Proceedings of the

1994 IEEE International
Frequency Control Symposium

(The 48th 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. 94CH3446-2

Library of Congress No. 87-654207

1-3 June 1994

Westin Hotel - Copley Place
Boston, Massachusetts, 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, 27 Congress Street, Salem, MA 01970. Instructors are permitted to photocopy isolated articles for non-commercial classroom use without fee. 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 ©1994 by the Institute of Electrical and Electronics Engineers, Inc.

IEEE Catalog Number: 94CH3446-2
ISBN 0-7803-1945-1 (Softbound)
ISBN 0-7803-1946-X (Casebound)
ISBN 0-7803-1947-8 (Microfiche)
ISSN 1075-6787
Library of Congress Number: 87-654207

1994 IEEE International Frequency Control Symposium

SYMPOSIUM CHAIRMEN

General Chairman.....Gary R. Johnson

Technical Program Chairman.....Lute Maleki

Finance Chairman.....Raymond L. Filler

Editorial and Awards Chairman.....John R. Vig

Tutorials Chairman.....David W. Allan

Local Arrangements Chairman.....R. Michael Garvey

TECHNICAL PROGRAM COMMITTEE VICE CHAIRMEN

Materials & Resonators.....Robert Smythe
Oscillators & Circuits.....Michael M. Driscoll
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 Electronic Systems
Errol EerNisse, Quartztronics, Inc.
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
Robert Kinsman, Motorola, Inc.
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
Gary Montress, Raytheon Research Division
Thomas Parker, Raytheon Research Division
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
Werner Weidemann, Ball Efratom Division
Joseph White, Naval Research Laboratory
Gernot Winkler, U.S. Naval Observatory
Lidia Zhourkina, Russian Committee IEC, Russia

TECHNICAL SESSION CHAIRMEN

PLENARY SESSION
Lute Maleki, Jet Propulsion Laboratory

PIEZOELECTRIC MATERIALS I: MATERIALS EVALUATION
John R. Vig, U.S. Army Research Laboratory

DIELECTRIC RESONATORS I: CRYOGENIC
G. John Dick, Jet Propulsion Laboratory

PIEZOELECTRIC MATERIALS II: CLASS 32 MATERIALS
Gary R. Johnson, Sawyer Research Products, Inc.

DIELECTRIC RESONATORS II
Fred L. Walls, NIST

SAW OSCILLATORS
Gary K. Montress, Raytheon Research Division

RADAR AND FREQUENCY SYNTHESIZERS
Michael M. Driscoll, Westinghouse Electronics Systems Group

THEORY AND DESIGN OF RESONATORS
John Kosinski, U.S. Army Research Laboratory

ADVANCED TECHNIQUES FOR ATOMIC FREQUENCY STANDARDS
Andrea DeMarchi, Politecnico di Torino, Italy

PROCESSING METHODS, TECHNIQUES AND PROCEDURES
Jack Kusters, Hewlett-Packard Company

MASERS AND TRAPPED ION FREQUENCY STANDARDS
John D. Prestage, Jet Propulsion Laboratory

SENSORS
Errol P. EerNisse, Quartztronics, Inc.

DILITHIUM TETRABORATE
Tadashi Shiosaki, Kyoto University, Japan

RESONATOR PROCESSING: NOVEL TECHNIQUES
K. Yamanouchi, Tohoku University, Japan

CRYSTAL OSCILLATORS
Robert G. Kinsman, Motorola, Inc.

POSTER SESSION
Resonators: *Marvin E. Frerking, Rockwell International/Collins*
Signal Generation & Frequency Sources: *Daniel S. Stevens, AT&T Bell Laboratories*
Filters: *Robert C. Smythe, Piezo Technology, Inc.*
Time Transfer & Atomic Frequency Standards: *J.S. Boulanger, NRC, Canada*
Quartz Resonators & Oscillators: *Gary R. Johnson, Sawyer Research Products*

SAW DEVICES AND MEASUREMENTS
Donald C. Malocha, University of Central Florida

SPACE BASED TIMING SYSTEMS
R. Michael Garvey, Frequency & Time Systems, Inc.

1/f NOISE AND AGING
Thomas E. Parker, Raytheon Research Division

ADVANCED TECHNIQUES FOR FREQUENCY CONTROL DISTRIBUTION & CHARACTERIZATION
Victor S. Reinhardt, Hughes Aircraft Company

TUTORIAL TOPICS & PRESENTERS*

"Introduction to High Stability SAW Oscillators: Design and Performance",
Dr. Thomas E. Parker and Dr. Gary K. Montress, Raytheon Research Division

"Introduction to Quartz Frequency Standards",
Dr. John R. Vig, U.S. Army Research Laboratory

"Introduction to Atomic and Molecular Frequency Standards",
Dr. Joseph White, Naval Research Laboratory

"Oscillator Design Techniques Using Calculated and Measured S Parameters",
Professor Robert J. Weber, Iowa State University

"Introduction to Telecommunications Timing Tools and Techniques",
Mr. George Zampetti, Telecom Solutions

"Introduction to Robust Statistics and Data Filtering",
Dr. Gernot M.R. Winkler, U.S. Naval Observatory

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

"Introduction to GPS as a Time and Frequency Tool",
Mr. David W. Allan, Allan's TIME

"Introduction to S-Parameter Measurements and Calibration Requirements of Quartz Crystals",
Mr. John A. Kusters and Mr. Charles A. Adams, Hewlett-Packard Company

"Clock Ensembles and Timing in Computer Networks",
Dr. Judah Levine, National Institute of Standards and Technology

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

"Introduction to Sapphire Microwave Resonators",
Dr. G. John Dick, Jet Propulsion Laboratory

---

* The Tutorial Session was held the day before the Symposium. 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

PLENARY SESSION PAPERS

Applications of Silicon Micromachining to Resonator Fabrication *(INVITED)* 2
R.T. Howe, University of California at Berkeley

1/f Noise Universality in High-Technology Applications *(INVITED)* 8
P.H. Handel, University of Missouri - St. Louis

PIEZOELECTRIC MATERIALS

Piezoelectric Resonator Materials *(INVITED)* 22
A. Ballato & J.G. Gualtieri, U.S. Army Research Laboratory

Measurements of Acoustic Wave Attenuation in $La_3Ga_5Si0_{14}$ Using HBAR Technique *(INVITED)* 35
G.D. Mansfeld, Institute of Radioengineering and Electronics RAS, Russia

Stress-Induced Optical Activity in Piezoelectric Crystals and Internal Stresses Method of Control in Quartz, Langasite, Lithium Niobate Crystals 40
V.S. Naumov, I.I. Kalashnikova, S&R Institute "Poljus" & S.S. Pashkov, J.S.C. "Phonon", Russia

Langasite ($La_3Ga_5Si0_{14}$) An Optical Piezoelectric: Growth & Properties 43
M.F. Dubovik, I.A. Andreyev & Yu.S. Shmaly, Institute for Single Crystals of Academy of Sciences, Ukraine

Study of Gallium Phosphate and Langasite Crystals and Resonators by X-Ray Topography 48
B. Capelle, A. Zarka, L.M.C.P. Universités P VI et P VII, CNRS, J. Détaint, J. Schwartzel, CNET Laboratoire de Bagneux, A. Ibanez, E. Philippot, L.P.C.M.S. Université de Montpellier, CNRS-URA & J.P. Denis, Crismatec, France

Bulk Wave Propagation and Energy Trapping in the New Thermally Compensated Materials with Trigonal Symmetry 58
J. Détaint, J. Schwartzel, France Telecom CNET PAB/BAG, A. Zarka, B. Capelle, L.M.C.P. Universités Paris VI et Paris VII, CNRS, J.P. Denis, CRISMATEC & E. Philippot, L.P.M.S. Université du Languedoc, CNRS, France

Growth of Crack-Free 3-Inch Diameter Lithium Tetraborate Single Crystals by Czochralski Method 72
R. Komatsu, S. Uda, K. Hikita & T. Sugihara, Mitsubishi Materials Corporation

Defects in the Bridgman Grown Lithium Tetraborate of Three-Inch Diameter 78
S-J. Fan, R-Y. Sun & J. Wang, Shanghai Institute of Ceramics, P.R. China

An Improved Resonator Method for the Determination of Piezoelectric Material Constants..... 85
J.A. Kosinski, A. Ballato, U.S. Army Research Laboratory & Y. Lu, Rutgers, The State University of New Jersey

A New Method for the Determination of Concentrations of Impurities in Quartz Crystals 91
P. Zecchini, K. Yamni, B. Viard & D. Dothée, Université de Franche-Comté, France

Growth of High Quality Quartz Crystal and its Application to Temperature Sensors 99
S. Ishigami, M. Sato, Tokyo Denpa Company Ltd., F. Uchiyama, K. Agatsuma & K. Tsukamoto, Electrotechnical Laboratory Umezono, Japan

Outgassing of Quartz..... 107
D.E. Pierce, R.A. Murray, R. Lareau, J.R. Vig, U.S. Army Research Laboratory & S. Laffey, Vitronics, Inc.

High-Temperature Acoustic Loss of AT-Cut Quartz Crystals 115
J.E. Ferris & J.J. Martin, Oklahoma State University

Generation of Micro-Domains in AT-Cut Quartz by Thermal Processing and the Effect on Resonator Modes..... 122
L. Weisenbach, S.J. Martin, G.C. Frye, T.V. Bohuszewicz & D.H. Doughty, Sandia National Laboratories

THEORY, DESIGN, PROCESSING AND TESTING OF RESONATORS

Design and Performance of CMOS Micromechanical Resonator Oscillators 127
C.T.-C. Nguyen & R.T. Howe, University of California at Berkeley

Micromachined Thin Film Bulk Acoustic Resonators..... 135
R. Ruby & P. Merchant, Hewlett-Packard Laboratories

Dynamic Deflection and Capacitance of Micromachined Piezoelectric Benders..... 139
J.G. Smits & W-S. Choi, Boston University

A Study of Quartz Tuning Fork Resonators in the Overtone Mode..... 142
E. Momosaki, T. Kaneko & T. Shimoda, Seiko Epson Corporation

Analysis of a Piezoelectric Composite Microcantilever for Accelerometer Applications 149
J.T. Stewart, U.S. Army Research Laboratory

Calculation of the Optimal Clip Dimensioning to Minimize the Influence of Fabrication Imperfections on the Acceleration Sensitivity of SC-Cut Quartz Resonators with Stiffened Rectangular Support Systems 165
B.J. Lwo & H.F. Tiersten, Rensselaer Polytechnic Institute

An Analysis of Transversely Varying Thickness Modes in Trapped Energy Resonators with Shallow Contours 172
H.F. Tiersten, B.J. Lwo, Rensselaer Polytechnic Institute & B. Dulmet, Ecole Nationale Supérieure de Mécanique et des Microtechniques, France

A Perturbation Method for Modeling the Thermal Sensitivity of Surface Transverse Wave (STW) Propagation on a Piezoelectric Substrate 184
S. Ballandras, E. Gavignet & E. Bigler, Laboratoire de Physique et Métrologie des Oscillateurs du CNRS, France

Two-Dimensional Equations for Guided Electromagnetic Waves in Anisotropic Dielectric Plates Surrounded by Free Space 192
P.C.Y. Lee & J.D. Yu, Princeton University

The Effect of Thermal Dissipation in the Vibrations of a Piezoelectric Body 201
J.S. Yang & R.C. Batra, University of Missouri-Rolla

Computer Analysis of the Characteristics of High-Frequency Quartz Resonators 209
I.I. Postnikov, RD Institute of Electronic Measurement Instruments, Russia

Reliable Quadratic for Frequency-Turnover Temperature vs. Orientation of Rotated Y-Cut Quartz Plate Resonator Oscillating in C-Mode 223
M. Nakazawa, A. Araki, Shinsyu University, S. Takakuwa & M. Hanji, Kinseki Co. Ltd., Japan

Inclusion of Non-Uniform Distribution of Motion Effects in the Transmission-Line Analogs of the Piezoelectric Plate Resonator: Theory and Experiment 229
J.A. Kosinski, A. Ballato, U.S. Army Research Laboratory, I. Mateescu & I. Mateescu, Institute of Physics and Technology of Materials, Romania

Improved Ω-Scan for Separate Measurement of True AT-Cutting Angles and X-Miscutting Angles for Round Quartz Blanks 237
B. Morys, H. Bradaczek & G. Hildebrandt, Universität Berlin, Germany

Alternatives to Ozone Depleting Substances in Electronic Cleaning 241
J.R. Stemniski, Charles Stark Draper Laboratory, Inc.

Polishing and Etching Langasite and Quartz Crystals 245
S. Laffey, Vitronics, Inc., M. Hendrickson & J.R. Vig, U.S. Army Research Laboratory

Submicron Fabrication Technology Using Electro-Chemical Effects and Application to GHz-range Unidirectional SAW Transducers 251
K. Yamanouchi, H. Odagawa, T. Meguro, Y. Wagatsuma & K. Yamamoto, Tohoku University, Japan

Measurement of Temperature Characteristics of Two-Dimensional Surface Charge Patterns of Spurious Modes in Rectangular Quartz Plates..... 255
Y. Watanabe, H. Sekimoto & Y. Oomura, Tokyo Metropolitan University, Japan

Origin and Measurement of Quartz Resonator Magnetic Sensitivity 260
R. Brendel, B. Cretin & A. El Habti, Laboratoire de Physique et Métrologie des Oscillateurs du CNRS, France

Measurement of Spurious Resonances of Crystal Units Using Network Analyzers with Error Correction 268
B.W. Neubig, Tele Quarz Group, Germany

Test Oscillator for Study of Drive Level Dependence of Quartz Crystals..... 273
L. Lindberg & G. Portnoff, Quartz Pro Instrument AB, Sweden

Method and Devices for Measuring Driving Power of Quartz Crystal Units..... 281
M.E. Karaulnik & A.E. Karaulnik, "Ligas", Russia

Implementing ISO9000 in a Small Company..... 286
E.P. EerNisse & C.D. Mercer, Quartzdyne, Inc.

