available, the Committee standardized the most widely used outlines and a specification for them, incorporating requirements, test and measuring methods (IEC 689). Revision of this publication is needed to take into account the present state-of-the art.

It has also become apparent, that IEC 122-2 dealing with the Guide to the use of quartz crystal units, the second edition of which has been compiled jointly by representatives of several National Committees cannot give complete recommendations on the use of products available in the industrial market. The present edition has been prepared by the following National Committees: The U.S. National Committee has provided a section on oscillator schematics, the Japanese National Committee has made a contribution to the section on unwanted responses, the U.K. National Committee has prepared a section on motional parameters, and Mr. E.W. Kentley, the Chairman, has added information on filter crystals and cost factors. The Swedish National Committee has submitted data on aging.

The decision has been made by TC 49 to prepare an additional section to the Guide, covering crystal units for microprocessor clock supply (IEC 122-2-1 (1991), which addresses some of the special problems encountered in practice in specifying and using crystal units for frequency control with integrated circuit oscillators, commonly incorporated in microprocessor silicon devices. These problems include phase delay and stray parameters of active circuits which often result in unpredictable behavior of the oscillator circuit. These applications commonly involve large volume manufacture and are very sensitive to both reliability and cost. Amendment to this standard released in 1993, addresses problems of limiting drive level of crystal units used with digital gates and on-chip oscillators up to permissible values (measurement of the crystal current and method of the reduction of the crystal current), and to describe a circuit enabling third overtone crystal units to be used in the frequency range above 25 MHz.

The main Guide (IEC 122-2) needs to be revised or obsoleted.

The work on piezoelectric filters, which commenced in 1966 and proceeded in the same way and under the same structure as for crystal units, has resulted in release of the Publication series: IEC 368-1, IEC 368-2 (IEC 368-2-1 and IEC 368-2-2)* and IEC 368-3. IEC 368-1, Test and measuring conditions, the third edition of which was printed in 1992. It recommends new methods of transmission characteristic measurements by using a network analyzer and describes return of attenuation and intermodulation distortion measurements, which were missing in the previous edition. These standards are applicable both to crystal and piezoelectric ceramic filters. Only guidance documents are different for them: IEC 368-2-1 applies to Guide to the use of crystal filters, IEC 368-2-2, which is at the voting stage now, will cover Guide to the use of piezoelectric ceramic filters. (It will replace the present edition in IEC 368B). Standards under the IECQ are released for piezoelectric

---

* Full titles of IEC Standards are given in the Specifications and Standards List Relating to Frequency Control placed at the end of Proceedings.

ceramic filters in IEC 1362-1, IEC 1362-2 and IEC 1362-2-1 series. Only qualification approval procedure is recommended for their certification. IECQ Standards for crystal filters are at the proposal stage so far, which is expected to be submitted within 1-2 months. The capability approval procedure will be used for crystal filter certification.

The work on piezoelectric ceramic resonators was completed last year: IEC 642, IEC 642-2, IEC 642-3 and IEC 483 are valid relating to this field of TC 49 activity. Standards under the IECQ are also released for piezoelectric ceramic resonators in IEC 1253-1, IEC 1253-2 and IEC 1253-2-1 series. Qualification approval approach is used for their certification. It is necessary to note here that IEC standardization is a continuing process aimed at improvement and revision of existing standards. Although IEC 483, Guide to dynamic measurement of piezoelectric ceramics with high electromechanical coupling, has been reconfirmed in 1994 for another period of 5 years. It needs to be revised as it is rather outdated. As soon as the proposal on its revision, assigned to the Japanese National Committee, is available, the work on piezoelectric ceramic resonators will resume.

The work on quartz crystal controlled oscillators began in 1971. All standards in accordance with the above-mentioned structure for devices have been released. They are: IEC 679-1, IEC 672-2 and IEC 679-3.

IECQ Specifications for quartz crystal controlled oscillators, covering both capability and qualification approval procedures, are under preparation for final voting prior to circulation as Draft International Standards (DIS). These specifications are also based on CECC Standards. IECQ Generic Specification when printed will replace IEC 679-1 completely. It is expected that all the series will be released next year. Further activity of this Working Group will be connected with the revision of the Guide to the use of crystal controlled oscillators (IEC 679-2).

The first request to incorporate surface acoustic wave (SAW) devices in the scope of TC 49 has been made by the Secretariat in 1975. Only SAW filters were mass-produced at that time. Therefore, SAW filters have been tentatively incorporated in the scope of WG2/Crystal Filters, and the work commenced with preparing SAW filter outline standard, i.e. Part 3. This situation hampered the work in view of differences in manufacturing technology. The real work on them began in 1983 and resulted in the release of IEC 862-3 (1986), Surface acoustic wave filters Part 3: Standard outlines and lead connections.

Part 1, covering Terms and definitions, standard values and test conditions has been printed as IEC 862-1 (1989) and Part 2, Guide to the use - as IEC 862-2 1993). The traditional publication series for SAW resonators is also completed. They are described by IEC 1019-1-1, IEC 1019-1-2, IEC 1019-2 and IEC 1019-3. No IECQ specifications are yet available for SAW filters, or for SAW resonators. Not a single SAW oscillator document is available. This should be a task for the future.

This Working Group has already started preparation of standards for dielectric resonators and filters. Two DIS are underway, both of them describing general information, terms and definitions and standard values. Standard values are withdrawn from the dielectric resonator document.

Test and measuring methods are also under discussion, but they have inferior status of Committee Drafts (CD) (former designation of Secretariat documents). In the opinion of the Secretariat this Working Group on SAW and dielectric devices needs to receive a sub-committee status in the future.

SCIENTIFIC APPROACH

Unlike other technical committees, TC 49 has always used a scientific approach to standards being prepared and problems being considered by it. This approach has become a tradition and has always been supported by both TC 49 Chairmen and the Secretariat.

Although IEC/TC 49 activity has no official liaison with the IEEE Symposium on Frequency Control, the most advanced achievements reported at the Symposia have always been borne in mind or taken into account or reflected in decisions made by TC 49. The spirit of the Symposium has always been present at TC 49 meetings. Moreover, the same highly scientific people have been or are involved both in the scientific work reported at the Symposia and serve on TC 49 Working Groups as Chairmen or experts and bear the major burden of responsibility for technical contents and high quality of IEC publications. This is evident from the list of TC 49 Working Groups, where the names of WG Chairmen are also given.

In order to inform the user of any device where he could find an additional information on the device or method standardized, TC 49 has always placed extensive lists of bibliographies in all Guide documents, containing many references to Proceedings of the IEEE Symposia on Frequency Control. Bibliographies are not recommended for inclusion in TC 49 documents, because they outdate quickly. To avoid this, a reference was made in IEC 122-2, stating that many articles on quartz crystal units have been published in the Proceedings of the Annual Symposium on Frequency Control with the indication of the address where they can be obtained.

In spite of these limitations imposed by the IEC, all TC 49 Guides contain extensive lists of literature (IEC 122-2, IEC 368-2, IEC 642-2, IEC 679-2, IEC 862-2, IEC 1019-2, IEC 758).

The main slogan of the IEC, the essence of which is, "time is enemy of standardization", TC 49 uses in conjunction with its own slogan, "never issue standards not being proven by practice", especially those which could have an impact on the crystal industry. Even by using this approach, TC 49 managed to do the work within time limits prescribed by the IEC and to run smoothly and logically.

Extensive scientific discussions during Main Committee meetings on the future of the crystal industry, perspectives of its growth, on measuring methods etc.have become a tradition in TC 49. Thus, Committee meetings have become a place where the activity and efforts of major manufacturing countries can be coordinated. Many problems have been satisfactorily solved based on the results of these extensive discussions and this is considered to be very important.