RESONATORS, FILTERS, SENSORS, SAW AND STW DEVICES

SAW Devices on Lithium Tetraborate ($Li_2B_4O_7$) *(INVITED)*..... 289
H. Abe, M. Ohmura & H. Saitou, Japan Energy Corporation, Japan

Leaky SAW Propagation Properties on $Li_2B_4O_7$ Substrates..... 296
M. Adachi, A. Kawabata, Toyama Prefectural University, T. Shiosaki & K. Ohtsuka, Kyoto University, Japan

Lithium Tetraborate as a Promising Material for BAW Filters 301
K.V. Shestopalov, V.A. Nefedov & B.I. Zadneprovsky, All-Russian Research Institute for the Synthesis of Materials (VNIISIMS), Russia

Sensitivity Analysis of One Port and Two Port BAW and SAW Resonator Model Parameters..... 308
B.D. Andersen, Piezo Technology, Inc., M. Cavin, Sawtek, Inc., M.A. Belkerdid & D.C. Malocha, University of Central Florida

An Analysis of the SAW Displacements in Quartz and Lithium Niobate by X-Ray Topography..... 315
A. Zarka, B. Capelle, M. Pilard, Université P. et M. Curie, J. Schwartzel, J. Detaint, France Telecom CNET/PAB & M. Solal, Thomson Sintra DTAS, France

Theoretical Analysis of Thermal Frequency Behavior of a SAW Device Being Caused by Laser Beam Irradiation and its Dependence on Angle of Cut..... 323
N. Oura & T. Nakase, Tokyo Institute of Technology, Japan

SAW Ring Filters with Insertion Loss of 1 dB 330
S.A. Dobershtein & V.A. Malyukhov, Omsk Research Electronic Equipment Institute, Russia

A SAW Resonator Filter Exploiting RMSCs 337
E. Danicki, Polish Academy of Sciences, Poland

Micro-Directional Couplers for Frequency Selection 343
E. Haruvi-Busnach, M. Zilberstein, N. Croitoru & A. Seidman, Tel-Aviv University, Israel

Extreme Pressure-Sensitive Cuts for Surface Acoustic Waves in α-Quartz 345
R.M. Taziev, E.A. Kolosovsky & A.S. Kozlov, Russian Academy of Sciences, Russia

Designing Smaller SAW Oscillators for Low Vibration Sensitivity 352
D. Andres & T.E. Parker, Raytheon Research Division

Dynamic and Static Pressure Sensitivities of AQP SAW Oscillators 359
J. Zhu, G.K. Montress & T.E. Parker, Raytheon Research Division

Design and Performance of an Extremely Low Noise Surface Acoustic Wave Oscillator 365
G.K. Montress & T.E. Parker, Raytheon Research Division

Impedance Element SAW Filters 374
J. Heighway, S.N. Kondratyev & V.P. Plessky, Advanced SAW Products SA, Switzerland

Surface Transverse Wave Oscillators with Extremely Low Thermal Noise Floors 379
I.D. Avramov, Institute of Solid State Physics, Bulgaria, F.L. Walls, T.E. Parker, National Institute of Standards & Technology & G.K. Montress, Raytheon Research Division

Highly Sensitive SAW Sensors *(INVITED)* 395
I.B. Yakovkin, Yu.N. Vlassov, A.S. Kozlov & S.M. Repinsky, Siberian Branch of Russian Academy of Sciences, Russia

Chemical/Biological Contaminant Detector for Aqueous Environments: Preliminary Report 401
R. McGowan, J. Foerster, D. Huynh, R. Lindenmuth, T. Lukaszek, U.S. Army Research Laboratory & Y. Lu, Rutgers University

High Frequency Monolithic Crystal Filters (MCF) Using Piezoelectric Elements Operating at the Fundamental Frequency 30-150 MHz and their Adjustment Method 405
S.V. Volkov, V.A. Mostiaev, L.V. Grigorjev, V.S. Frolov & Y.M. Ermakov, Scientific Research Institute "Phonon", Russia

Use of Active Networks to Widen the Spectrum of Application of Piezoelectric Filters 411
A.T. Orlov & V.F. Skripov, Kiev Polytechnic Institute, Ukraine

Monolithic Quartz Structure Vibrating Beam Accelerometer (VBA) 415
W.C. Albert, Crystal Gage, Inc.

DIELECTRIC RESONATORS AND OSCILLATORS

Temperature Compensated Sapphire Resonator for Ultra-Stable Oscillator Capability at Temperatures Above 77 Kelvin 421
G.J. Dick, D.G. Santiago & R.T. Wang, Jet Propulsion Laboratory

Low Noise Microwave Oscillators Based on High-Q Temperature Stabilized Sapphire Resonators 433
M.E. Tobar, E.N. Ivanov, R.A. Woode, University of Western Australia & J.H. Searls, Poseidon Scientific Instruments Pty. Ltd., Australia

Cryogenic Sapphire Resonator-Oscillator with Exceptional Stability: An Update 441
A.N. Luiten, A.G. Mann, M.E. Costa & D.G. Blair, University of Western Australia, Australia

Sapphire Disk Dielectric Resonator Temperature Coefficient of Frequency Dependence on Temperature, Disk Configuration and Resonant Mode 447
D.P. Tsarapkin, Moscow Power Engineering Institute, Russia

Low Phase Noise Sapphire Disk Dielectric Resonator Oscillator with Combined Stabilization 451
D.P. Tsarapkin, Moscow Power Engineering Institute, Russia

High Tuning Coefficient Whispering Gallery Modes in a Sapphire Dielectric Resonator Transducer 459
H. Peng & D.G. Blair, University of Western Australia, Australia

Ceramic Dielectric Resonator Oscillator Aging 466
M. Mizan, U.S. Army Research Laboratory

Development of a Low Noise L-Band Dielectric Resonator Oscillator 472
W.J. Tanski, United Technologies Research Center

Supporting Structures Effects on High Q Dielectric Resonators for Oscillator Applications 478
J-F. Liang, C. Wang, Q-F. Sun, K.A. Zaki, University of Maryland & R.W. Weinert, Westinghouse Science and Technology Center

Experimental Test and Application of a 2-D Finite Element Calculation for Whispering Gallery Sapphire Resonators 482
D.G. Santiago, R.T. Wang, G.J. Dick, Jet Propulsion Laboratory, R.A. Osegueda, J.H. Pierluissi, L.M. Gil, A. Revilla & G.J. Villalva, The University of Texas at El Paso

Stress Induced Frequency Shift of Dielectric Resonators with Magnetic Wall... 486
J.S. Yang, Princeton University & X.Y. Wu, Carleton University, Canada

2D Finite Element Program to Compute Resonant Frequencies and 3D Visualization of Electromagnetic Modes in Cylindrical Resonators 493
A. Revilla, R.A. Osegueda, J.H. Pierluissi, G.J. Villalva, The University of Texas at El Paso & G.J. Dick, Jet Propulsion Laboratory

Sapphire Loaded Microwave Resonators with Enhanced Quality Factor 500
E.N. Ivanov, D.G. Blair, University of Western Australia & V.I. Kalinichev, Moscow Institute of Energy, Russia

OSCILLATORS, CIRCUITRY, INSTABILITIES, AND APPLICATIONS

Fundamental Limits on the Frequency Instabilities of Quartz Crystal Oscillators *(INVITED)*..... 506
J.R. Vig, U.S. Army Research Laboratory & F.L. Walls, National Institute of Standards and Technology

1/f Frequency Fluctuations and Nonlinearity of Quartz Resonators and Quartz Crystal Oscillators 524
V.N. Kuleshov, Moscow Power Engineering Institute & G.D. Janushevsky, VEGA-M Scientific and Production Corporation, Russia

1/f Noise in Surface Acoustic Wave (SAW) Resonators 530
T.E. Parker & D. Andres, Raytheon Research Division

Analysis of Quantum 1/f Effects in Frequency Standards..... 539
P.H Handel, University of Missouri - St. Louis & F.L. Walls, National Institute of Standards and Technology

Irreversible Frequency Variations in Time of Precision Quartz Crystal Units 541
A.G. Smagin, "Crystal" Company, Russia

Study on Characterization of Frequency Stability in Time Domain..... 544
S. Hongwei, Shaanxi Astronomical Observatory, P.R. China

Design and Optimization of Low-Noise Oscillators Using Nonlinear CAD Tools 548
U.L. Rohde, C-R. Change & J. Gerber, Compact Software, Inc.

Nonlinear Oscillator Design for Maximum Power 555
M.C.E. Yagoub & H. Baudrand, Laboratoire d'Électronique, ENSEEIHT, France

Computer Aided Design of Quartz Crystal Oscillators 559
R. Brendel, G. Marianneau, T. Blin, Laboratoire de Physique et Métrologie des Oscillateurs du CNRS & M. Brunet, Centre National d'Études Spatiales, France

A New 5 and 10 MHz High Isolation Distribution Amplifier..... 567
C.W. Nelson, Spectra Dynamics, F.L. Walls, National Institute of Standards and Technology, M. Sicarrdi & A. DeMarchi, Politecnico di Torino, Italy

A No-Drift and Less than 1 X 10^{-13} Long-Term Stability Quartz Oscillator Using a GPS SA Filter 572
J.A. Kusters, K. Ho, Hewlett-Packard Co., R.P. Giffard, L.S. Cutler, Hewlett-Packard Laboratories & D.W. Allan, Allan's TIME

Dual-Mode Crystal Oscillators with Resonators Excited on B and C Modes..... 578
A.V. Kosykh, I.V. Abramson & V.P. Bagaev, Omsk Radio-engineering Company Ltd., Russia

Temperature Compensation of Crystal Oscillators Using Microcontroller - μCTCXO 587
D. Habic & D. Vasiljevic, University of Belgrade, Yugoslavia

TCXO's Employing NS-GT Cut Quartz Crystal Resonators for Cellular Telephones 594
E. Takahashi, O. Teshigawara & K. Yamashita, Japan Radio Co., Ltd., Japan

Operation of a Neural Network Controlled Crystal Oscillator 600
D.B. Opie, C.T. Butler, J. Yates, M.C. Sharp, Physical Sciences Inc., W.M. Golding, SFA Inc. & F. Danzy, Naval Research Laboratory

The Functions of the Precision Frequency Measurement Technique in MCXO..... 604
W. Zhou & M. Wang, Xidian University, P.R. China

Application of LFE SC-Cut Resonators at High Stability Oscillators..... 608
A. Masiukiewicz, B. Gniewinska & W. Szulc, Tele and Radio Research Institute, Poland

Spectral Properties of DDFS: Computer Simulations and Experimental Verifications 613
V.F. Kroupa, Academy of Sciences, Czech Republic

Design and Analysis Methods of a DDS-Based Synthesizer for Military Spaceborne Applications..... 624
T.J. Endres, R.B. Hall & A.M. Lopez, Hughes Space and Communications Company

Frequency Control Requirements of Radar *(INVITED)* 633
N. Slawsby, ANRO Engineering, Inc.

Frequency Control Devices Applicable to Radar Systems 641
S.P. Caldwell, Westinghouse Electric Corporation

Design and Performance of an Ultra-Low Phase Noise, Radar Exciter 647
M.M. Driscoll, A.C. Hazzard & D.G. Opdycke, Westinghouse Electronics Systems Group

ATOMIC FREQUENCY STANDARDS, AND TIME TRANSFER

A Factor 1600 Increase in Neutral Atomic Beam Intensity Using Laser Cooling 651
M.D. Hoogerland, National Institute of Standards and Technology, J.P.J. Driessen, E.J.D. Vredenbregt, H.J.L. Megens, M.P. Schuwer, H.C.W. Beijerinck & K.A.H. van Leeuwen, Eindhoven University of Technology, The Netherlands

Optical Lattices for Atomic Fountain Frequency Standards 655
W.D. Phillips, W.M. Golding, A. Kastberg, S.L. Rolston & R.J.C. Spreeuw, National Institute of Standards and Technology

Velocity Distributions of Atomic Beams by Gated Optical Pumping 658
W.D. Lee, J.H. Shirley & R.E. Drullinger, National Institute of Standards and Technology

A Hybrid Digital/Analog Servo for the NIST-7 Frequency Standard 662
J.P. Lowe, W.D. Lee, F.L. Walls & R.E. Drullinger, National Institute of Standards and Technology

High-Frequency Oscillators Using Phase-Locked Arrays of Josephson Junctions 666
S.P. Benz & P.A.A. Booi, National Institute of Standards and Technology

Atomic Hydrogen Spin-Exchange Frequency Shift Cross Section at 0.5 Kelvin 670
M.E. Hayden, M.D. Hürlimann & W.N. Hardy, University of British Columbia, Canada

Cryogenic Hydrogen Maser at 10 Kelvin 677
D.R. McAllaster, J.J. Krupczak, A.L. Cole, A.J. Kerman & S.B. Crampton, Williams College

A Time Transfer Technique Using a Space-Borne Hydrogen Maser and Laser Pulse Timing 684
E.M. Mattison, L.M. Coyle, R.F.C. Vessot, D. Arnold, Harvard-Smithsonian Center for Astrophysics & R.C. Smith, Los Alamos National Laboratory

Radiation Resistant and Reproducible Fluorine-Based Coatings of H-Maser Storage Box for a Possible Use in Space 687
A. Kalachev, Institute of Petrolchemical Synthesis, V. Marihin, L. Myasnikova, V. Svetlov, Joffe-Phisico-Technical Institute of Russian Academy of Science, D. Muzhdaba & A. Myasnikov, Russian Institute of Radionavigation and Time, Russia

Frequency Control of Hydrogen Masers Using High Accuracy Calibrations 695
J.-S. Boulanger & R.J. Douglas, National Research Council, Canada

A Hydrogen Maser for Long-Term Operation in Space 709
R.F.C. Vessot, D.A. Boyd, L.M. Coyle, R.F. Jendrock, E.M. Mattison, G.U. Nystrom, Harvard-Smithsonian Center for Astrophysics & T.E. Hoffman, Hoffman Design and Development

Developments of Rubidium Frequency Standards at Neuchatel Observatory 716
P. Rochat, H. Schweda, G. Mileti & G. Busca, Observatoire Cantonal de Neuchatel, Switzerland

The Double Bulb Rubidium Maser 724
W.M. Golding, A. Frank, SFA Inc., R. Beard, J. White, F. Danzy & E. Powers, Naval Research Laboratory

Performance of a Prototype Microwave Frequency Standard Based on Trapped $^{171}Yb^{+}$ Ions 731
P.T.H. Fisk, M.J. Sellars, M.A. Lawn, C. Coles, CSIRO Division of Applied Physics, A.G. Mann & D.G. Blair, University of Western Australia, Australia

Recent Stability Comparisons with the JPL Linear Trapped Ion Frequency Standards 739
R.L. Tjoelker, J.D. Prestage, G.J. Dick & L. Maleki, Jet Propulsion Laboratory

Progress on a Cryogenic Linear Trap for $^{199}Hg^{+}$ Ions 744
M.E. Poitzsch, J.C. Bergquist, W.M. Itano & D.J. Wineland, National Institute of Standards and Technology

Space Flyable Hg^{+} Frequency Standards 747
J.D. Prestage & L. Maleki, Jet Propulsion Laboratory

Progress Report on the Improved Linear Ion Trap Physics Package 755
J.D. Prestage, R.L. Tjoelker, G.J. Dick & L. Maleki, Jet Propulsion Laboratory

Progress Toward the Development of a Ytterbium Ion Standard 761
D.J. Seidel & L. Maleki, Jet Propulsion Laboratory

Second Harmonic Level Monitors in Cesium Atomic Frequency Standards 769
B. Jaduszliwer & J.P. Hurrell, The Aerospace Corporation

Glonass On-Board Time/Frequency Standard - Architecture and Operations 776
P.P. Bogdanov, Y.G. Gouzhva, A.G. Gevorkyan, A.B. Bassevich & A.E. Tyulyakov, Russian Institute of Radionavigation and Time (RIRT), Russia

Effect of Space Vehicle Temperature on the Frequency of On-Orbit Navstar Clocks 782
W.G. Reid, T.B. McCaskill, O.J. Oaks, U.S. Naval Research Laboratory, J.A. Buisson & H.E. Warren, Sachs Freeman Associates, Inc.

Wavelet Analysis for Synchronization and Timekeeping 791
D.A. Howe, National Institute of Standards and Technology & D.B. Percival, University of Washington

Accumulation of Random Noise in a Chain of Slave Clocks 798
G.M. Garner, AT&T Bell Laboratories

PROCEEDINGS ORDERING INFORMATION 812

SPECIFICATIONS AND STANDARDS RELATED TO FREQUENCY CONTROL . 813

AUTHOR INDEX 816

1994 AWARD WINNERS

W.G. Cady Award

The W.G. Cady Award was presented to Thomas E. Parker, Raytheon Research Division, "for outstanding contributions to the theory and applications of high-precision SAW oscillators." The award was presented by Arthur Ballato, U.S. Army Research Laboratory.

I.I. Rabi Award

The I.I. Rabi Award was presented to Jacques Vanier, National Research Council of Canada, "for his contributions to the application of quantum theory to the development of atomic frequency standards, and for his leadership in promoting the whole field and making it widely accessible to students and junior colleagues." The award was presented by Robert Vessot, Harvard-Smithsonian Center for Astrophysics.

C.B. Sawyer Memorial Award

The C.B. Sawyer Memorial Award was presented to Roger W. Ward, Quartzdyne, Inc., "for technical and industrial contributions to the design and fabrication of quartz crystal resonators." The award was presented by Charles Adams, Hewlett-Packard Company.



Jacques Vanier, I.I. Rabi Award winner; Thomas E. Parker, W.G. Cady Award winner; and Roger W. Ward, C.B. Sawyer Memorial Award winner.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

APPLICATIONS OF SILICON MICROMACHINING TO RESONATOR FABRICATION

Roger T. Howe

University of California at Berkeley
Berkeley Sensor & Actuator Center
Department of Electrical Engineering and Computer Sciences
231 Cory Hall # 1770, Berkeley, California 94720-1770 USA
Internet: howe@EECS.Berkeley.EDU

Abstract

This paper surveys recent developments in using silicon micromachining processes to fabricate micromechanical resonant structures for sensors, filters, and oscillators. Single-crystal silicon and polycrystalline silicon are excellent mechanical materials with quality factors approaching those of quartz. By etching sacrificial layers from a sandwich of thin films, microstructures can be surface micromachined from a deposited polycrystalline silicon (polysilicon) film. This technology can be applied to fabricate micromachined vacuum encapsulation for the resonant structure, as well. Surface micromachining has been merged with CMOS processes to enable fabrication of integrated micro electromechanical systems (MEMS) on a single silicon chip. An alternative process fabricates crystalline silicon resonators by means of wafer bonding and etch-back. These microstructures can be over 10 μm thick -- much thicker than surface-micromachined polysilicon structures. Integration with this technology is feasible by processing CMOS electronics on wafers containing sealed cavity microstructures, with the microresonators defined at the end by reactive-ion etching. Trimming silicon microresonators is possible using micromechanical fusing and welding structures, or by electrical tuning.

Introduction

Silicon micromachining technology is a collection of processes derived from or related to integrated-circuit fabrication. Over the past twenty years, a variety of approaches have developed for micromachining structures from single crystal silicon or from thin films deposited on the silicon substrate. These technologies are the focus of intensive research and development in the US, Japan, and Europe for commercialization in low-cost, high-reliability silicon sensors and actuators. Early applications include accelerometers for air-bag deployment, accelerometers and gyroscopes for vehicle control systems, and valves for fuel injectors.

Micromechanical resonators have been a major focus of research in silicon micromachining, starting with the pioneering plated-metal cantilever beams of Nathanson in the 1960s [1]. Frequency modulation is an attractive means of transducing a measurand into an electrical signal [2], since frequency is a quasi-digital output that can be easily measured to high precision. Silicon resonant sensors have been demonstrated for a wide variety of measurands, including linear acceleration, angular rotation rate, pressure, force, and flow [3].

The core element of resonant sensors is the vibrating mechanical structure. Unlike quartz, silicon is not piezoelectric. Many approaches have been investigated for the excitation and detection of resonance in micromachined resonators. Piezoelectric films, such as ZnO or PZT, can be sputter-deposited and sandwiched between conductors built into the resonator [4]. Thermal excitation is feasible to quite high frequencies, over 100 kHz for some microstructures [5]. Piezoresistors built into the resonator are effective, low-impedance vibration pickups [6]. Magnetic excitation has the advantage of minimizing the interconnections needed to the resonator, although an external bias magnetic field is needed [7]. Optical drive and detection is convenient for an optical fiber-based measurement system [8].

Electrostatic drive and detection has the advantage of simplicity, since tightly controlled gaps between the conducting silicon or thin-film structure and fixed plates are a by-product of most silicon micromachining processes [2]. With electrostatic excitation, the microstructure can be a single conducting layer instead of a sandwich of films, with the result being a higher quality factor and improved frequency stability. Due to space limitations, this paper will discuss this class of resonant structures exclusively.

At present, there are two main branches of silicon micromachining: the polysilicon surface micromachining process and wafer-bonding based processes. For resonator fabrication, the latter is much more attractive than the more mature silicon bulk micromachining technology, which is described in Petersen's review paper [9]. Using electrostatically driven and sensed resonators as examples, the basic features of these two technologies will be outlined. Merging microresonator and electronic fabrication is a current research activity in both processes and will be discussed.

Surface Micromachining

The key step in surface micromachining is to etch selectively a sacrificial layer underlying an etch-resistant thin film. Originally demonstrated with metal films [1], the dominant material is now polycrystalline silicon (polysilicon) [10]. Polysilicon is a high-Q mechanical material that is deposited conformally by low-pressure chemical vapor deposition (LPCVD) at temperatures from 570°C to 620°C. The sacrificial layer of choice is SiO_2, since it is etched by solutions of hydrofluoric acid (HF), which attacks polysilicon very slowly. Phosphosilicate glass (PSG) deposited by LPCVD etches fastest among the oxide films and is a commonly used oxide sacrificial layer.

0-7803-1945-1/94 $3.00 © 1994 IEEE

The lateral comb-drive resonator shown schematically in Fig. 1 is a relevant example of a polysilicon surface microstructure [12]. Unlike parallel-plate capacitors, the interdigitated capacitors apply an electrostatic force to the movable plate that is independent of its displacement, to first order. This feature enables high-amplitude vibration with far less frequency pulling, which is very useful in some resonant sensors. The folded-beam suspension improves the linearity of the suspension and also allows relaxation of any residual stress in the polysilicon film. The vibration of the plate, which is biased with a DC voltage, is sensed by detecting the motional current induced in the other stationary comb. Equivalent circuits for this resonant structure have been derived [2, 13].

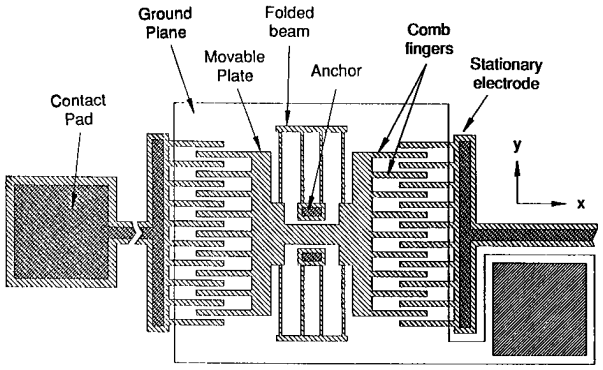


Figure 1. Lateral polysilicon resonator [12].

The fabrication of this polysilicon microstructure requires four masking steps, as shown in cross section in Fig. 2. After a blanket n^+ diffusion, which defines the substrate ground plane, the wafer is passivated with a layer of 1500 A-thick LPCVD nitride deposited on top of a layer of 5000 A-thick thermal SiO_2. Contact windows to the substrate ground plane are then opened, followed by the deposition and definition of an *in situ* phosphorus doped polysilicon layer in step (3) in Fig. 2. This layer serves as a second electrode plane and the interconnection to the n^+ diffusion and the microstructure.

A 2 μm-thick LPCVD sacrificial phosphosilicate glass (PSG) layer is deposited and patterned with the third mask, which defines the anchors of the microstructures in step (4) in Fig. 2. The 2 μm-thick polysilicon structural layer is then deposited by LPCVD (undoped) at 605 °C. The structural layer is doped by depositing another layer of 3000 A-thick PSG and then annealing at 950 °C for one hour. This doping process is designed to dope the polysilicon symmetrically by diffusion from the top and the bottom layers of PSG, thereby avoiding stress gradients that could cause warpage of the released structures.

Following a further stress-annealing step at 1050 °C for 30 minutes, the top PSG layer is stripped and the plates, beams, and electrostatic comb drive and sense structures are defined in the final masking in step (6) in Fig. 2. The structures are anisotropically patterned in a CCl_4 plasma by reactive-ion etching, in order to achieve nearly vertical sidewalls. Finally, the wafer is immersed in 10:1 diluted HF to etch the sacrificial PSG, rinsed, and dried. In order to avoid stiction of the microstructure to the substrate due to capillary forces during drying, the wafer can be dried using a supercritical CO_2 process [14].

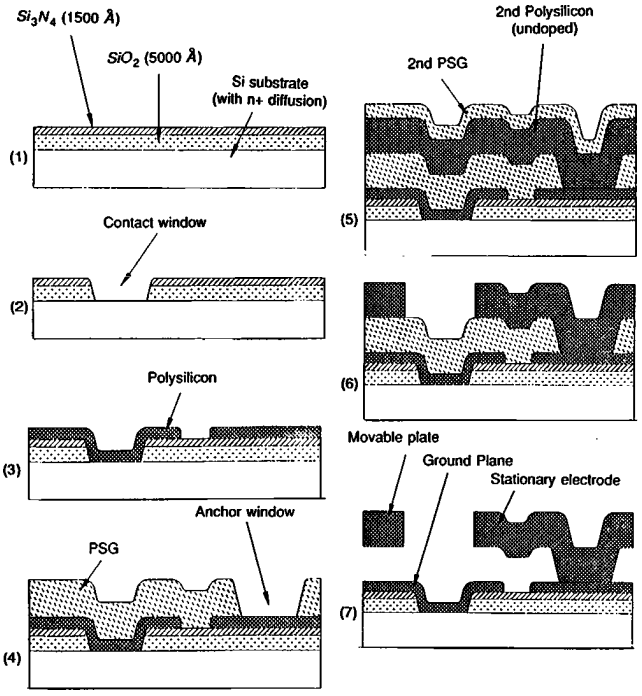


Figure 2. Polysilicon lateral resonator fabrication [12].

Figure 3 is an SEM of a torsional microresonator fabricated by this process. The spiral suspension has a 2 μm x 2 μm cross section. Polysilicon has been studied extensively over the past decade and the connection between deposition conditions and its mechanical properties is understood, at least empirically [15]. Quality factors of the order of 75,000 for resonator frequencies under 100 kHz have been measured at low pressures, for both PSG-doped polysilicon (as in Fig. 2) [13] and undoped, fine-grained polysilicon [16].

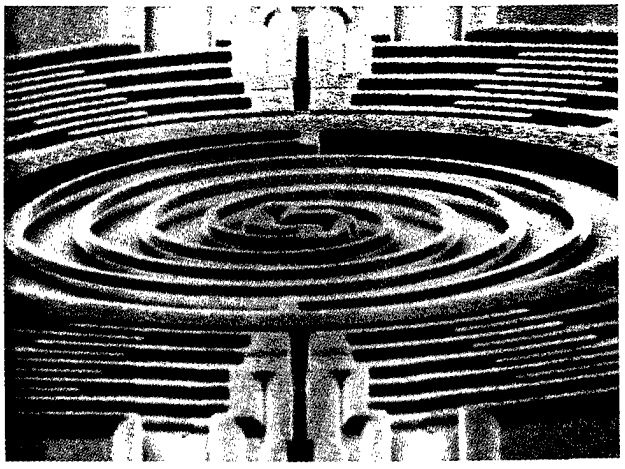


Figure 3. SEM of polysilicon torsional microresonator [12].

Micro Vacuum Chambers

In order to achieve high-Q operation, microresonators must be operated in a vacuum. Fabrication of a thin-film shell around the structure to provide the hermetically sealed, evacuated ambient can be achieved with a few additional masking steps. The implication is that standard electronic packages could be used for precision microresonators, since the critical encapsulation is accomplished by an extension of the micromachining process.

Figure 4 is an SEM of an encapsulated polysilicon resonant structure [16,17]. After deposition and patterning of the resonator, an additional sacrificial layer is deposited to form a spacer between the structure and the shell layer. Access holes are photolithographically defined at the perimeter to allow removal of the sacrificial layer. A final structural layer is deposited, which seals the access holes and forms the micro vacuum chamber. This process, pioneered at the University of Wisconsin, is being developed by Honeywell for precision resonant pressure sensors and accelerometers [18, 19]. Electrostatically driven polysilicon resonators encapsulated in thin-film vacuum chambers have short term stability of less than 0.02 Hz, for a resonant frequency of 625 kHz [19]. Failure-free operation for over 3 years with less than 0.4 ppm long-term frequency variations demonstrates conclusively the suitability of polysilicon for precision applications. Recently, a lateral comb-drive resonator has been encapsulated in a transparent LPCVD silicon nitride film [20].