An important tool for obtaining agreements is the Round Robin Tests. This method was undertaken by TC 49 to prove certain standard claims or measuring methods in practice. The tools to find out whether or not standards issued by TC 49 have been used or implemented in National Standards are in practice, are Secretariat questionnaires sent to National Committees. The replies on which are considered by TC 49 at its meetings under the item of the agenda "National Standards". TC 49 always insist that TC 49 standards are in wide usage.

By the end of seventies, the crystal industry had difficulties with measurement methods. The absence of an internationally accepted standard hampered trade. There came a pressure from the industry and TC 49 began their consideration. Since that time it has always paid particular attention to measuring methods and problems connected with them. In view of large volume of documents on measuring methods a special Working Group was established in 1968 to consider them. The role of this Working Group in relation to the others is to resolve problems of measuring methods encountered in the latter upon request of any Working Group of TC 49 dealing with devices.

The first proposal on measuring methods came in TC 49 from the Swedish National Committee. This proposal concerned IEEE Standard 177-57, describing a classical resistive π-network. The proposal has been accepted for frequencies below 30 MHz and has been issued within 1 year as IEC 302 (1969). It is still valid and remains the basic terminological document on crystal units. It is under revision now by the U.S. National Committee.

Two years later TC 49 has standardized a zero phase method over the frequency range 1 MHz to 125 MHz. The first Round Robin Experiment was executed in October 1970 in Netherlands to prove the method in practice. A set of 20 crystals covering the frequency range 1-125 MHz has been chosen. For each frequency the sum of temperature coefficients of crystal units was near to zero.They have been measured during 5 days in Netherlands by WG members. After that, crystals have been sent around and the results obtained were compared with those obtained by WG Chairman. The fractional frequency accuracy of the method claimed was ranging between 10^{-6} and 10^{-8} depending on the type of crystal and an accuracy of resonance resistance was $\pm 2\%$. This method was published in IEC 444 (1973).

The Round Robin continued during two years even after the release of the Standard. TC 49 intended to make the method in IEC 444 a basic one for the measurement of resonance frequency and resonance resistance by zero phase technique in π-network.

The third IEC method is a bridge method that has not been accepted by TC 49 as an industrial standard because of its high price. Thus, the method in IEC 444 has been used for about 15 years as the basic one. In the mean time TC 49 tried to fill in gaps and drawbacks in it. Additional publications have been issued in IEC 444-2 describing C_1 - measurements, in IEC 444-3 describing C_0 - compensation, the frequency range of the method has been extended up to 200 MHz, new edition has been released in IEC 444-1 recommending another π-network configuration. Load resonance frequency measurements have been described in the frequency range up to 30 MHz in IEC 444-4. It became necessary again to extend the frequency range of the method, which is now claimed to be 500 MHz, to improve its accuracy, to describe network, to find calibrating resistances. This method is good for measurement of medium grade resonators only and is still used in many countries.

With the advent of new developments, data processing, programmable devices appearance, a strong need appeared to pursue a new approach to measuring methods. This approach accepted by TC 49 involved preparation of:

1) Guide to the measurement of equivalent electrical parameters of quartz crystal units,

2) Separate independent standardization of several measuring methods and various parameters, each describing the composition, construction and use of a particular method of measurement.

This Guide has been prepared and released in IEC 1080 (1991). It describes equivalent network, impedance/admittance characteristics, parameters to be measured, attributes to measurement systems and then the active methods (active or other closed-loop methods, bridge methods, automated network analyzer methods), commercially available measurement systems. Parameter determination with automated S-parameter system, and comparison of characteristics of various measurement systems, with precautions for making precise measurements, and measurement methods in the production environment are included. Classification of quartz crystal unit measurement

systems and tabulation of measuring methods to facilitate the preliminary choice of the method is also given. This Guide has been prepared based on the tutorial of Mr. W.L. Smith at the EIA Piezoelectric Devices Conference held in Kansas City in September 1982.

Further standard to be released by TC 49 concerned methods for the determination of equivalent electrical parameters using automatic network analyzer techniques and error correction. To prove this standard in practice and compare it with network method the experimental approach has been used again.

This International Round Robin Exercise on parameter measurement was well prepared, because the previous Round Robin exercises have shown that sending crystal units around the world was time consuming and a strong possibility exists that crystals will be damaged. A separate WG meeting was held in London in March 1988 to consider details of the experiment. The actual measurement Workshop was held in Bled, Yugoslavia in October 1988.

The purpose of Workshop was to compare the performance of a variety of measurement systems using a suitably chosen set of standard test crystals. The programme of measurement was performed using a variety of combinations of the software systems, crystal test fixtures and test equipments. The following has been used in the experiment:

Software systems:
United Kingdom S-parameter
United States S-parameter
United Kingdom π-network
German π-network (extended)
German π-network (classical)
Yugoslav S-parameter (reflection)

Test equipment:
Hewlett Packard 35 77 A network analyzer
Hewlett Packard 875 3A network analyzer
Rohde and Schwarz ZVP vector analyzer
and SM PD signal generator

A set of 14 crystals for different frequencies submitted by TELEQUARTZ, STC, NDK CEPE and U.S.A. was chosen. The range of the results obtained for each crystal was quite large, typically several p.p.m. for frequency and several percent for resistance and motional capacitance.

The possible causes for measurement discrepancies were suggested and various areas for further investigations were considered. Complete analysis of the measurement data was made by Mr. Peach (United Kingdom). The conclusions have been also presented by Messrs. B. Neubig (Germany), R. J. Williamson (United Kingdom), and Mr. W.L. Smith (U.S.A.) as Working Group Chairman made his Statement on the draft IEC document as follows: "The results of the measurements Workshop held in Bled provided data concerning some remaining problems in the measurement of crystal unit frequencies above about 300 MHz. These problems include questions about the correct mathematical model of a crystal unit to be employed at these high frequencies, and the effects of test fixtures."

The data obtained also demonstrated that amplitude/phase measurements made of crystal units in an unmatched transmission fixture could be corrected for system errors with sufficient accuracy to yield results comparable to those obtained with scattering parameter method.

For these reasons, the document available has been once more revised and submitted to TC 49 for further technical consideration.

After revision of the format and contents of the document in accordance with the format and contents agreed by TC 49 it has been voted and released as IEC 444-5. The methods presented in 444-5 provide state-of-the-art measurements of quartz crystal units and other electroacoustic resonance devices over the frequency range from 1 kHz to 1 GHz or even higher. Within narrow frequency ranges, other IEC Standards are in existence and can be optionally specified. IEC Publications 444-1, 444-2, 444-3, 444-4 are specified for use over the range from 1 MHz to 200 MHz and IEC 302 may be specified at frequencies below 30 MHz.

A very important standard on crystal unit measurements recently published is IEC 444-6, Measurement of drive level dependence (DLD), of quartz crystal units describing two methods. Method A, based on π-network according to IEC 444-1, can be used in the complete frequency range covered by this Publication (200 MHz at present). Method B, an oscillator method, is suitable for measurement of fundamental mode crystals in large quantities with fixed conditions; on maximum R_r during start-up. It can be applied as a 100% final inspection or as a 100% incoming inspection. It can also be used as an instrument to judge if the crystal unit meets the requirements on R_r given in the detail specification.

Both the list of standards available (8 standards and a report dealing with unwanted response measurements) and the approach applied by TC 49 to the measurement method problems are impressive.