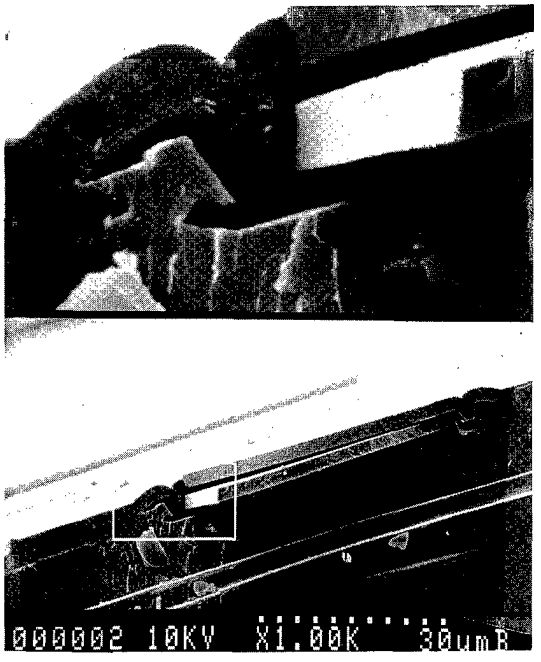


Figure 4. SEM showing cleaved wafer, with microencapsulated polysilicon beam. (courtesy of Prof. H. Guckel, Univ. of Wisconsin - Madison).

Surface micromachining is a flexible technology that can fabricate complex micromechanical elements. For example, a lateral resonator has been fabricated on top of a temperature controlled micro platform [21, enabling on-chip oven stabilization. In addition, self-aligned process techniques borrowed from VLSI fabrication are being exploited to make submicron resonators from single-crystal silicon [22], as well as hollow polysilicon microstructures [23].

Wafer Bonding

Since the mid-1980s, a number of academic and industrial groups in the US and Europe have been developing a micromachining technology based on silicon-silicon or silicon-glass wafer bonding. In contrast to surface micromachining, wafer-bonded microstructures are single crystal silicon, which has the benefit that the mechanical properties of the resonator are well known. Single crystal silicon resonant structures have very low internal damping [24, 25], with room temperature quality factors of over 600,000 having been reported. Wafer bonding processes also can fabricate thicker microstructures than surface micromachining processes, which can be advantageous for some applications.

Lateral resonant microstructures have been fabricated using a silicon-on-glass process developed at the University of Michigan [26], as shown in a perspective view in Fig. 5.

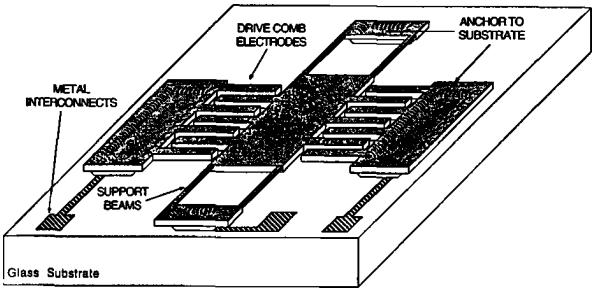


Figure 5. Silicon -on-glass lateral resonator (courtesy of Prof. K. Najafi, University of Michigan)

The fabrication sequence is shown in Fig. 6 below. Recesses are etched in the silicon wafer using wet etching in KOH or a plasma etch process to a depth of 3-5 µm. A deep boron diffusion is then performed, which defines the thickness of the micromechanical structures. Their lateral dimensions are defined by a reactive-ion etch (RIE) using a nickel mask, yielding the cross section in Fig. 6(c). It is important to optimize the etch chemistry for vertical sidewalls [26]. The nickel mask is then removed and the silicon wafer process is complete. The glass substrate, having had partially recessed nickel interconnects patterned on it, is then electrostatically bonded to the front (micromachined) side of the silicon wafer. Immersion of the silicon-glass sandwich in EDP, a standard silicon etchant, dissolves all of the silicon wafer except for the heavily boron doped microstructure. Note carefully that the structural layer is inverted between Fig. 6(c) and Fig. 6(d) and that the bulk of the silicon wafer (actually 50 x thicker than the microstructure) is dissolved. in the final process step.

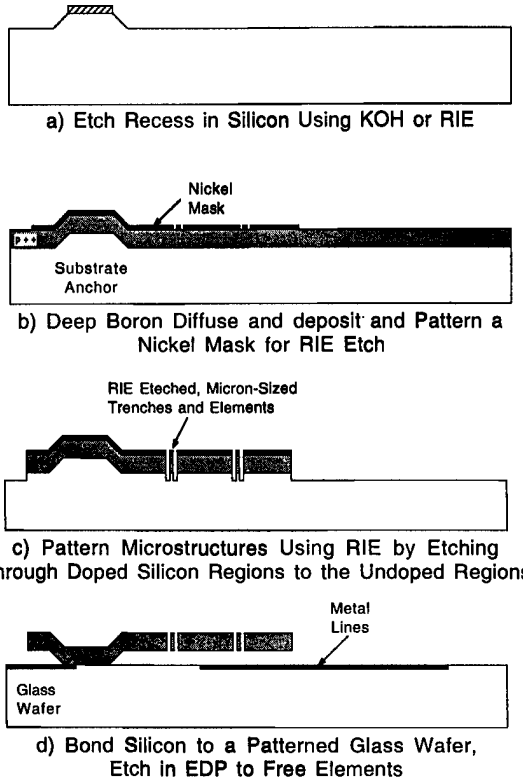


Figure 6. Fabrication sequence for the silicon-on-glass lateral resonator (courtesy of Prof. K. Najafi, University of Michigan).

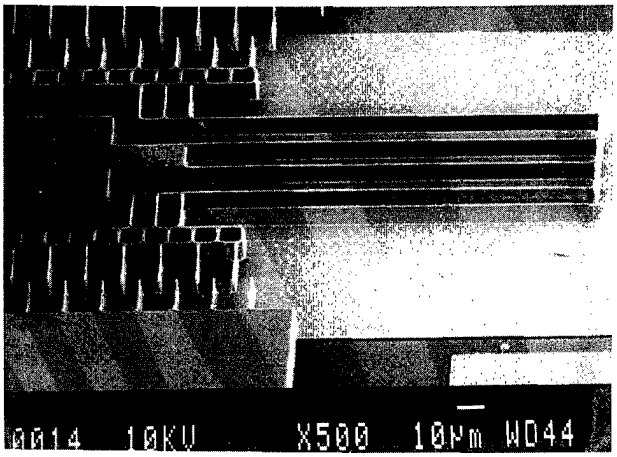


Figure 7. SEM of silicon-on-glass lateral resonator. (courtesy of Prof. K. Najafi, University of Michigan)

Figure 7 shows an SEM of a portion of the lateral resonator. This structure is 10 μm thick, with 1 μm-thick suspension beams and comb fingers. Silicon-on-glass technology has been applied by C. S. Draper Laboratory to fabricate a tuning-fork gyroscope using an electrostatic comb drive [27]. A significant advantage of the silicon-on-glass technology is that the parasitic capacitances are substantially lower than for surface-micromachined devices. One disadvantage is that the heavily boron-doped structural layer has a built-in residual tensile strain and may have degraded long-term stability, when compared with lightly doped single crystal silicon.

Silicon-silicon bonding is a related process to silicon-on-glass technology. After the demonstration of silicon-silicon direct bonding in the mid-1980s, researchers in silicon micromachining applied the technique to fabricate pressure sensors [28] and more recently, to a variety of micromechanical structures [29-31]. The critical step is precisely thinning the bonded wafer to leave the microstructural layer. There are several techniques for accomplishing this goal without the need for heavy boron doping of the structural film. Figure 8 is an SEM of a 10 μm-thick, wafer-bonded comb drive lateral resonator. This structure was found to have essentially zero residual strain, which enabled the resonant frequency to be accurately predicted from its dimensions and from the single crystal silicon mechanical properties. For this reason, wafer-bonding processes are attractive for high-precision resonator fabrication.

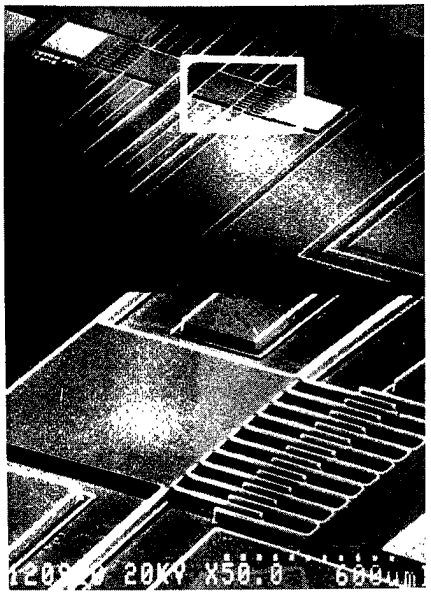


Figure 8. SEM of wafer-bonded tuning fork gyroscope. (courtesy of Prof. M. A. Schmidt, MIT)

Integrated Processes

Microfabricated resonators could have applications as on-chip building blocks for VLSI microsystems, if they can be compatibly fabricated with CMOS electronics. Some potential applications of lateral resonators include integrated mechanical filters [32], integrated frequency references [33], and mixer-demodulators for communication systems. A modular process for integrating CMOS with polysilicon micromechanics has been under development at the University of California at Berkeley for the past five years [34-36]. Figure 9is a cross section of the Berkeley MICS technology, which incorporates two structural polysilicon films integrated with a 3 μm p-well CMOS process. The "poly-3" layer is useful for controlling out-of-plane motion of the structure, which could, for example, pro-

vide a servo-controlled output in a vibratory gyroscope. Micro vacuum chambers have yet to be demonstrated in the MICS process. Analog Devices, Inc. is manufacturing an air-bag deployment accelerometer using a single polysilicon structural film integrated with an analog bipolar/MOS process [37].

Silicon wafer bonding processes can also achieve the goal of integration of resonant microstructures with CMOS. Wafers having pre-fabricated sealed cavities are run through a standard CMOS fabrication sequence [38]. Microstructures such as the lateral resonator in Fig. 8 are then released by reactive ion etching through the single crystal silicon diaphragm that seals the cavity. In order to fabricate an upper electrode, a cover wafer must be bonded to the completed integrated CMOS/microstructure wafer. Interconnections between the wafers and the reliable bonding of pre-processed wafers remain areas for further research [29].

Conclusions

Silicon micromachining is a rapidly maturing technology that has great potential for frequency control applications. In addition to resonant sensors for a variety of physical variables, the merger of microstructure and electronic processes makes feasible the integration of frequency-reference and signal processing functions onto silicon VLSI systems.

Silicon microresonator frequency adjustment will be necessary for these applications to be practical. Given the dimensional control on material properties and deposition and etching processes, a polysilicon lateral resonator can be fabricated to within around 1% of its design frequency. With silicon wafer bonding processes, a tighter distribution is expected due to the larger dimensions and predictable mechanical properties of the single crystal structures. For polysilicon microstructures, frequencies can be adjusted by altering spring constants using fusing or welding structures [39]. Electrical adjustment of the resonator frequency response is convenient for lateral resonators, since the extra electrical port can be added without additional process complexity. For example, a vibrating-ring gyroscope uses a lateral parallel-plate capacitor for mode balancing [40]. In the case of micromechanical filters, the quality factor can also be adjusted using a third port on the lateral resonator [13]. Given the variety of options, cost-effective and stable means for frequency adjustment can be developed that will enable the manufacture of silicon microresonator sensors and systems.

Acknowledgements

The author acknowledges the major contributions of his former graduate student W. C. Tang, as well his current graduate students C. T.-C. Nguyen, and G. K. Fedder to the design and fabrication technology for polysilicon microresonators. In addition, he is grateful to Profs. H. Guckel, K. Najafi, and M. A. Schmidt for supplying SEMs. Microresonator research at Berkeley has been sponsored by the Berkeley Sensor & Actuator Center, an NSF/Industry/University Cooperative Research Center.

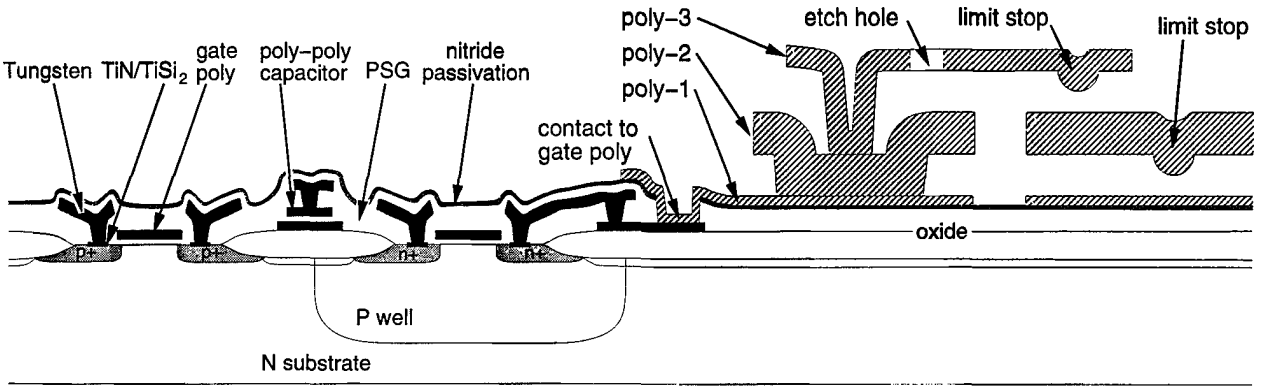


Figure 9. Cross section of the MICS technology [36].

References

[1] H. C. Nathanson, W. E. Newell, R. A. Wickstrom, and J. R. Davis, Jr., "The resonant gate transistor," *IEEE Trans. Electron Devices*, **ED-14**, 117-133, (1967).

[2] R. T. Howe, "Resonant microsensors," *4th Int. Conf. on Solid-State Sensors and Actuators (Transducers '87)*, Tokyo, Japan, June 1987, 843-848.

[3] A. Prak, *Silicon Resonant Sensors: Operation and Response*, Ph.D. Thesis, Oct. 1993, University of Twente, MESA Research Institute, Box 217, 7500-AE Enshede, Neth. ISBN 90-9006407-9. Extensive bibliography.

[4] D. B. Hicks, *et al*, "Piezoelectrically activated resonant bridge microaccelerometer," *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, S.C., June 1994, 225-228. ISBN 0-9640024-0-X.

[5] T. S. J. Lammerink, *et al*, "Performance of thermally excited resonators," *Sensors and Actuators*, **A21-23**, 352-356 (1990).

[6] J. D. Zook, *et al*, "Characteristics of polysilicon resonant microbeams," *Sensors and Actuators*, **A35**, 51-59 (1992).

[7] A. D. Nikolich and S. D. Senturia, "A wafer-bonded load cell operating in the tensioned wire regime," *IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head Island, S.C., June 1992, 157-160.

[8] L. M. Zhang, D. Utaamchandani, and B. Culshaw, "Excitation of silicon microresonators using short optical pulses," *Sensors and Actuators*, **A21-23**, 391-393 (1990).

[9] K. E. Petersen, "Silicon as a mechanical material," *Proc. of the IEEE*, **70**, 420-457 (1982).

[10] R. T. Howe, "Surface micromachining for microsensors and microactuators," *J. Vac. Sci. and Tech., B*, **6**, 1809-1813 (1988).

[11] D. J. Monk, D. S. Soane, and R. T. Howe, "A chemical reaction mechanism and kinetics for hydrofluoric acid etching of silicon dioxide thin films," *Thin Solid Films*, **232**, 1-12, (1993).

[12] W. C. Tang, T.-C. H. Nguyen, and R. T. Howe, "Laterally drive polysilicon resonant microstructures," *Sensors and Actuators*, **20**, 25-32 (1989).

[13] C. T.-C. Nguyen and R. T. Howe, "Quality factor control for micromechanical resonators," *IEEE Int. Electron Devices Meeting*, San Francisco, Calif., Dec. 1992, 505-508.

[14] G. T. Mulhern, D. S. Soane, and R. T. Howe, "Supercritical carbon dioxide drying of microstructures," *7th Int. Conf. on Solid-State Sensors and Actuators (Transducers '93)*, Yokohama, Japan, June 1993, 296-299.

[15] P. Krulevitch, G. C. Johnson, and R. T. Howe, "Stress and microstructure in phosphorus doped polycrystalline silicon," *Materials Research Society Symposium Proceedings*, **276**, 79-84 (1992).

[16] H. Guckel, *et al*, "The application of fine-grained tensile polysilicon to mechanically resonant transducers," *Sensors and Actuators*, **A21-23**, 346-351 (1990).

[17] J. J. Sniegowski, *Design and Fabrication of the Polysilicon Resonating Beam Force Transducer*, Ph.D. Thesis, Dept. of Electrical and Computer Engineering, Univ. of Wisconsin - Madison, 1991.

[18] H. Guckel, *et al*, "Polysilicon resonant microbeam technology for high performance sensor applications," *IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head Island, S.C., June 1992, 153-156.

[19] D. W. Burns, *et al*, "A digital pressure sensor based on resonant microbeams," *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, S.C., June 1994, 221-224.

[20] C. T.-C. Nguyen and R. T. Howe, "Microresonator frequency control using an integrated micro oven," *7 th Int. Conf. on Solid-State Sensors and Actuators (Transducers '93)*, Yokohama, Japan, June 1993, 1040-1043.

[21] L. Lin, *et al*, "Vacuum-encapsulated lateral microresonators," *7 th Int. Conf. on Solid-State Sensors and Actuators (Transducers '93)*, Yokohama, Japan, June 1993, 270-273.

[22] R. E. Mihailovich, *et al*, "Single-crystal silicon torsional resonators," *IEEE Micro Electro Mechanical Systems Workshop (MEMS-93)*, Ft. Lauderdale, Florida, Jan. 1993, 184-188.

[23] M. W. Judy and R. T. Howe, "Polysilicon hollow beam lateral resonators," *IEEE Micro Electro Mechanical Systems Workshop (MEMS-93)*, Ft. Lauderdale, Florida, Jan. 1993, 265-271.

[24] T. V. Roszhart, "The effect of thermoelastic internal friction on the Q of micromechanical silicon resonators," *IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head Island, S.C., June 1990, 13-16.

[25] R. A. Buser, *Theoretical and Experimental Investigations on Silicon Single Crystal Resonant Structures*, Ph.D. Thesis, Inst. of Microtechnology, University of Neuchatel, CH-2000 Neuchatel, Switzerland, July 1989.

[26] Y. Gianchandani and K. Najafi, "Micron-sized, high aspect ratio bulk silicon micromechanical devices," *IEEE Micro Electro Mechanical Systems Workshop (MEMS-92)*, Travemuende, Germany, Feb. 1992, 208- 213.

[27] J. Bernstein, *et al*, "A micromachined comb-drive tuning fork rate gyroscope," *IEEE Micro Electro Mechanical Systems Workshop (MEMS-93)*, Ft. Lauderdale, Florida, Jan. 1993, 143-148.

[28] K. E. Petersen, *et al*, "Silicon fusion bonding for pressure sensors," *IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head Island, S.C., June 1990, 144-147.

[29] M. A. Schmidt, "Silicon wafer bonding for micromechanical devices," *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, S.C., June 1994, 127-131.

[30] C. H. Hsu and M. A. Schmidt, "Micromachined structures fabricated using a wafer-bonded sealed cavity process," *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, S.C., June 1994, 151-155.

[31] K. Warren, "Navigation grade silicon accelerometers with sacrificially etch SIMOX and BESOI structure," *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, S.C., June 1994, 69-72.

[32] L. Lin, *et al*, "Micro Electromechanical Filters," *IEEE Micro Electro Mechanical Systems Workshop (MEMS-92)*, Travemuende, Germany, Feb. 1992, 226-231.

[33] C. T.-C. Nguyen and R. T. Howe, "CMOS Micromechanical Resonator Oscillator," *IEEE Int. Electron Devices Meeting*, Washington, D.C., Dec. 1993, 199-202.

[34] W. Yun, W. C. Tang, and R. T. Howe, "Fabrication technologies for integrated microdynamic systems," in F. Harashima, ed., *Integrated Micro-Motion Systems: Micromachining, Control, and Applications*. Amsterdam: Elsevier. 1990, 297-312.

[35] W. Yun, R. T. Howe, and P. R. Gray, "Surface micromachined, digitally force-balanced accelerometer with integrated CMOS detection circuitry," *IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head Island, S.C., June 1992, 126-131.

[36] G. K. Fedder and R. T. Howe, "Integrated testbed for multi-mode digital control of suspended microstructures," *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, S.C., June 1994, 145-150.

[37] R. S. Payne and K. A. Dinsmore, "Surface Micromachined Accelerometer: A Technology Update," *Soc. Auto. Eng. P-242*, Detroit, Mich., Feb. 1991, 127-135.

[38] L. Parameswaran, *et al*, "Sealed-cavity microstructure using wafer bonding technology," *7 th Int. Conf. on Solid-State Sensors and Actuators (Transducers '93)*, Yokohama, Japan, June 1993, 274-277.

[39] G. K. Fedder, J. C. Chang, and R. T. Howe, "Thermal microassembly of narrow-gap electrostatic comb-drive structures," *IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head Island, S.C., June 1992, 63-68.

[40] M. W. Putty and K. Najafi, "A micromachining vibrating ring gyroscope," *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, S.C., June 1994, 213-220.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

1/f NOISE UNIVERSALITY IN HIGH-TECHNOLOGY APPLICATIONS

Peter H. Handel

Department of Physics and Astronomy, University of Missouri - St. Louis
St. Louis, MO 63121

Abstract

Quantum mechanics and electrodynamics require the presence of fundamental quantum fluctuations of physical currents, cross sections, and process rates. The present paper calculates these fluctuations, shows that they have a simple universal 1/f spectrum, and shows how they invariably come to the foreground in high-technology applications, because in them all other fluctuations and sources of instability have been either eliminated, or otherwise discounted and put under control.

I. Introduction

Fluctuations with a spectral density proportional to 1/f are found in a large number of systems in science, technology and everyday life. These fluctuations are known as 1/f noise in general. They have first been noticed by Johnson[1] in early amplifiers, have limited the performance of vacuum tubes in the thirties and forties, and have later hampered the introduction of semiconductor devices.

1/f noise is present as a limitation in most modern high-technology devices. It is present in the resonance frequency of quartz resonators, in SAW devices, in junction and MIS infrared detectors, in SQUIDs, in electron beams, etc. Outside the domain of electrophysics 1/f noise is present in the rate of radioactive decay, in the flow rate of sand in an hourglass, in the flux of cars on an expressway, in the frequency of sunspots, in the light output of quasars, in the flow rate of the Nile over the last 2,000 years, in the water current velocity fluctuations at a depth of 3,100 meters in the Pacific ocean, and in the loudness and pitch fluctuations of classical music. It has also been found below 10^{-8} Hz in the angular velocity of the earth's rotation, and below 10^{-4} Hz in the relativistic neutron flux in the terrestrial atmosphere[2].

The present paper is focused on the general origin of fundamental 1/f noise as a universal form of chaos, and on the cause of its ubiquity. It starts with a *special case* of the general 1/f noise phenomenon, the Quantum 1/f Effect (with its conventinal and coherent contributions) which is as fundamental as time and space. The paper then presents the *general case* of 1/f fluctuations as a necessary consequence of the mathematical homogeneity of the dynamical (or physical) equations describing the motion of an arbitrary chaotic or stochastic nonlinear system. A sufficient criterion is derived, which indicates if an arbitrary system governed by a given system of differential equations will exhibit 1/f noise. The criterion is then applied to several particular systems, and is used to predict the fundamental quantum 1/f effect as a *special case.*

II. Conventinal Quantum 1/f Effect

This effect[3-8] is present in any cross section or process rate involving charged particles or current carriers. The physical origin of quantum 1/f noise is easy to understand. Consider for example Coulomb scattering of current carriers, e.g., electrons on a center of force. The scattered electrons reaching a detector at a given angle away from the direction of the incident beam are described by DeBroglie waves of a frequency corresponding to their energy. However, some of the electrons have lost energy in the scattering process, due to the emission of bremsstrahlung. Therefore, part of the outgoing DeBroglie waves is shifted to slightly lower frequencies. When we calculate the probability density in the scattered beam, we obtain also cross terms, linear both in the part scattered with and without bremsstrahlung. These cross terms oscillate with the same frequency as the frequency of the emitted bremsstrahlung photons. The emission of photons at all frequencies results therefore in probability density fluctuations at all frequencies. The corresponding current density fluctuations are obtained by multiplying the probability density fluctuations by the velocity of the scattered current carriers. Finally, these current fluctuations present in the scattered beam will be noticed at the detector as low frequency current fluctuations, and will be interpreted as fundamental cross section fluctuations in the scattering cross section of the scatterer. While incoming carriers may have been Poisson distributed, the scattered beam will exhibit super-Poissonian statistics, or bunching, due to this new effect which we may call quantum 1/f effect. The quantum 1/f effect is thus a many-body or collective effect, at least a two-particle effect, best described through the two-particle wave function and two-particle correlation function.

Let us estimate the magnitude of the quantum 1/f effect semiclassically by starting with the classical (Larmor) formula $2q^2a^2/3c^3$ for the power radiated by a particle of charge q and acceleration **a**. The acceleration can be approximated by a delta function $\mathbf{a}(t) = \Delta\mathbf{v}\delta(t)$ whose Fourier transform $\Delta\mathbf{v}$ is constant and is the change

0-7803-1945-1/94 $3.00 © 1994 IEEE

in the velocity vector of the particle during the almost instantaneous scattering process. The one-sided spectral density of the emitted bremsstrahlung power $4q^2(\Delta v)^2/3c^3$ is therefore also constant. The number $4q^2(\Delta v)^2/3hfc^3$ of emitted photons per unit frequency interval is obtained by dividing with the energy hf of one photon. The probability amplitude of photon emission $[4q(\Delta v)^2/\ 3hfc^3]^{1/2}e^{i\gamma}$ is given by the square root of this photon number spectrum, including also a phase factor $e^{i\gamma}$. Let ψ be a representative Schrödinger catalogue wave function of the scattered outgoing charged particles, which is a single-particle function, normalized to the actual scattered particle concentration. The beat term in the probability density $\rho=|\psi|^2$ is linear both in this bremsstrahlung amplitude and in the non-bremsstrahlung amplitude. Its spectral density will therefore be given by the product of the squared probability amplitude of photon emission (proportional to 1/f) with the squared non-bremsstrahlung amplitude which is independent of f. The resulting spectral density of fractional probability density fluctuations is obtained by dividing with $|\psi|^4$ and is therefore

$$|\psi|^{-4}S_{|\psi|^2}(f) = 8q^2(\Delta v)^2/3hfNc^3 = 2\alpha A/fN = j^{-2}S_j(f), \qquad (1)$$

where $\alpha = e^2/\hbar c = 1/137$ is the fine structure constant and $\alpha A = 4q^2(\Delta v)^2/3hc^3$ is known as the infrared exponent in quantum field theory, and is known as the quantum 1/f noise coefficient, or Hooge constant, in electrophysics.

The spectral density of current density fluctuations is obtained by multiplying the probability density fluctuation spectrum with the squared velocity of the outgoing particles. When we calculate the spectral density of fractional fluctuations in the scattered current j, the outgoing velocity simplifies, and therefore Eq. (1) also gives the spectrum of current fluctuations $S_j(f)$, as indicated above. The quantum 1/f noise contribution of each carrier is independent, and therefore the quantum 1/f noise from N carriers is N times larger; however, the current j will also be N times larger, and therefore in Eq. (1) a factor N was included in the denominator for the case in which the cross section fluctuation is observed on N carriers simultaneously.

The fundamental fluctuations of cross sections and process rates are reflected in various kinetic coefficients in condensed matter, such as the mobility μ and the diffusion constant D, the surface and bulk recombination speeds s, and recombination times τ, the rate of tunneling j_t and the thermal diffusivity in semiconductors. Therefore, the spectral density of fractional fluctuations in all these coefficients is given also by Eq. (1).

When we apply Eq. (1) to a certain device, we first need to find out which are the cross sections σ or process rates which limit the current I through the device, or which determine any other device parameter P, and then we have to determine both the velocity change Δv of the scattered carriers and the number N of carriers simultaneously used to test each of these cross sections or rates. Then Eq. (1) provides the spectral density of quantum 1/f cross section or rate fluctuations. These spectral densities are multiplied by the squared partial derivative $(\partial I/\partial\sigma)^2$ of the current, or of the device parameter P of interest, to obtain the spectral density of fractional device noise contributions from the cross sections and rates considered. After doing this with all cross sections and process rates, we add the results and bring (factor out) the fine structure constant α as a common factor in front. This yields excellent agreement with the experiment[9] in a large variety of samples, devices and physical systems.

Eq. (1) was derived in second quantization, using the commutation rules for boson field operators. For fermions one repeats the calculation replacing in the derivation the commutators of field operators by anticommutators, which yields[6,7]

$$\rho^{-2}S_\rho(f) = j^{-2}S_j(f) = \sigma^{-2}S_\sigma(f) = 2\alpha A/f(N-1). \qquad (2)$$

This causes no difficulties, since $N\geq 2$ for particle correlations to be defined, and is practically the same as Eq. (10), since usually $N>>1$. Eqs. (1) and (2) suggest a new notion of physical cross sections and process rates which contain 1/f noise, and express a fundamental law of physics, important in most high-technology applications[7].