A new proposal of the German National Committee on amendment of IEC 444-1 dealing with updating formulae used in the standard is available. The proposal takes into account the modified calibrating procedure of the π-network with a reference resistor $R_r = 25$ ohm instead of a short. Two other proposals are under preparation: on the revision of IEC 302 and measurement of load resonance frequency f_L.

TC 49 activity on materials began after 1968 when the Working Group to deal with synthetic quartz had been established. The first standard covering Specification and Guide to the use of synthetic quartz crystal (IEC 758) was printed in 1983 and a year later

supplemented by a Specification for lumbered synthetic quartz crystal. The quality of the material was estimated in it by a Q-value of resonator at 5 MHz, 5th overtone and frequency vs temperature characteristic. In the second edition of IEC 758 printed in 1993, the IR absorption method of quartz material quality indication is recommended. To prove the method in practice the scientific approach has been used again. An International Round Robin on the extinction coefficient measurement was undertaken and executed in 1990-1992 by Working Group 5. The IRR was in extinction coefficient measurement with three wave numbers 3410, 3500 and 3585cm-1, which is a primary indicator of synthetic quartz.

The quartz samples measured were 26 slices that had come from 11 different quartz producers in 5 different countries. The measurements have been completed on 18 different IR-instruments in 6 different countries. An enormous volume of work has been done by Dr. Sawyer in organization and implementation of the IRR. The interim and final results obtained were reported by Dr. Sawyer at the Annual Symposia on Frequency Control.

The slices used for the experiment are considered as standards and are kept at Sawyer Research Products by Mr. G.R. Johnson, Chairman of Working Group on materials.

The title of the Working Group has been amended in 1993 to "Materials for frequency control and selection". The scope was also enlarged to cover other piezoelectric, dielectric, thin film and piezoelectric ceramic materials used for frequency control and selection applications.

The list of materials to be standardized has been prepared. It is expected that commercially available materials such as lithium niobate and lithium tantalate will be standardized in the near future. Moreover, further refinements of IEC 758 will be needed, e.g. radiation resistance requirements and measurements, special requirements to quartz for SAW application. National proposals are expected on this.

Further field of TC 49 activity is outlines and lead connections of piezoelectric devices used for frequency control and selection. IEC publications covering Part 3 are available for all the devices now within the scope of TC 49 (IEC 122-3, IEC 368-3, IEC 679-3 etc).

They were originally prepared in the pertinent Working Group dealing with devices and considerably differed in their presentation. For this purpose a new Working Group was established in 1984, which has prepared Guidance for the standardization of outline drawings for frequency control and selection devices. This guide was later incorporated in all existing outline standards, preliminary revised in accordance with its requirements.

The work on outlines cannot be considered as a standardization, as such. Existing IEC Publications merely constitute a collection of national experience gained in their manufacturing practice. The rules for submission of outline proposals have been changed several times. At present the questionnaire is circulated with the request to National Committees on the use of all outlines contained in TC 49 Publications in their countries. It is aimed at removing outdated outlines and nonproliferation of types used in one or two countries only. The acceptance criteria will be the use by at least in three countries.

The most important activity on outlines at present is standardization of outlines of surface mounted devices (SMD). A new standard IEC 1240 (1994), Piezoelectric devices - Preparation of outline drawings of surface mounted devices (SMD) for frequency control and selection - General rules, has been released. The experience gained in semiconductor devices was used for its preparing. The Chairman and Secretary of TC 49 are members of IEC Surface Mounting Harmonization Group (SMHG) and materials are available to them from these Committees. In accordance with the information received from the IEC Central Office this standard is included in the plan of CECC for acceptance as European Standard in 1995. Work on preparing TC 49 Standard on outlines of surface mounted devices is progressing. Outline drawings of all devices will be included in a single standard.

One more aspect of TC 49 activity relates to preparation of Chapter 561, Piezoelectric devices for frequency control and selection, of the International Electrotechnical Vocabulary (IEV), the sections of which, at the measure of preparation and discussion in the Working Group, are submitted by the Secretariat of TC 49 to the to the Secretariat of TC 1, Terminology, under the responsibility of TC 1, where linguistical aspect is taken into account. Finalizing these sections is effected in collaboration of TC 1 Secretary, TC 49 Secretary and TC 49/WG 8 Chairman. Five sections dealing with quartz crystal units, their characteristics, piezoelectric filters, crystal oscillators and synthetic quartz material have been printed in 1990. Sections 6 and 7 dealing with SAW filters and characteristics of SAW devices are also completed.

The preparation of the IEV Chapter is a lengthy time-consuming process and terms and definitions might be outdated before they are printed. TC 49 has established WG 11 to prepare a standard for internal use. The purpose of this decision is to have a single source of terminology in TC 49 and to up-date terms and definitions in it quickly as the need arises. Input for this Working Group should be provided by all Chairmen of other Working Groups.

CONCLUSION

As it follows from the foregoing, the complex of problems connected with piezoelectric, dielectric devices and materials to be standardized internationally by TC 49, is very large. The principal attention is paid to IECQ specification preparation, methods of device parameter measurements, piezo-electric crystal and ceramic materials, surface mounting technology. The nearest tasks of TC 49 activity are as follows:

For quartz crystal units:

-revision of IEC 122-2 dealing with the Guide to the use of quartz crystal units. The possibility to use ASFC tutorial for this can be considered.

-co-ordination of IECQ activity within other Working Groups

For piezoelectric filters:

-preparation of IECQ specifications for crystal filters by using capability approval procedure;

-revision of the Guide to the use of crystal filters.

For piezoelectric ceramic resonators & filters:

-completion of the Guide to the use of piezoelectric filters (DIS- stage) revision of IEC 483, Guide to dynamic measurement of piezoelectric ceramics with high electromechanical coupling factor, by the Japanese National Committee

For materials:

-preparation of proposals on new materials (e.g. lithium tantalate, lithium niobate) and further improvements of IEC 758, incorporating new aspects of synthetic quartz application. Proposals from National Committees are invited on this.

For measuring methods:

-revision of IEC 444-1 based on the German proposal available and further extension of its frequency range;

-revision of IEC 302: by the U.S. National Committee

- preparation of a proposal on load resonance frequency measurement of quartz crystal units.

For crystal oscillators:

-progressing with IECQ specification series under both capability and qualification approval procedures;

-revision of IEC 679-2 dealing with the Guide to the use of crystal oscillators.

For piezoelectric devices outlines:

-preparation of the first Committee Draft (a Secretariat document) on SMD devices;

-preparation of a proposal on SMT crystal unit measurements.

For SAW and dielectric devices:

-finalizing Draft International Standards for waveguide type resonators and filters using wave guide type resonators;

-preparation of a Committee Draft describing method of measurement of relative permittivity of materials at microwave frequencies.

For Glossary:

-preparation of a Committee Draft, including terms and definitions describing all aspects of TC 49 activity.

During the period of the work TC 49 has finalized over 80 IEC Standards, 140 Draft International Standards, over 250 Secretariat drafts, of which 24 IEC Standards have been issued during the last two and a half years with the advent of new IEC/ISO Directives.

ACKNOWLEDGMENTS

The author would like to thank Mr. D. C. Bradley for guidance provided to TC 49 and the Secretariat, all WG Chairmen and Messrs. W. L. Smith, J.D. Holmbeck, B. Sawyer, C. Adams (U.S.A.), Messrs. F. Fick, B. Neubig, G. Riha (Germany), Messrs. E. W. Kentley, R. Williamson, S. Morris (United Kingdom), Messrs. M.Onoe, M. Takagi and S. Fujishima (Japan) for their considerable personal contribution to TC 49 activity and support of the Secretariat.