We conclude that the conventional quantum 1/f effect can be explained in terms of interference beats between the part of the outgoing DeBroglie waves scattered without bremsstrahlung energy losses above the detection limit (given in turn by the reciprocal duration T of the 1/f noise measurement) on one hand, and the various parts scattered with bremsstrahlung energy losses; but there is more to it than that: exchange between identical particles is also important. This, of course, is just one way to describe the reaction of the emitted bremsstrahlung back on the scattered current. This reaction, itself an expression of the nonlinearity introduced by the coupling of the charged-particle field to the electromagnetic field, thus reveals itself as the cause of the quantum 1/f effect, and implies that the effect can not be obtained with an independent boson model. The effect, just like the classical turbulence-generated 1/f noise[12], is a result of the scale-invariant nonlinearity of the equations of motion describing the coupled system of matter and field. Ultimately, therefore, this nonlinearity is the source of the 1/f spectrum in both the classical and quantum form of the theory. We can say that the quantum 1/f effect is an infrared divergence phenomenon, this divergence being the result of the same nonlinearity. The quantum 1/f effect is, in fact, the first time-dependent infrared radiative correction. Finally, it is

also deterministic in the sense of a well determined wave function, once the initial phases γ of all field oscillators are given. In quantum mechanical correspondence with its classical turbulence analog[12], the new effect is therefore a quantum manifestation of classical chaos which we can take as the definition of a certain type of quantum chaos.

III. Derivation of the Coherent Quantum 1/f Effect

The coherent quantum 1/f effect is a quantum fluctuation effect present in any extended current due to the definition of the physical electron as a system composed of the bare particle plus its electomagnetic field which is in a coherent state and has therefore indefinite energy. This in turn causes the state to be non-stationary. The non-stationarity is expressed by the coherent quantum 1/f effect. An elementary physical derivation is given in Sec. VII.

The present derivation is based on the well-known new propagator $G_s(x'-x)$ derived relativistically[13] in 1975 in a new picture required by the infinite range of the Coulomb potential. The corresponding nonrelativistic form[14] was provided by Zhang and Handel:

$$-i<\Phi_o|T\psi_{s'}(x')\psi_s^+(x)|\Phi_o> \equiv \delta_{ss'}\, G_s(x'-x)$$

$$= (i/V)\sum_{\mathbf{p}}\{\exp i[\mathbf{p}(\mathbf{r}-\mathbf{r}')-\mathbf{p}^2(t-t')/2m]/\hbar\}\, n_{\mathbf{p},s}$$

$$\times\{-i\mathbf{p}(\mathbf{r}-\mathbf{r}')/\hbar+i(m^2c^2+\mathbf{p}^2)^{1/2}(t-t')(c/\hbar)\}^{\alpha/\pi}. \qquad (3)$$

Here $\alpha=e^2/\hbar c=1/137$ is Sommerfeld's fine structure constant, $n_{\mathbf{p},s}$ the number of electrons in the state of momentum $\mathbf{p}$ and spin s, m the rest mass of the fermions, $\delta_{ss'}$ the Kronecker symbol, c the speed of light, $x=(\mathbf{r},t)$ any space-time point and V the volume of a normalization box. T is the time-ordering operator which orders the operators in the order of decreasing times from left to right and multiplies the result by $(-1)^P$, where P is the parity of the permutation required to achieve this order. For equal times, T normal-orders the operators, i.e., for t=t' the left-hand side of Eq. (3) is $i<\Phi_o|\psi_s^+(x)\psi_{s'}(x')|\Phi_o>$. The state Φ_o of the N electrons is described by a Slater determinant of single-particle orbitals.

The resulting spectral density coincides with the result $2\alpha/\pi fN$, first derived[10] directly from the coherent state of the electromagnetic field of a physical charged particle. The connection with the conventional quantum 1/f effect was suggested later[11].

To calculate the current autocorrelation function we need the density correlation function, which is also known as the two-particle correlation function. The two-particle correlation function is defined by

$$<\Phi_o|T\psi_s^+(x)\psi_s(x)\psi_{s'}^+(x')\psi_{s'}(x')|\Phi_o> =$$
$$<\Phi_o|\psi_s^+(x)\psi_s(x)|\Phi_o><\Phi_o|\psi_{s'}^+(x')\psi_{s'}(x')|\Phi_o> -$$
$$<\Phi_o|T\psi_{s'}(x')\psi_s^+(x)|\Phi_o><\Phi_o|T\psi_s(x)\psi_{s'}^+(x')|\Phi_o>. \qquad (4)$$

The first term can be expressed in terms of the particle density of spin s, $n/2 = N/2V = <\Phi_o|\psi_s^+(x)\psi_s(x)|\Phi_o>$, while the second term can be expressed in terms of the Green function (1) in the form

$$A_{ss'}(x-x') \equiv <\Phi_o|\psi_s^+(x)\psi_{s'}^+(x')\psi_{s'}(x')\psi_s(x)|\Phi_o> = (n/2)^2 + \delta_{ss'}\, G_s(x'-x)G_s(x-x'). \qquad (5)$$

The "relative" autocorrelation function $A(x-x')$ describing the normalized pair correlation independent of spin is obtained by dividing by n^2 and summing over s and s'

$$A(x-x') = 1 - (1/n^2)\sum_s G_s(x-x')G_s(x'-x) = 1 -$$

$$(1/N^2)\sum_s\sum_{\mathbf{p}\mathbf{p}'}\{\exp i[(\mathbf{p}-\mathbf{p}')(\mathbf{r}-\mathbf{r}')-(\mathbf{p}^2-\mathbf{p}'^2)(t-t')/2m]/\hbar\}\, n_{\mathbf{p},s}n_{\mathbf{p}',s}$$
$$\times\{\mathbf{p}(\mathbf{r}-\mathbf{r}')/\hbar-(m^2c^2+\mathbf{p}^2)^{1/2}(t-t')(c/\hbar)\}^{\alpha/\pi}$$
$$\times\{\mathbf{p}'(\mathbf{r}-\mathbf{r}')/\hbar-(m^2c^2+\mathbf{p}'^2)^{1/2}(t-t')(c/\hbar)\}^{\alpha/\pi}. \qquad (6)$$

Here we have used Eq. (1). We now consider a beam of charged fermions, e.g., electrons, represented in momentum space by a sphere of radius p_F, centered on the momentum $\mathbf{p}_o$ which is the average momentum of the fermions. The energy and momentum differences between terms of different $\mathbf{p}$ are large, leading to rapid oscillations in space and time which contain only high-frequency quantum fluctuations. The low-frequency and low-wavenumber part A_l of this relative density autocorrelation function is given by the terms with $\mathbf{p}=\mathbf{p}'$

$$A_l(x-x') = 1 - (1/N^2)\sum_s\sum_{\mathbf{p}} n_{\mathbf{p},s}$$
$$\times\{\mathbf{p}(\mathbf{r}-\mathbf{r}')/\hbar-(m^2c^2+\mathbf{p}^2)^{1/2}(t-t')(c/\hbar)\}^{2\alpha/\pi} \qquad (7)$$

$$\approx 1 - (1/N)|\mathbf{p}_o(\mathbf{r}-\mathbf{r}')/\hbar-mc^2\tau/\hbar|^{2\alpha/\pi}$$
$$\text{for } p_F<<|p_{o3}-mc^2\tau/z|. \qquad (8)$$

Here we have used the mean value theorem, considering the $2\alpha/\pi$ power as a slowly varying function of $\mathbf{p}$ and neglecting $\mathbf{p}_o$ in the coefficient of $\tau \equiv t-t'$, with $z\equiv|\mathbf{r}-\mathbf{r}'|$. Using the identity[15], with arbitrarily small cutoff ω_o, we obtain from Eq. (8) with $\theta \equiv |\mathbf{p}_o(\mathbf{r}-\mathbf{r}')/\hbar-mc^2\tau/\hbar|$ the exact form

$$A_l(x-x') = 1 + [(2\alpha/\pi N)\int_{\omega_o}^{\infty}(mc^2/\hbar\omega)^{2\alpha/\pi}\cos(\theta\omega)d\omega/\omega]$$
$$\times\{\cos\alpha + (2\alpha/\pi)\sum_{n=0}^{\infty}(\theta\omega_o)^{2n-2\alpha/\pi}[(2n)!(2n-2\alpha/\pi)]^{-1}\}^{-1}. \qquad (9)$$

This indicates a $\omega^{-1-2\alpha/\pi}$ spectrum and a 1/N dependence of the spectrum of fractional fluctuations in density n and current j, if we neglect the curly bracket in the denominator which is very close to unity for very small ω_o. The fractional autocorrelation of current fluctuations δj is obtained by multiplying Eq. (5) on both sides with ep_o/m, and dividing by $(enp_o/m)^2$ which is the square of the average current density j, instead of just dividing by n^2. It is the same as the fractional autocorrelation for quantum density fluctuations. Then Eq. (9) for the coherent Quantum Electrodynamical chaos process in electric currents can be written also in the form

$$S_{\delta j/j}(k) \approx [2\alpha/\pi\omega N][mc^2/\hbar\omega]^{2\alpha/\pi} \approx 2\alpha/\pi\omega N$$

$$= 0.00465/\omega N. \qquad (10)$$

Being observed in the presence of a constant applied field, these fundamental quantum current fluctuations are usually interpreted as mobility fluctuations[3]. Most of the conventional quantum 1/f fluctuations in physical cross sections and process rates are also mobility fluctuations, but some are also in the recombination speed or tunneling rate.

IV. Sufficient Criterion for Fundamental 1/f Noise

In spite of the practical success of our quantum 1/f theory in explaining electronic 1/f noise in most high tech devices, and in spite of the conceptual success of our earlier classical turbulence approach to 1/f noise, the question about the origin of nature's omnipresent 1/f spectra remained unanswered. During the last three decades, we have claimed repeatedly that nonlinearity is a general cause of 1/f noise. The present paper proves that nonlinearity always leads to a 1/f spectrum if homogeneity is also present in the equation(s) of motion. Specifically, if the system is described in terms of the dimensionless vector function $\mathbf{Y}(\mathbf{x},t)$ by the m^{th} order nonlinear differential equation

$$\partial\mathbf{Y}/\partial t + \mathbf{F}(\mathbf{x}, \mathbf{Y}, \partial\mathbf{Y}/\partial x_1 ... \partial\mathbf{Y}/\partial x_n, \partial^2\mathbf{Y}/\partial x_1^2 ..... \partial^m\mathbf{Y}/\partial x_n^m) = 0 \qquad (11)$$

a 1/f spectrum is obtained if the nonlinear function $\mathbf{F}$ satisfies the homogeneity condition

$$\mathbf{F}[\lambda x, \mathbf{Y}, \partial\mathbf{Y}/(\lambda\partial x_1) ... \partial\mathbf{Y}/(\lambda\partial x_n), \partial^2\mathbf{Y}/(\lambda\partial x_1)^2 ..... \partial^m\mathbf{Y}/(\lambda\partial x_n)^m] = \lambda^{-p}\mathbf{F}(x, \mathbf{Y}, \partial\mathbf{Y}/\partial x_1 ... \partial\mathbf{Y}/\partial x_n, \partial^2\mathbf{Y}/\partial x_1^2 ..... \partial^m\mathbf{Y}/\partial x_n^m), \qquad (12)$$

for any real number λ. The order of homogeneity is the number -p. Performing a Fourier transformation of Eq. (10) with respect to the vector $x(x_1, x_2, ...x_n)$, we get in terms of the Fourier-transformed wavevector $\mathbf{k}$ the nonlinear integro-differential equation

$$\partial\mathbf{y}(\mathbf{k},t)/\partial t + \mathbf{G}[\mathbf{k}, \mathbf{y}(\mathbf{k},t), k_1\mathbf{y}(\mathbf{k},t) ... k_n\mathbf{y}(\mathbf{k},t), k_1^2\mathbf{y}(\mathbf{k},t) ..... k_n^m\mathbf{y}(\mathbf{k},t)] = 0, \qquad (13)$$

where $\mathbf{y}(\mathbf{k},t)$ is the Fourier transform of $\mathbf{Y}(\mathbf{x},t)$. Due to Eq. (12), the nonlinear integro-differential operator $\mathbf{G}$ satisfies the relation

$$\mathbf{G}[\lambda\mathbf{k}, \mathbf{y}, \lambda k_1\mathbf{y} ... \lambda k_n\mathbf{y}, (\lambda k_1)^2\mathbf{y} ..... (\lambda k_n)^m\mathbf{y}] = \lambda^p\mathbf{G}[\mathbf{k}, \mathbf{y}, k_1\mathbf{y} ... k_n\mathbf{y}, k_1^2\mathbf{y} ..... k_n^m\mathbf{y}], \qquad (14)$$

where the integration differentials dk, dk', etc., are excepted from replacement with λdk, λdk', etc. Eq. (13) can thus be rewritten in the form

$$d\mathbf{y}/d(t/\lambda^p) + \mathbf{G}[\lambda\mathbf{k}, \mathbf{y}, \lambda k_1\mathbf{y} ... \lambda k_n\mathbf{y}, (\lambda k_1)^2\mathbf{y} ..... (\lambda k_n)^m\mathbf{y}] = 0, \qquad (15)$$

Taking $\lambda = 1/k$, where $k = |\mathbf{k}| = (k_1^2 + ... + k_n^2)^{1/2}$, and setting $k^p t = z$, we notice that k has been eliminated from the dynamical equation, and only $\mathbf{k}/k$ is left. This means that there is no privileged scale left for the system in $\mathbf{x}$ or $\mathbf{k}$ space, other than the scale defined by the given time t, and expressed by the dependence on z. We call this property of the dynamical system "*sliding-scale invariance*".