BIBLIOGRAPHY

[1] L.A.Lubimov, L.I.Zhourkina "Piezoelectric devices" IEC Bulletin, vol. IX, No. 31, January 1975

[2] "TC 49 in top gear on the road to faster standards production", IEC Bulletin No.137, Nov./ Dec. 1992

[3] Publication QC 0001002 (1986) - Rules of Procedure of the IEC Quality Assessment System for Electronic Components (IECQ)

1995 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Monofrequency Control Of Low-Frequency Quartz Resonators By Holographic Interferometry

Pyetr Borovtsov

Abstract-**The holographic interferometry method has been used for analyzing spectral characteristics of XYS/+3^0 cut piezoelectric elements of low-frequency quartz resonators. The study of piezoelectric elements of longitudinal vibrators at the resonant frequency f_r=128 kHz has shown that all piezoelectric elements have side resonance's at frequencies of 375 kHz, 585 kHz and 635 kHz. The existence of such resonances can be explained by physical nature of acoustic waves propagation in anisotropic bodies and, namely, by the presence of the angle between the acoustic wave front and the propagation direction. Varying the geometric parameters (thickness, width and length) of piezoelectric elements of quartz resonators, it is impossible to eliminate the side resonances, but their amplitude may by attenuated slightly.**

The holographic time-averaging method has allowed to visualize the process of volumetric acoustic wave propagation and interaction in piezoelectric elements of low-frequency quartz resonators as well as to analyze the effect of piezoelectric elements geometry on monofrequency [1-4].

1. Introduction

The vibrations of different modes and directions are set up in a quartz resonator piezoelectric element due to anisotropy, i.e. the quartz resonator is a system of infinite number of degrees of freedom. Such system may be considered as a monofrequent one only conditionally, so that connection of operating vibration and other vibration is sufficiently slight.

Under excitation of quartz resonators the stress strains of various modes are set up at once due to elastic connections. The main vibrations and their harmonics correspond to those stress strains. Moreover, a system of standing waves of a definite frequency of natural vibrations is produced in piezoelectric elements of finite dimensions.

Thus, apart from the main (operating) frequency, the quartz resonators have an inherence of definite spectrum of resonance frequencies of undesirable, various intensity vibrations expressed by the amplitude-frequency characteristic (AFC) of resonators. The propagation studies have shown that during acoustic wave movement in any anisotropic body an angle other that 0 or 90 exists between the wave front and the wave direction, i.e. the waves are "nearly longitudinal" or "nearly transverse". This means that the elastic waves in piezoelectric elements are longitudinal or transverse only under definite specific orientation.

The method available till now for determining the quartz resonator characteristics gave the highly accurate values of spectrum frequencies and attenuation amplitudes of the main and side resonances with respect to the main resonance amplitude, but those method did not allow to determine the vibration mode precisely. The holographic interferometry method makes it possible to control the behavior of vibration modes of low-frequency piezoelectric elements of quartz resonator. Spectral characteristics of piezoelectric elements of low-frequency quartz resonator have been studied by interferograms. The existence of the same side resonatores in piezoelectric elements of one and the same cut, but of different geometric dimensions can be explained by physical nature of acoustic waves propagation in an anisotropic body [1], [2], [3], [4].

2. Analysis of Acoustic Wave Moment In Anisotropic Body

The equation for wave moment free of attenuation in the anisotropic body may be presented in the tensor form as follows:

$$\rho \frac{\partial^2 U_i}{\partial t^2} = \frac{\partial^2 T_{ij}}{\partial X_i^{\,2}}, \qquad (1)$$

where:

ρ is the density of surroundings (in this case: the quartz density):

T_{ij} - the components of symmetrical, mechanical stress tensor in second order;

0-7803-2500-1/95 $4.00 © 1995 IEEE

U_i - the displacement vector components.

T_{ij} and U_i are the functions of time t and coordinates X_i. The summation of identical indexes in one member is taken in the same may. All indices are numbered as 1,2,3. With the Hook's law T_{ij} may be eliminated from equation (1):

$$T_{ij} = C_{ijkl} \bullet S_{kl}. \qquad (2)$$

The components of symmetrical tensor in second order for mechanical stress is determined by the following expression:

$$S_{kl} = \frac{1}{2}\left(\frac{\partial U_k}{\partial X_l} + \frac{\partial U_l}{\partial X_k}\right). \qquad (3)$$

It may be proved that for the modules of elasticity C_{ijkl} the following is valid:

$$C_{ijkl} = C_{ijlk} = C_{jilk} = C_{klij} = \\ = C_{lkij} = C_{klji} = C_{lkji} \quad . \qquad (4)$$

So only 21 modules out of 81 modules are linearly independent. Substituting aquatints (2) and (3) into equation (1), we obtain /1/:

$$\rho\frac{\partial^2 U_i}{\partial t^2} = \frac{1}{2}C_{ijkl}\frac{\partial^2 U_k}{\partial X_j \partial X_l} + \frac{1}{2}C_{ijkl}\frac{\partial^2 U_l}{\partial X_j \partial X_k} \qquad (5)$$

Both members in the right side of the equation are equal due to the fact that the indicis i, k and l are imaginary in accordance with rule of summation and $C_{ijlk} = Cijkl$. Hence we obtain:

$$\rho\frac{\partial^2 U_i}{\partial t^2} = C_{ijkl}\frac{\partial^2 U_l}{\partial X_j \partial X_k}. \qquad (6)$$

In case of the plan waves propagation depended only on the new coordinate S (the distance between the plane wave surface and the reference point X_j of coordinates) measured in direction of the normal N to the wave surface. Therefore,

$$S = X_i \cdot \cos(X_i, N). \qquad (7)$$

Substituting expression (7) into equation (6)

$$\frac{\partial^2 U_l}{\partial X_j \partial X_k} = \frac{\partial^2 U_l}{\partial S^2} \cdot \cos(X_k, N) \cdot \cos(X_j, N) \qquad (8)$$

we obtain the equation of movement:

$$\rho\frac{\partial^2 U_i}{\partial t^2} = \Gamma_{ij}\frac{\partial^2 U_l}{\partial S^2}, \qquad (9)$$

with

$$\Gamma_{il} = C_{ijkl} \cdot \cos(X_j, N) \cdot \cos(X_k, N) = \Gamma_{li}. \qquad (10)$$

If a_i is the direction of cosines of total displacement $U = a_i \cdot U_i$ then from equation (8) after multiplication by a_i and summation we obtain

$$\rho\frac{\partial^2 U_i}{\partial t^2} = a_i \cdot \Gamma_{il}\frac{\partial^2 U_l}{\partial S^2}. \qquad (11)$$

Chose the E value with the dimension of module of elasticity such that

$$a_i\Gamma_{il} = E \cdot a_l. \qquad (12)$$

Then the equation for the plane acoustic wave is written in the following form:

$$\rho\frac{\partial^2 U}{\partial t^2} = E \cdot \frac{\partial^2 U}{\partial S^2}. \qquad (13)$$

The wave velocity is written in the form of

$$V = \sqrt{\frac{E}{\rho}} \qquad (14)$$

Equation (12) is satisfied if the following relationship takes place:

$$\begin{vmatrix} \Gamma_{11}-E & \Gamma_{21} & \Gamma_{31} \\ \Gamma_{12} & \Gamma_{22}-E & \Gamma_{32} \\ \Gamma_{13} & \Gamma_{23} & \Gamma_{33}-E \end{vmatrix} = 0 \quad (15)$$

Expression (15) has three roots: $E^{(1)}, E^{(2)}, E^{(3)}$ Each root makes a_i to be solved with the use of equation (12) together with the relationship:

$$a_i \cdot a_j = 1. \quad (16)$$

Let's examine the elastic wave propagation in quartz. We consider the wave with the normal N to the wave surface in the plane YZ. As an axial system X_i let's choose the system of axes rotated at the angle θ about the X axis, referred to crystallographic axes, so X_2 has the direction of N. Then

$$\begin{aligned} &\cos(X_1, N) = \cos(X_3, N) = 0 \\ &\cos(X_2, N) = 1 \end{aligned} \quad (17)$$

Therefore, the members Γ_{il} of equation (18)

$$\Gamma_{il} = C_{i221} \quad (18)$$

are reduced to

$$\begin{aligned} &\Gamma_{11} = C_{66}, \Gamma_{22} = C_{22}, \Gamma_{23} = C_{24} \\ &\Gamma_{33} = C_{44}, \Gamma_{12} = C_{13} = 0 \end{aligned} \quad (19)$$

Substituting the values of (19) into equation (15), we obtain:

$$E^{(1)} = C_{66}, \quad (20)$$

$$E^{(22)} = \frac{1}{2}(C_{22} + C_{44} + \sqrt{(C_{22} - C_{44})^2 + 4 \cdot C_{24}^2}) \quad (21)$$

$$E^{(22)} = \frac{1}{2}(C_{22} + C_{44} - \sqrt{(C_{22} - C_{44})^2 + 4 \cdot C_{24}^2}) \quad (22)$$

The solutions of equations (20), (21) and (22) substituted into equation (12) together with expression (16) reduce to the following cases:

1) $E = E^{(1)}$ which gives

$$a_1 = 1, a_2 = a_3 = 0 \quad (23)$$

The displacement U takes place towards the direction of the axis $X_1 \perp N$, the wave is purely transverse;

2) $E = E^{(2)}$. Now displacement U takes place in the plane $X_2 X_3$, as

$$a_1 = 0 \quad (24)$$

Moreover,

$$\frac{a_2}{a_3} = \frac{E^{(2)} - C_{44}}{C_{24}} = \frac{C_{22} - C_{44} + \sqrt{(C_{22} - C_{44})^2 + 4 \cdot C_{24}^2}}{2 \cdot C_{24}} \quad (25)$$

$$a_3^2 + a_2^2 = 1$$

As $a_i = \cos(U, X_i)$, from equation (25) we find

$$ctg(U, N) = Ctg(U, X_2) = \frac{C_{22} - C_{44} + \sqrt{(C_{22} - C_{44})^2 + 4 \cdot C_{24}^2}}{2 \cdot C_{24}} \quad (26)$$

3) $E = E^{(3)}$, in this case a_1 is equal to zero as well, but

$$ctg(U,N) = \frac{C_{22} - C_{44} - \sqrt{(C_{22} - C_{44})^2 + 4 \cdot C_{24}^2}}{2 \cdot C_{24}} \quad (27)$$

From equation (25), (26) and (27) it follows that for the value of the displacement U those tree relationships are orthogonal relatively to one another. In case 2 and 3 the wave is not a pure compression wave or transverse wave. So, a wave at which the angle between U and N is small will be called a "Nearly Longitudinal Wave"(NLW), in the other case it will be called a "Nearly Transverse Wave"(NTW).

3. Study Of Spectral Characteristic Of Low-Frequency Quartz Resonator Piezoelectric Elements By Interferograms

The methods available till now for determining the quartz resonator characteristics gave the highly accurate value of spectrum frequencies and attenuation amplitudes of the main and side resonances with respect to the main resonance amplitude, but those methods did not allow to determine the vibration mode precisely. Amplitude-frequency characteristics of longitudinal vibrations of X-cut quartz plates have been studied. These vibration are more complicated then isotropic ones, since the compressive-shearings are interconnected. In X-cut plate tension not only the plate elongation and transverse compression take place, but the displacement in the plate plane occurs as well. This connection being of elastic character makes the vibration mode more complicated and affects the frequency and its temperature-dependence.

With $b/l = 0.23$ the anomaly typical for connected vibrations is observed. It is explained by the fact that the frequency of longitudinal vibrations in the first order coincides with the frequency of bending vibrations in the width of the second order and the side resonance is revealed nearby. This eliminates the possibility of using the plates such dimension-relation. The longitudinal vibrations may interact with the even-numbered bending vibrations in the higher orders, and with their frequencies close with one another the anomalies being the same as those set forth are observed.

The undesirable connection with the vibrations of displacement disappears when the plate is rotated around the X-axis at the angle of -18.5^0 . With such orientation the plate vibrations are purely longitudinal. The amplitude-frequency characteristics (AFC) of quartz resonators with the XYS/+3^0 cut piezoelectric elements with the longitudinal vibration mode the resonant frequency of 128 kHz are show on Fig. 1 and Fig. 2

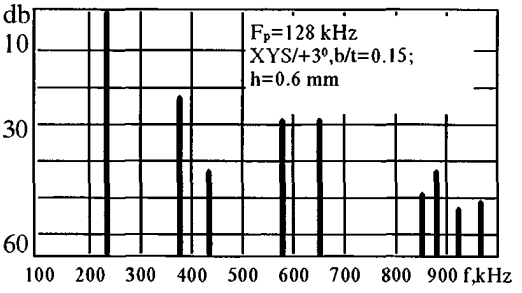


Fig. 1

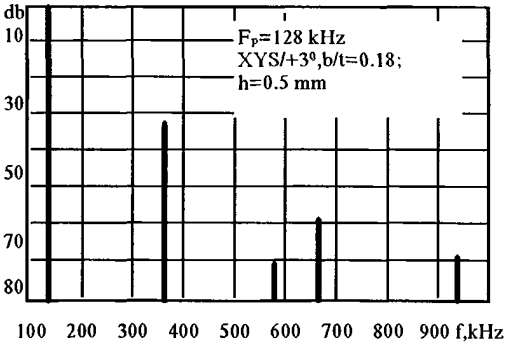


Fig. 2

The piezoelectric elements of these resonators have different geometric parameters: width, length and thickness. But all piezoelectric elements have the width-length relation

$b/l < 0.23$, i.e. this relation is under the critical one with which the frequency of longitudinal vibrations in the first order coincides with the frequency of bending vibrations in the width of the second order.

The XYS/+3^0 cut piezoelectric elements with the longitudinal vibration mode at the resonant frequency of 128 kHz excited in the side resonant frequency rage up to 635 kHz are shown in the interferograms obtained by the holographic time-averaging method. (The side resonance's are of 375 kHz and 635 kHz). The existence of such resonances can be explained by the physical nature of volumetric acoustic waves propagation in anisotropic surroundings and, namely, by the presence of the angle between the volumetric acoustic wave front and the propagation direction. This angle is seen in the interferograms as the slope of interferentinal lines. Varying the geometric parameters (thickness, width and length) of piezoelectric elements of quartz resonators, it is impossible to eliminate the presence of these side resonance's, but the amplitude of the side vibration modes be attenuated slightly. The spectral characteristic of low-frequency quartz resonator can be improved by the use of XYS/+18.5^0 cut piezoelectric elements. There is no slope of interferential lines in the interferogram of XYS/-18.5^0 cut piezoelectric elements, so they have the "pure" mode of longitudinal vibrations and, in consequence, their amplitude-frequency characteristic is free of side vibrations.

The holographic time-averaging method has allowed to visualize the process of volumetric wave propagation in piezoelectric elements of low-frequency quarts resonators.

The interferograms show not only the character of volumetric acoustic wave propagation, but the interaction of different vibration modes as well. With the interferogram in can be seen that in the width-length plane the bending vibration mode is superimposed on the longitudinal vibration mode and the effect of this superimposition will depend on geometric parameters which is visible from the interferenial pattern change.