In certain conditions, instabilities of a solution of Eq. (10) may generate chaos, or turbulence. In a sufficiently large system described by the local dynamical equation (10), in which the boundary conditions become immaterial, homogeneous, isotropic turbulence, (chaos) can be obtained, with a spectral density determined only by Eq. (10). The stationary autocorrelation function $A(\tau)$ is defined as an average scalar product, the average being over the turbulent ensemble

$$A(\tau) = \langle\mathbf{Y}(\mathbf{x},t)\mathbf{Y}(\mathbf{x},t+\tau)\rangle = \int\langle\mathbf{y}(\mathbf{k},t)\mathbf{y}(\mathbf{k},t+\tau)\rangle d^n\mathbf{k}$$

$$= \int u(\mathbf{k},z)d^n\mathbf{k}. \qquad (16)$$

Here we have introduced the scalar

$$u(\mathbf{k},z) = \langle\mathbf{y}(\mathbf{k},t)\mathbf{y}(\mathbf{k},t+\tau)\rangle \qquad (17)$$

of homogeneous, isotropic chaos (turbulence), which depends only on $|\mathbf{k}|$ and $z = k^p\tau$. All integrals are from minus infinity to plus infinity. The chain of integro-differential equations for the correlation functions of any order obeys the same sliding-scale invariance which we have noticed in the fundamental dynamical equation above. *Therefore, in isotropic, homogeneous, conditions, u can only depend on k and z.* Furthermore, the direct dependence on k must reflect this sliding-scale invariance,

and is therefore of the form

$$u(k,z) = k^{-n}v(z). \qquad (18)$$

Indeed, only this form insures that $u(k,z)d^nk$ and therefore also the corresponding integrals and multiple convolutions in k space have the necessary sliding-scale invariance.

According to the Wiener-Khintchine theorem, the spectral density is the Fourier-transform of A(t),

$$S_y(f) = \int e^{2\pi i f\tau} A(\tau)d\tau = (1/f)\int e^{2\pi i t'}\int k'^{-n}v(z)d^nk'dt' = C/f, \qquad (19)$$

where we have set $f\tau=t'$, $k^n=fk'^n$, $z=k^n\tau=k'^nt'$, and the integral

$$C = \int e^{2\pi i t'}\int k'^{-n}v(z)d^nk'dt' = \int e^{2\pi i t'}\int k''^{-n}v(k''^n)d^nk''dt' \qquad (20)$$

is independent of f. We have defined the vector $\mathbf{k}''=t'^{1/n}\,\mathbf{k}$.

The general form of our criterion considers a system described in terms of the integro-differential system of equations

$$\Phi[t, \mathbf{x}, \mathbf{Y}, \partial\mathbf{Y}/\partial t, \partial\mathbf{Y}/\partial x_1 ... \partial\mathbf{Y}/\partial x_n, \partial^2\mathbf{Y}/\partial t^2, \partial^2\mathbf{Y}/\partial x_1^2 ..... \partial^m\mathbf{Y}/\partial x_n^m) = 0 \qquad (21)$$

where the vector function Φ may be nonlinear in any of its arguments. *If a number θ exists such that Eq. (21) implies*

$$\Phi[\lambda^\theta t, \lambda\mathbf{x}, \mathbf{Y}, \partial\mathbf{Y}/\lambda^\theta\partial t, \partial\mathbf{Y}/\lambda\partial x_1 ... \partial\mathbf{Y}/\lambda\partial x_n, \partial^2\mathbf{Y}/\lambda^{2\theta}\partial t^2, \partial^2\mathbf{Y}/\lambda^2\partial x_1^2 ..... \partial^m\mathbf{Y}/\lambda^m\partial x_n^m) = 0 \qquad (22)$$

for any real number λ, the power spectral density of any chaotic solution for the vector function Y defined by Eq. (11) is proportional to 1/f.

Here we have assumed that there are no boundary conditions associated with Eq. (22), or that any boundary conditions included would satisfy the same homogeneity conditions.

In conclusion, nonlinearity + homogeneity = 1/f noise, provided the system is chaotic. The ultimate cause of the ubiquitous 1/f noise in nature is the omnipresence of nonlinearities (no matter how weak) and homogeneity. The latter is finally related to rotational (or Lorentz) invariance and therefore to the isotropy of space (or space-time). All our four specific theories of 1/f chaos in nonlinear systems are just special cases to which this criterion is applicable. They include our magneto-plasma theory of turbulence for current carriers in intrinsic symmetric semiconductors[12] (1966), our similar theory for metals[12] (1971), the quantum 1/f theory[3-11] (pure QED, 1975), and the spectral theory of Musha's highway traffic turbulence results[16] (1989). Applied to the motion of a nonlinearly interacting chain of atoms, it predicts no 1/f spectrum. Starting from a wrong defining equation of the chain, both our criterion and direct calculation allowed for 1/f noise in a special case[16], but the correct defining equation does not fulfill the criterion, and no 1/f spectrum is expected. However, 1/f fluctuations in phonon number, in frequency, and in phase are predicted by the criterion, are derived directly[17] with the quantum 1/f theory, and have been experimentally verified[17,18], in piezoelectric crystals.

V. Application to QED; Quantum 1/f Effect as a Special Case

The nonlinearity causing the 1/f spectrum of turbulence in both semiconductors and metals is caused by the reaction of the field generated by charged particles and their currents back on themselves. The same nonlinearity is present in quantum electrodynamics (QED), where it causes the infrared divergence, the infrared radiative corrections for cross sections and process rates, and the quantum 1/f effect. We shall prove this on the basis of our sufficient criterion for 1/f spectral density in chaotic systems.

Consider a beam of charged particles propagating in a well-defined direction which we shall call the x direction, so that the one-dimensional Schrödinger equation describes the longitudinal fluctuations in the concentration of particles. Considering the non-relativistic case which is encountered in most quantum 1/f noise applications, we write in second quantization the equation of motion for the Heisenberg field operators ψ of the in the form

$$i\hbar\partial\psi/\partial t = (1/2m)[-i\hbar\nabla - (e/c)\,A]^2\psi, \qquad (23)$$

With the non-relativistic form $J = -i\hbar\psi^*\nabla\psi/m$ + hermitian conjugate, and with

$$A(x,y,z,t) = (\hbar/2cmi)\int\frac{[\psi^*\nabla\psi-\psi\nabla\psi^*]}{|x-x'|}dx' \qquad (24)$$

we obtain

$$i\hbar\partial\psi/\partial t = (1/2m)\Big[-i\hbar\nabla - (e\hbar/2c^2mi)\int\frac{[\psi^*\nabla\psi-\psi\nabla\psi^*]}{|x-x'|}dx'\Big]^2\psi. \qquad (25)$$

At very low frequencies or wave numbers the last term in rectangular brackets is dominant on the r.h.s., leading to

$$i\hbar\partial\psi/\partial t = (-1/2m)\left[(e\hbar/2c^2m)\int\frac{[\psi^*\nabla\psi-\psi\nabla\psi^*]}{|x-x'|}dx'\right]^2\psi. \quad (26)$$

For x replaced by λx, and x' replaced by $\lambda x'$, we obtain

$$i\hbar\frac{\partial\psi}{\partial t} = (-1/2m)\left[(e\hbar/2c^2m)\int\frac{[\psi^*\nabla/\lambda\psi-\psi\nabla/\lambda\psi^*]}{\lambda|x-x'|}\lambda^3dx'\right]^2\psi$$
$$= \lambda^2H\psi = \lambda^{-p}H\psi. \quad (27)$$

This satisfies our homogeneity criterion with p=-2. Our sufficient criterion only requires homogeneity, with any value of the weight p, for the existence of a 1/f spectrum in chaos. Therefore, we expect a 1/f spectrum of quantum current-fluctuations, i.e., of cross sections and process rates in physics, as derived in detail earlier[3-8,10-11]. This is in agreement with the well-known, and experimentally verified, results of the Quantum 1/f Theory.

In conclusion, we realize that, both in classical and quantum mechanical nonlinear systems, the limiting behavior at low wave numbers is usually expressed by homogeneous functional dependences, leading to fundamental 1/f spectra on the basis of our criterion.

VI. Derivation of the Conventional Quantum 1/f Effect in Second Quantization

The simplified description of quantum 1/f noise was presented above in the elementary terms of Schrödinger's statistical catalogue model, without using second quantization. This approach is natural in view of the close connection between this new effect and diffraction which is usually treated without second quantization, in the statistical catalogue model based on the single-particle solution of the Schrödinger equation, normalized to the number of particles N. Just as the superposition of elementary phase-shifted waves allows for the simplest and most intuitive description of diffraction through a slit, the description of quantum 1/f noise in terms of interference beats between slightly frequency-shifted scattered partial waves with brems-strahlung energy losses will always provide the simplest and most elementary quantitative derivation of the quantum 1/f effect, easily accessible even at the undergraduate level.

Below we now present the derivation of the Quantum 1/f Effect in a general form which determines the scattered current j from the observation of a sample of N outgoing particles. The minimal outgoing sample for defining particle-particle correlations in the scattered wave consists of two particles, and therefore the effect can be calculated for the case of two outgoing particles.

We start with the expression of the Heisenberg representation state $|S\rangle$ of N identical bosons of mass M emerging at an angle θ from some scattering process with various undetermined bremsstrahlung energy losses reflected in their one-particle waves $\varphi_i(\xi_i)$

$$|S\rangle = (N!)^{-1/2}\,\Pi_i\int d^3\xi_i\varphi_i(\xi_i)\psi^+(\xi_i)\,|0\rangle$$
$$= \Pi_i\int d^3\xi_i\varphi_i(\xi_i)\,|S^o\rangle, \quad (28)$$

where $\psi^+(\xi_i)$ is the field operator creating a boson with position vector ξ_i and $|0\rangle$ is the vacuum state, while $|S^o\rangle$ is the state with N bosons of position vectors ξ_i with i = 1.....N. All products and sums in this Section run from 1 to N, unless otherwise stated.

To calculate the particle density autocorrelation function in the outgoing scattered wave, we need the expectation value of the operator

$$O(x_1,x_2) = \psi^+(x_1)\psi^+(x_2)\psi(x_2)\psi(x_1), \quad (29)$$

known as the operator of the pair correlation. This operator corresponds to a density autocorrelation function. The presence of two-particle coordinates in the operator O does not mean that we are considering two-particle interactions, it only means that the expectation value which we are calculating depends on the relative position of the particles. Using the well known commutation relations for boson field operators

$$\psi(x)\psi^+(y) - \psi^+(y)\psi(x) = \delta(x - y), \quad (30a)$$

$$\psi(x)\psi(y) - \psi(y)\psi(x) = 0, \quad (30b)$$

$$\psi^+(x)\psi^+(y) - \psi^+(y)\psi^+(x) = 0, \quad (30c)$$

we first calculate the matrix element:

$$N!\langle S^o|O|S^o\rangle =$$
$$\Sigma'_{\mu\nu}\Sigma'_{mn}\delta(\eta_\nu-x_1)\delta(\eta_\mu-x_2)\delta(\xi_n-x_1)\delta(\xi_m-x_2)\Sigma_{(i,j)}\Pi'_{ij}\delta(\eta_j-\xi_i), \quad (31)$$

where $|S^o\rangle$ is the state with well defined particle coordinates. Here the prime excludes $\mu=\nu$ and m=n in the summations and excludes i=m, i=n, j=μ and j=ν in the product. The summation $\Sigma_{(i,j)}$ runs over all permutations of the remaining N-2 values of i and j. On the basis of this result we now calculate the complete matrix element

$$\langle S|O|S\rangle = [1/N(N-1)]\,\Sigma'_{\mu\nu}\Sigma'_{mn}\int d^3\eta_\mu\int d^3\eta_\nu\int d^3\xi_m\int d^3\xi_n$$

$$\times\varphi_\mu^*(\eta_\mu)\varphi_\nu^*(\eta_\nu)\varphi_m(\xi_m)\varphi_n(\xi_n)\delta(\eta_\nu-x_1)\delta(\eta_\mu-x_2)\delta(\xi_n-x_1)\delta(\xi_m-x_2) = [1/N(N-1)]$$

$$\times\Sigma'_{\mu\nu}\Sigma'_{mn}\varphi_\mu^*(x_2)\varphi_\nu^*(x_1)\varphi_m(x_1)\varphi_n(x_2). \quad (32)$$

The one-particle states are spherical waves emerging from the scattering center located at x=0:

$$\varphi(x) = (C/x)e^{iKx}\,[1 + \Sigma_{kl}b(k,l)e^{-iqx}\,a^+_{k,l}]. \quad (33)$$

Here C is an amplitude factor, K the boson wave vector magnitude, b(**k**,l) the bremsstrahlung amplitude for emission of photons of wave vector **k** and polarization l, while $a^+_{\mathbf{k},l}$ is the corresponding photon creation operator, allowing the emitted photon state to be created from the vacuum if Eq. (33) is inserted into Eq. (32). The momentum magnitude loss $\hbar q = Mc\mathbf{k}/K \equiv 2\pi Mf/K$ is necessary for energy conservation in the Bremsstrahlung process. The reader can derive this expression of $q=\delta p$ from the canonical equation $v = \partial H/\partial p$ which yields $\delta p = \delta H/v = M\delta H/p = M\delta H/\hbar K$. Substituting Eq. (33) into Eq. (32), we obtain

$$<S|O|S> = |C/x|^4\{N(N-1) + 2(N-1)\Sigma_{\mathbf{k},l}|b(\mathbf{k},l)|^2[1+\cos q(x_1-x_2)]\}, \quad (34)$$

where we neglected a small term of higher order in b(**k**,l). To perform the angular part of the summation in Eq. (34), we calculate the current expectation value of the state in Eq. (33), and compare it to the well known cross section without and with bremsstrahlung

$$\mathbf{j} = (\hbar K/Mx^2)[1 + \Sigma_{kl}|b(k,l)|^2] = \mathbf{j}_o[1 + \alpha A\int df/f], \quad (35)$$

where the quantum fluctuations have disappeared, $\alpha = e^2/\hbar c$ is the fine structure constant, $\alpha A = (2\alpha/3\pi)(\Delta v/c)^2$ is the fractional bremsstrahlung rate coefficient, also known in QED as the infrared exponent, and where the 1/f dependence of the bremsstrahlung part displays the well-known infrared catastrophe, i.e., the emission of a logarithmically divergent number of photons in the low frequency limit. Here Δv is the velocity change $\hbar(\mathbf{K}-\mathbf{K}_o)/M$ of the scattered boson, and $f = ck/2\pi$ the photon frequency. Eq. (34) thus gives

$$<S|O|S> = |C/x|^4\{N(N-1) + 2(N-1)\alpha A\int[1+\cos q(x_1-x_2)]df/f\}, \quad (36)$$

which is the pair correlation function, or density autocorrelation function along the scattered beam with $df/f = dq/q$. The spatial distribution fluctuations along the scattered beam will also be observed as fluctuations in time at the detector, at any frequency f. According to the Wiener-Khintchine theorem, we obtain the spectral density of fractional scattered particle density ρ, (or current j, or cross section σ) fluctuations in terms of frequency f or wave number q by dividing the coefficient of the cosine by the constant term N(N-1):

$$\rho^{-2}S_\rho(f) = j^{-2}S_j(f) = \sigma^{-2}S_\sigma(f) = 2\alpha A/fN, \quad (37)$$

where N is the number of particles or current carriers used to define the current j whose fluctuations we are studying. Quantum 1/f noise is thus a fundamental 1/N effect. The exact value of the exponent of f in Eq. (37) can be determined by including the contributions from all real and virtual multiphoton processes of any order (infrared radiative corrections), and turns out to be αA-1, rather than -1, which is important only philosophically, since $\alpha A \ll 1$. The spectral integral is thus convergent at f=0.