Thus, the holographic time-averaging method allows to determine precisely the vibration mode, the amplitude value, the vibration zone and piezoelectric element optimum geometric dimensions.

References

[1] Van der Veen, "One the Angle Between Wave Front And Displacement of Plane Acoustic Waves in Quarts", Philips Research v. 12, pp.273-280, 1957.

[2] C.M.Vest, D.W.Sweetney, "Measurement of Vibrational Amplitude by Modulation of Projected Fringes", Applied Optics, vol.11, pp.449-454, 1972.

[3] H.J. Tiziany, J. Klenk, K. Leonhardt, "Holography and Speckle Techniques For Real-time Displacement Deformations And Vibration Analysis", Proc. Soc. Phoyo-Opt. Instrum. Eng., vol. 236, pp. 62-67, 1980.

[4] П.В. Боровцов, В.Б. Грузиенко, С.Ф. Михеев, "Применение голографической интерферометрии для исследования колебаний пьезоэлементов низкочастотных кварцевых резонаторов", Электронная техника, Сер. 5, вып. 5, с. 116-119, 1975

IEEE Frequency Control Symposium+
PROCEEDINGS ORDERING INFORMATION

NO.	YEAR	DOCUMENT NO.	OBTAIN FROM*	PRICE MICROFICHE / HARDCOPY
10	1956	AD-298322	NTIS	$9.00 /$61.00
11	1957	AD-298323	NTIS	9.00 / 77.00
12	1958	AD-298324	NTIS	9.00 / 77.00
13	1959	AD-298325	NTIS	9.00 / 84.00
14	1960	AD-246500	NTIS	9.00 / 52.00
15	1961	AD-265455	NTIS	9.00 / 44.00
16	1962	AD-285086	NTIS	9.00 / 52.00
17	1963	AD-423381	NTIS	9.00 / 77.00
18	1964	AD-450341	NTIS	9.00 / 77.00
19	1965	AD-471229	NTIS	9.00 / 77.00
20	1966	AD-800523	NTIS	9.00 / 77.00
21	1967	AD-659792	NTIS	9.00 / 61.00
22	1968	AD-844911	NTIS	9.00 / 77.00
23	1969	AD-746209	NTIS	9.00 / 44.50
24	1970	AD-746210	NTIS	9.00 / 44.50
25	1971	AD-746211	NTIS	9.00 / 44.50
26	1972	AD-771043	NTIS	9.00 / 44.50
27	1973	AD-771042	NTIS	9.00 / 52.00
28	1974	AD-A011113	NTIS	9.00 / 52.00
29	1975	AD-A017466	NTIS	9.00 / 52.00
30	1976	AD-A046089	NTIS	9.00 / 61.00
31	1977	AD-A088221	NTIS	9.00 / 77.00
32	1978	AD-A955718	NTIS	19.50 / 61.00
33	1979	AD-A213544	NTIS	19.50 / 84.00
34	1980	AD-A213670	NTIS	17.50 / 61.00
35	1981	AD-A110870	NTIS	20.50 / 84.00
36	1982	AD-A130811	NTIS	9.00 / 61.00
37	1983	AD-A136673	NTIS	9.00 / 52.00
38	1984	AD-A217381	NTIS	12.50 / 44.50
39	1985	AD-A217404	NTIS	19.00 / 61.00
40	1986	AD-A235435	NTIS	17.50 / 52.00
41	1987	AD-A216858	NTIS	17.50 / 61.00
42	1988	AD-A217275	NTIS	17.50 / 61.00
43	1989	AD-A235629	NTIS	19.50 / 77.00
44	1990	90CH2818-3	IEEE	90.00 / 90.00
45	1991	91CH2965-2	IEEE	110.00 / 110.00
46	1992	92CH3083-3	IEEE	120.00 / 120.00
47	1993	93CH3244-1	IEEE	120.00 / 120.00
48	1994	94CH3446-2	IEEE	120.00 / 120.00
49	1995	95CH3575-2	IEEE	120.00 / 120.00 (tentative)

*NTIS - National Technical Information Service
5285 Port Royal Road, Sills Building
Springfield, VA 22161, U.S.A.

*IEEE - Inst. of Electrical & Electronics Engineers
445 Hoes Lane
Piscataway, NJ 08854, U.S.A.
Tel: 800-678-4333 or 908-981-0060

+ Prior to 1992, the Symposium's name was the "Annual Symposium on Frequency Control," and "International" was not yet in the name in 1992.

SPECIFICATIONS AND STANDARDS RELATING TO FREQUENCY CONTROL

INSTITUTE OF ELECTRICAL & ELECTRONIC ENGINEERS (IEEE)

Order from: IEEE Service Center
445 Hoes Lane
Piscataway, NJ 08854
Telephone: (908) 981-0060

176-1987 (ANSI/IEEE) Standard on Piezoelectricity (SH11270)

177-1966 Standard Definitions & Methods of Measurements for Piezoelectric Vibrators

180-1986 (ANSI/IEEE) Definitions of Primary Ferroelectric Crystal Terms (SH10553)

319-1971 (Reaff 1978) Piezomagnetic Nomenclature (SH02360)

1139-1988 Standard Definitions of Physical Quantities for Fundamental Frequency & Time Metrology (SH12526)

DEPARTMENT OF DEFENSE

Order from: Naval Pubs & Form Center
5801 Tabor Avenue
Philadelphia, PA 19120
Telephone: (215) 697-2000

MIL-C-3098 Crystal Unit, Quartz, Gen Spec for

MIL-C-24523 (SHIPS), Chronometer, Quartz Crystal

MIL-F-15733 Filters & Capacitors, Radio Interference, Gen Spec for

MIL-F-18327 Filters, High Pass, Band Pass Suppression and Dual Processing, Gen Spec for

MIL-F-28861 Filters and Capacitors, Radio Frequency Electro-magnetic Interference Suppression, Gen Spec for

MIL-F-28811 Frequency Standard, Cesium Beam Tube

MIL-H-10056 Holders (Encl), Crystal, Gen Spec for

MIL-O-55310 Oscillators, Crystal, Gen Spec for

MIL-O-39021 Oven, Crystal, Gen Spec for

MIL-S-49433 Surface Acoustic Wave Devices, Bandpass Filter

MIL-S-49433(ER) Surface Acoustic Wave Devices, Gen Spec for

MIL-STD-683Crystal Units, Quartz/Holders, Crystal

MIL-STD-188-115 Interoperability & Performance Standards for Communications, Timing & Synchronization Subsystems

MIL-STD-1395 Filters & Networks, Selection & Use of

MIL-T-28816(EC) Time Frequency Standard, Disciplined, AN/URQ-23, Gen Spec for

MIL-W-46374D Watch wrist: General Purpose

MIL-W-87967 Watch wrist: Digital

GENERAL SERVICES ADMINISTRATION

Order from: Naval Publication & Form Center **or** General Services Administration Business Service Centers in major U.S. cities.

FED-STD-1002 - Time & Frequency Reference Information in Telecommunication Systems

ELECTRONIC INDUSTRIES ASSOCIATION

Order from: Electronic Industries Assoc.
2001 Eye Street, NW
Washington, DC 20006
Telephone: (202) 457-4900

(a) Holders and Sockets

EIA-192-A, Holder Outlines and Pin Connections for Quartz Crystal Units (Standard Dimensions for Holder Types).

EIA-367, Dimensional & Electrical Characteristics Defining Receiver Type Sockets (including crystal sockets).

EIA-417, Crystal Outlines (Standard dimensions and pin connections for current quartz crystal units-1974).

(b) Production Tests

EIA-186-E, (All Sections), Standard Test Methods for Electronic Component Parts.

EIA-512, Standard Methods for Measurement of Equivalent Electrical Parameters of Quartz Crystal Units, 1 kHz to 1 GHz, 1985.

EIA/IS-17-A, Assessment of Outgoing Nonconforming Levels in Parts Per Million (PPM).

EIA-IS-18, Lot Acceptance Procedure for Verifying Compliance with the Specified Quality Level (SQL) in PPM.

(c) Application Information

EIA Components Bulletin No. CB6-A, Guide for the Use of Quartz Crystal Units for Frequency Control, Oct 1987.

(d) EIA-477, Cultured Quartz (Apr. 81)

EIA-477-1, Quartz Crystal Test Methods (May 1985)

INTERNATIONAL ELECTRO-TECHNICAL COMMISSION (IEC)

Order from: American Nat'l. Standard Inst. (ANSI)
1430 Broadway
New York, NY 10018
Telephone: (212) 354-3300

IEC PUBLICATIONS PREPARED BY TC 49

122: Quartz crystal units for frequency control and selection

122-2 (1983) Part 2: Guide to the use of quartz crystal units for frequency control and selection

122-2-1 (1991) Section One: Quartz crystal units for microprocessor clock supply. Amendment 1 (1993)

122-3 (1977) Part 3: Standard outlines and pin connections. Amendment No. 2 (1991), Amendment No. 3 (1992), Amendment No. 4 (1993)

283 (1986) Methods for the measurement of frequency & equivalent resistance of unwanted resonances of filter crystal units

302 (1969) Standard definitions and methods of measurement for piezoelectric vibrators operating over the frequency range up to 30 MHz

314 (1970) Temperature control devices for quartz crystal units. Amendment No. 1 (1979)

314A (1971) First supplement

368: Piezoelectric Filters

368-1 (1992) Part 1: General information, standard values and test conditions

368-2 (1973) Part 2: Guide to the use of piezoelectric filters

368-2-1 (1988) Section One - Quartz crystal filters

368B (1975) Second supplement
368-3 (1991) Part 3: Standard outlines

444: Measurement of quartz crystal unit parameters

444-1 (1986) Part 1: Basic method for the measurement of resonance frequency and resonance resistance of quartz crystal units by zero phase technique in a π-network.

444-2 (1980) Part 2: Phase offset method for measurement of motional capacitance of quartz crystal units.

444-3 (1986) Part 3: Basic method for the measurement of two-terminal parameters of quartz crystal units up to 200 MHz by phase technique in a π-network with compensation of the parallel capacitance C_o.

444-4 (1988) Part 4: Method for the measurement of the load resonance frequency f_L, load resonance resistance R_L and the calculation of other derived values of quartz crystal units, up to 30 MHz.

483 (1976) Guide to dynamic measurements of piezoelectric ceramics with high electromechanical coupling.

642 (1979) Piezoelectric ceramic resonators and resonator units for frequency control and selection. Chapter I: Standard values and conditions. Chapter II: Measuring and test conditions.

642-2 (1994) Part 2: Guide to the use of piezoelectric ceramic resonator units.

642-3 (1992) Part 3: Standard outlines.

679: Quartz crystal controlled oscillators.

679-1 (1980) Part 1: General information, test conditions & methods. Amendment No. 1 (1985).

679-2 (1981) Part 2: Guide to the use of quartz crystal controlled oscillators.

679-3 (1989) Part 3: Standard outlines and lead connections. First supplement (1991). Amendment 1 (1994).

689 (1980) Measurements and test methods for 32 kHz quartz crystal units for wrist watches and standard values.

758 (1993) Synthetic quartz crystal - specifications and guide to the use.

862: Surface acoustic wave (SAW) filters.

862-1 (1989) Part 1: General Information, standard values and test conditions. Chapter I: General information and standard values - Chapter II: Test conditions.

862-2 (1991) Part 2: Guide to the use of surface acoustic wave filters (Chapter III).

862-3 (1986) Part 3: Standard outlines (Chapter IV).

1019: Surface acoustic wave (SAW) resonators.

1019-1-1 (1990) Part 1: General information, standard values and test conditions. Section 1 - General information and standard values.

1019-1-2 (1993) Section 2: Test conditions.

1019-3 (1991) Part 3: Standard outlines and lead connections.

1080 (1991) Guide to the measurement of equivalent electrical parameters of quartz crystal units.

1178-1 (1993) Quartz crystal units - A specification in the IEC Quality Assessment System for Electronic Components (IECQ) Part 1: Generic specification.

1178-2 (1993) Part 2: Sectional specification - Capability approval.

1178-2-1 (1993) Part 2: Sectional specification - Capability approval. Section 1: Blank detail specification.

1178-3 (1993) Part 3: Sectional specification - Qualification approval.

1178-3-1 (1993) Part 3: Sectional specification - Qualification approval - Section 1: Blank detail specification.

1240 (1994) Piezoelectric devices - Preparation of outline drawings of surface-mounted devices (MSD) for frequency control and selection. General rules.

1253: Piezoelectric ceramic resonators - A specification in the IEC quality assessment system for electronic components (IECQ).

1253-1 (1993) Part 1: Generic specification - Qualification approval.

1253-2 (1993) Part 2: Sectional specification - Qualification approval.

1253-2-1 (1993) Section 1: Blank detail specification - Assessment level E.

1261: Piezoelectric ceramic filters for use in electronic equipment - A specification in the IEC quality assessment system for electronic components (IECQ).

1261-1 (1994) Part 1: Generic specification - Qualification approval.

1261-2 (1994) Part 2: Sectional specification - Qualification approval.

1261-2-1 (1994) Part 2: Section 1: Blank detail specification - Assessment level E.

AUTHOR INDEX

Abbott, B.P. 452
Abdrafikov, S.N. 685
Abramson, I.V. 838
Adachi, T. 565
Agnieray, P. 122
Alhayek, I. 832
Aliev, V.S. 459
Almar, R.C. 476,486
Aman, J. 185
Anderson, C.L. 519
Ando, M. 679
Atsumi, K. 33
Audoin, C. 178
Auriol, A. 122
Avramov, I.D. 459

Bagayev, V.P. 574
Bagby, J.S. 519
Bahoura, M. 60
Ballandras, S. 469
Ballato, A. 486,499,507,794
Barillet, R. 178
Barrett, R.L. 519
Bastiaans, G.J. 876
Bauer, N.M. 685
Beard, R.L. 133
Behagi, A.A. 427
Bergquist, J.C. 110
Berkeland, D.J. 110
Besson, R.J. 590
Bezdelkin, V.V. 657
Bigler, E. 469
Blair, D.G. 433
Boggs, C.K. 367
Borovtsov, P. 903
Boy, J-J. 590
Brendel, R. 10
Brosig, J.A. 885
Broughton, J. 870
Brunet, M. 122
Buisson, J.A. 133
Burns, S. 876,879
Busca, G. 53

Cambril, E. 469
Cantor, S.G. 3
Cantor, S.R. 3,20
Capelle, B. 629,716
Carosella, C. 848
Cavin, M.S. 476
Celikov, A. 153
Chang, K-W. 190
Cherpoukhina, G.N. 657
Chiba, K. 33
Christophorova, T.V. 810
Clairon, A. 60
Cochet-Muchy, D. 629
Coles, C. 66
Couplet, C. 53
Cruz, F.C. 110
Cutler, L.S. 86

D'Evelyn, L. 153
Daniau, W. 623
Denissenko, S. 459,469
Detaint, J. 629,716
DeYoung, J.A. 86
Dick, G.J. 82,175,212,397
Doak, A.D. 367
Dow, G.S. 190
Dresin, D.I. 657
Driscoll, M.M. 401,514
Drullinger, R.E. 113
Dubovik, M.F. 638
Dulmet, B. 557

Emmons, D.A. 600
Evenson, K.M. 205
Everard, J.K.A. 374

Fagg, H. 118
Fan, S-J. 667
Ferre-Pikal, E.S. 294,305
Filler, R.L. 852
Fisk, P.T.H. 66
Fox, R.W. 185

Galliou, S. 557
Gamidov, R. 149
Gankin, V.Z. 140
Garvey, R.M. 600
Gavignet, E. 469
Geier, G.J. 266
Gevorkyan, A.G. 140
Ghezali, S. 60
Giffard, R.P. 86
Girardet, E. 623
Gjelsvik, A.M. 20
Goli, J. 794
Gotalskaja, A.N. 657
Gouzhva, Yu.G. 140
Greenhall, C.A. 346
Grimm, M.J. 212
Grossman, E.N. 205

Hackman, C. 275
Hamaguchi, K. 679
Hamell, R.L. 212
Hartnett, R.J. 259
Hashimoto, K-Y. 442,452
He, J.W. 118
Helmcke, J. 153
Hickernell, F.S. 530
Hickernell, T.S. 530
Hivert, B. 10
Hollberg, L. 185
Hou, J. 755
Howald, R. 250
Howe, D.A. 321,337
Huynh, D. 494

Ioakimidis, T.E. 20
Ionov, B.P. 574
Ishii, O. 818
Ishizaki, A. 746,761
Itano, W.M. 110
Ivanov, E.N. 314,420

Jefferts, S.R. 275,294

Kalia, R. 870
Kant, R.A. 848
Karlquist, R. 217
Katrunov, K.A. 638
Kawashima, H. 778,798
Kennedy, H.L. 266
Kent, G.W. 207
Kersten, P. 153
Kikuta, S. 679
Kim, Y. 852
King, T.M. 266
Kline, G.R. 827
Knuth, H.D. 530
Kochemasov, V.N. 236
Kornishov, O.N. 140
Korshikova, T.I. 638
Kosinski, J.A. 486,494,499,507
Kosykh, A.V. 542
Koyama, M. 620
Koyama, Y. 33
Kozlov, A.S. 459
Kubik, A.J. 86
Kuleshov, V.N. 282,288

Laffey, S. 696
Lakin, K.M. 827
Lantz, E. 330
Lareau, R. 696
Laurent, Ph. 60
Lawn, M.A. 66
Lea, S.N. 60
Lee, P.C.Y. 705,769
Lee, W.D. 113
Leshukov, B.E. 282
Lider, V. 642
Lin, C.F. 118
Liu, G. 876,879
Liu, H.Y. 282,288
Lo, D.C-W. 190
Luiten, A.N. 433

Mackie, N. 185
Maleki, L. 74,79,82,161,169
Mamaev, N.A. 685
Mang, L. 530
Mann, A.G. 433
Marguier, J. 623
Marianneau, G. 623
Marquardt, J.H. 185
Marra, T. 250
Martin, D.W. 600
Matsakis, D.N. 86
Matsuura, H. 33
McCarron, K.T. 827
McCaskill, T.B. 133
McClelland, T. 39
McDonald, N.J. 433
McGowan, R. 494,499
McHugh, M. 330
McNamara, B.R. 266
Medvedev, A.V. 642,647
Meli, C. 870
Mileti, G. 53
Mill, B.V. 653
Miller, J.D. 110
Miller, L. 259
Morita, T. 818
Mourey, M.M. 557,590
Mulford, C.D. 494
Murray, R.A. 696
Myasnikov, A.L. 140

Nagai, K. 679
Nakajima, Y. 33
Nakazawa, Y. 818
Naumenko, N.F. 671
Nelson, C.W. 305
Norton, J.R. 614
Nouel, F. 122

Oaks, O.J. 133
Odagawa, H. 537
Oomura, Y. 620
Ottman, G. 259

Page-Jones, M.A. 374
Parker, T.E. 275
Parzen, B. 569
Pascaru, I. 39
Pavasovic, A. 548
Pendrill, L. 185
Peterson, B. 259
Pierce, D.E. 696
Pierkarz, R. 494
Pisarevsky, Yu. 642,647,653
Pogrebnyak, A.P. 685
Poitzsch, M.E. 110
Popov, P.A. 653
Popovic, J. 548
Porter, M.D. 876,879
Postnikov, I.I. 804,810
Prestage, J.D. 79,82,169

Reid, W.G. 133
Riehle, F. 153
Riley, W.J. 360
Robinson, H.G. 185
Rochat, P. 53
Rohde, U.L. 379
Rokos, H. 546

Saito, T. 818
Sakharov, S.A. 642,647
Santarelli, G. 60
Santiago, D.G. 397
Sautenkov, V. 149
Saveleva, N.I. 657
Schegolkova, S.N. 657
Schweda, H. 53
Seidel, D.J. 74
Sekimoto, H. 620,746,761
Sellars, M.J. 66
Senushencov, P.A. 642,647,653
Sheng, N-H. 207
Shirley, J.H. 113
Shmaly, Yu.S. 579
Shtaerman, I. 39
Shulgin, B.V. 685
Simon, E. 60
Smagin, A.G. 843
Smits, G.J. 794
Stephens, M. 185
Stewart, J.T. 486,494,499,507
Stone, C. 39
Stowers, D.A. 175,212
Sunaga, K. 798
Suzuki, C.K. 679
Sydnor, R.L. 212
Szekely, C. 39
Szymaniec, K. 60

Tajima, D. 746
Takeuchi, M. 537
Taki, S. 679
Tanaka, M. 537
Taskin, I. 149
Taslakov, M.A. 438
Teplova, S.D. 140
Thielen, C. 879
Thomann, P. 53
Thomas, R.D. 266
Tiersten, H.F. 727,740
Tjoelker, R.J. 79,82,169
Tobar, M.E. 314,420
Tsarapkin, D.P. 413
Tsuzuki, Y. 565
Turner, S.D. 427
Ulitko, I.A. 786
Ungvichian, V. 519

Vale, L.R. 205
VanBaak, D.A. 185
Vashishta, P. 870
Vasiljevic, D. 548
Velichansky, V. 153
Venot, D. 178
Vernotte, F. 330
Vig, J.R. 696,852
Voissem, J. 832

Walls, F.L. 113,294,305,367
Waltman, S. 185
Wang, H. 190
Wang, J. 667,705
Wang, R.T. 397
Warren, H.E. 133
Watanabe, Y. 620,746
Weimer, C.S. 185
Weinert, R.W. 401
Wilson, W.E. 118
Wineland, D.J. 110
Woode, R.A. 314,420
Wyse, R. 879

Xia, J. 879
Xu, Y-B. 667
Xuan, Z. 354
Xue, T. 879

Yamaguchi, M. 442
Yamanouchi, K. 537
Yang, J.S. 727
Yao, S. 161
Yoda, Y. 679
Yong, Y-K. 755
Yu, J. 354
Yu, J.D. 769

Zacharski, J. 39
Zalamansky, G. 330
Zarka, A. 629,716
Zavyalov, S.A. 542,574
Zhai, Z.C. 118
Zhang, J.W. 565
Zhang, X.W. 679
Zhang, Z. 755
Zharii, O.Yu. 750
Zharov, A.N. 236
Zheng, Y. 716
Zhou, W. 354
Zhourkina, L.I. 894
Zibrov, A.S. 153,185
Zibrova, T.P. 185
Zink, L.R. 205
Zinner, G. 153
Zivkovic-Dzunja, Z. 548