For fermions we repeat the calculation replacing in the derivation of Eq. (34) the commutators of field operators by anticommutators, which finally yields in the same way

$$\rho^{-2}S_\rho(f) = j^{-2}S_j(f) = \sigma^{-2}S_\sigma(f) = 2\alpha A/f(N-1), \quad (38)$$

The denominator causes no difficulties, since $N \geq 2$ for particle correlations to be defined, and which is practically the same as Eq. (37), since usually $N \gg 1$. Eqs. (37) and (38) suggest a new notion of physical cross sections and process rates which contain 1/f noise, and express a fundamental law of physics, important in most high-technology applications[9].

We conclude that the conventional quantum 1/f effect can be explained in terms of interference beats between the part of the outgoing DeBroglie waves scattered without bremsstrahlung energy losses above the detection limit (given in turn by the reciprocal duration T of the 1/f noise measurement) on one hand, and the various parts scattered with bremsstrahlung energy losses; but there is more to it than that: exchange between identical particles is also important. This, of course, is just one way to describe the reaction of the emitted bremsstrahlung back on the scattered current. This reaction thus reveals itself as the cause of the quantum 1/f effect, and implies that the effect can not be obtained with the independent boson model. The effect, just like the classical turbulence-generated 1/f noise[12], is a result of the scale-invariant nonlinearity of the equations of motion describing the coupled system of matter and field. Ultimately, therefore, this nonlinearity is the source of the 1/f spectrum in both the classical and quantum form of the author's theory. We can say that the quantum 1/f effect is an infrared divergence phenomenon, this divergence being the result of the same nonlinearity. The new effect is, in fact, the first time-dependent infrared radiative correction. Finally, it is also deterministic in the sense of a well determined wave function, once the initial phases γ of all field oscillators are given. In quantum mechanical correspondence with its classical turbulence analog, the new effect is therefore a quantum manifestation of classical chaos which we can take as the definition of a certain type of quantum chaos.

We turn now to the connection to the coherent Quantum 1/f Effect. The coherent state in a conductor or semiconductor sample is the result of the experimental efforts directed towards establishing a steady and constant current, and is therefore the state defined by the collective motion, i.e. by the drift of the current carriers. It is expressed in the Hamiltonian by the magnetic energy E_m, per unit length, of the current carried by the sample. In very small samples or electronic devices, this magnetic

energy

$$E_m = \int (B^2/8\pi) d^3x = [nevS/c]^2 \ln(R/r) \qquad (39)$$

is much smaller than the total kinetic energy E_k of the drift motion of the individual carriers

$$E_k = \sum mv^2/2 = nSmv^2/2 = E_m/s. \qquad (40)$$

Here we have introduced the magnetic field B, the carrier concentration n, the cross sectional area S and radius r of the cylindrical sample (e.g., a current carrying wire), the radius R of the electric circuit, and the "coherece ratio"

$$s = E_m/E_k = 2ne^2S/mc^2 \ln (R/r) \approx 2e^2N'/mc^2, \qquad (41)$$

where N' = nS is the number of carriers per unit length of the sample and the natural logarithm ln(R/r) has been approximated by one in the last form. We expect the observed spectral density of the mobility fluctuations to be given by a relation of the form

$$(1/\mu^2)S_\mu(f) = [1/(1+s)][2\alpha A/fN] + [s/(1+s)][2\alpha/\pi fN] \quad (42)$$

which can be interpreted as an expression of the effective Hooge constant if the number N of carriers in the (homogeneous) sample is brought to the numerator of the left hand side. In this equation $\alpha A = 2\alpha(\Delta v/c)^2/3\pi$ is the usual nonrelativistic expression of the infrared exponent, present in the familiar form of the conventional quantum 1/f effect[3-8]. This equation is limited to quantum 1/f mobility (or diffusion) fluctuations, and does not include the quantum 1/f noise in the surface and bulk recombination cross sections, in the surface and bulk trapping centers, in tunneling and injection processes, in emission or in transitions between two solids.

Note that the coherence ratio s introduced here equals the unity for the critical value N' = N" = $2 \cdot 10^{12}$/cm., e.g. for a cross section S = $2 \cdot 10^{-4}$ cm^2 of the sample when n = 10^{16}. For small samples with N'<<N" only the first term survives, while for N'>N" the second term in Eq. (42) is dominant.

VII. Physical Derivation of the Coherent Quantum 1/f Effect

This effect arises in a beam of electrons (or other charged particles propagating freely in vacuum) from the definition of the physical electron as a bare particle plus a coherent state of the electromagnetic field. It is caused by the energy spread characterizing any coherent state of the electromagnetic field oscillators, an energy spread which spells non-stationarity, i.e., fluctuations. To find the spectral density of these inescapable fluctuations which are known to characterize any quantum state which is not an energy eigenstate, we use an elementary physical derivation based on Schrödinger's definition of coherent states, which supplements the rigorous derivation which was given in Sec. III from a well-known quantum-electrodynamical propagator. The chaotic character of these fluctuations was discussed in Sec. V.

The coherent quantum 1/f effect will be derived in three steps: first we consider a hypothetical world with just one single mode of the electromagnetic field coupled to a beam of charged particles; considering the mode to be in a coherent state, we calculate the autocorrelation function of the quantum fluctuations in the particle-density (or concentration) which arise from the nonstationarity of the coherent state. Then we calculate the amplitude with which this one mode is represented in the field of an electron, according to electrodynamics. Finally, we take the product of the autocorrelation functions calculated for all modes with the amplitudes found in the previous step.

Let a mode of the electromagnetic field be characterized by the wave vector **q**, the angular frequency $\omega = cq$ and the polarization λ. Denoting the variables q and λ simply by q in the labels of the states, we write the coherent state of amplitude $|z_q|$ and phase arg z_q in the form

$$|z_q\rangle = \exp[-(1/2)|z_q|^2] \exp[z_q a_q^+] |0\rangle$$
$$= \exp[-(1/2)|z_q|^2] \sum_{n=0}^{\infty} (z_q^n)/n! \; |n\rangle. \qquad (43)$$

Here a_q^+ is the creation operator which adds one energy quantum to the energy of the mode. Let us use a representation of the energy eigenstates in terms of Hermite polynomials $H_n(x)$

$$|n\rangle = (2^n n! \sqrt{\pi})^{-1/2} \exp[-x^2/2] \, H_n(x) \, e^{in\omega t}. \quad (44)$$

This yields for the coherent state $|z_q\rangle$ the representation

$$\Psi_q(x) = \exp[-(1/2)|z_q|^2] \exp[-x^2/2] \sum_{n=0}^{\infty} \frac{[z_q e^{i\omega t}]^n}{[n!(2^n\sqrt{\omega})]^{1/2}} H_n(x)$$

$$= \exp[-|z_q|^2/2] \exp[-x^2/2] \exp[-z_q^2 e^{-2i\omega t} + 2xz_q e^{i\omega t}]. \quad (45)$$

In the last form the generating function of the Hermite polynomials was used. The corresponding autocorrelation function of the probability density function, obtained by averaging over the time t or the phase of z_q, is, for $|z_q|<<1$,

$$P_q(\tau,x) = \langle |\Psi_q|_t^2 \, | \, \Psi_q|_{t+\tau}^2 \rangle$$
$$= \{1 + 8x^2|z_q|^2[1 + \cos \omega\tau] - 2|z_q|^2\} \exp[-x^2/2]. \quad (46)$$

Integrating over x from $-\infty$ to ∞, we find the autocorrelation function

$$A^1(\tau) = (2)^{-1/2}\{1 + 2|z_q|^2\cos\omega\tau\}. \quad (47)$$

This result shows that the probability distribution contains a constant background with small superposed oscillations of frequency ω. Physically, the small oscillations in the total probability describe self-organization or bunching of the particles in the beam. They are thus more likely to be found in a measurement at a certain time and place than at other times and places relative to each other along the beam. Note that for $z_q = 0$ the coherent state becomes the ground state of the oscillator which is also an energy eigenstate, and therefore stationary and free of oscillations.

We now determine the amplitude z_q with which the field mode q is represented in the physical electron. One way to do this is to let a bare particle dress itself through its interaction with the electromagnetic field, i.e. by performing first order perturbation theory with the interaction Hamiltonian

$$H' = A_\mu j^\mu = -(e/c)\mathbf{v}\cdot\mathbf{A} + e\phi, \quad (48)$$

where $\mathbf{A}$ is the vector potential and ϕ the scalar electric potential. Another way is to Fourier expand the electric potential e/r of a charged particle in a box of volume V. In both ways we obtain the well-known result

$$|z_q|^2 = \pi(e/q)^2(\hbar c q V)^{-1}. \quad (49)$$

Considering now all modes of the electromagnetic field, we obtain from the single - mode result of Eq. (47)

$$A(\tau) = C\,\Pi_q\{1 + 2|z_q|^2\cos\omega_q\tau\} = C\{1 + \Sigma_q\, 2|z_q|^2\cos\omega_q\tau\}$$

$$= C\{1 + 4(V/2^3\pi^3)\int d^3q\, |z_q|^2\cos\omega_q\tau\} \quad (50)$$

Here we have again used the smallness of z_q and we have introduced a constant C. Using Eq. (49) we obtain

$$A(\tau) = C\{1 + 4\pi(V/2^3\pi^3)(4\pi/V)(e^2/\hbar c)\int(dq/q)\cos\omega_q\tau\}$$

$$= C\{1 + 2(\alpha/\pi)\int\cos(\omega\tau)d\omega/\omega\}. \quad (51)$$

Here $\alpha = e^2/\hbar c$ is Sommerfeld's fine structure constant 1/137. The first term in curly brackets is unity and represents the constant background, or the d.c. part of the current carried by the beam of particles through vacuum. The autocorrelation function for the relative (fractional) density fluctuations, or for the current density fluctuations in the beam of charged particles is obtained therefore by dividing the second term in curly brackets by the first term. The constant C drops out when the fractional fluctuations are considered. According to the Wiener-Khintchine theorem, the coefficient of $\cos\omega\tau$ is the spectral density of the fluctuations, $S_{|\psi|^2}$ for the particle concentration, or S_j for the current density $j = e(k/m)|\psi|^2$

$$S_{|\psi|^2}\langle|\psi|^{-2}\rangle = S_j\langle j\rangle^{-2} = 2(\alpha/\pi fN) = 4.6\cdot10^{-3}\, f^{-1}N^{-1}. \quad (52)$$

Here we have included the total number N of charged particles which are observed simultaneously in the denominator, because the noise contributions from each particle are independent. This result is related to the conventional Quantum 1/f Effect considered in the next section. A similar calculation yields the gravidynamical quantum 1/f effect (QGD 1/f effect) by substituting gravitons for the photons considered so far as infraquanta.

VIII. Derivation of Mobility Quantum 1/f Noise in n+-p Diodes

For a diffusion limited n^+-p junction the current is controlled by diffusion of electrons into the p - region over a distance of the order of the diffusion length $L = (D_n\tau_n)^{1/2}$ which is shorter than the length w_p of the p - region in the case of a long diode. If N(x) is the number of electrons per unit length and D_n their diffusion constant, the electron current at x is

$$I_{nd} = -eD_n dN/dx, \quad (53)$$

where we have assumed a planar junction and taken the origin x = 0 in the junction plane. Diffusion constant fluctuations, given by kT/e times the mobility fluctuations, will lead to local current fluctuations in the interval Δx

$$\delta\Delta I_{nd}(x,t) = I_{nd}\Delta x\delta D_n(x,t)/D_n. \quad (54)$$

The normalized weight with which these local fluctuations representative of the interval Δx contribute to the total current I_d through the diode at x = 0 is determined by the appropriate Green function and can be shown to be $(1/L)\exp(-x/L)$ for $w_p/L >> 1$. Therefore the contribution of the section Δx is

$$\delta\Delta I_d(x,t) = (\Delta x/L)\exp(-x/L)I_{nd}\delta D_n(x,t)/D_n, \quad (55)$$

with the spectral density

$$S_{\Delta Id}(x,f) = (\Delta x/L)^2\exp(-2x/L)\, I_{nd}^2\, S_{Dn}(x,f)/D_n^2. \quad (56)$$

For mobility and diffusion fluctuations the fractional spectral density is given by $\alpha_{Hnd}/fN\Delta x$, where α_{Hnd} is determined from quantum 1/f theory according to Secs II and VI. With Eq. (53) we obtain then

$$S_{\Delta Id}(x,f) = (\Delta x/L^2)\exp(-2x/L)(eD_n dN/dx)^2\,\alpha_{Hnd}/fN. \quad (57)$$

The electrons are distributed according to the solution of the diffusion equation, i.e.

$$N(x) = [N(0) - N_p] \exp(-x/L);$$
$$dN/dx = -\{[N(0) - N_p]/L\} \exp(-x/L). \quad (58)$$

Substituting into Eq. (57) and simply summing over the uncorrelated contributions of all intervals Δx, we obtain

$$S_{Id}(f) = \alpha_{Hnd}(eD_n/L^2)^2 \int_0^{w_p} [N(0) - N_p]^2 e^{-4x/L} dx / \{[N(0) - N_p] e^{-x/L} + N_p\}. \quad (59)$$

We note that $eD_n/L^2 = e/\tau_n$. With the expression of the saturation current $I_0 = e(D_n/\tau_n)^{1/2}N_p$ and of the current $I = I_0[\exp(eV/kT) - 1]$, we can carry out the integration

$$S_{Id}(f) = \alpha_{Hnd}(eI/f\tau_n) \int_0^1 a^2u^3du/(au + 1)$$
$$= \alpha_{Hnd}(eI/f\tau_n)\, F(a). \quad (60)$$

Here we have introduced the notations

$$u = \exp(-x/L), \quad a = \exp(eV/kT) - 1,$$

$$F(a) = 1/3 - 1/2a + 1/a^2 - (1/a^3)\ln(1+a). \quad (61)$$

Eq. (60), obtained by van der Ziel and Anderson, gives the diffusion noise as a function of the quantum 1/f noise parameter αH_{nd}. A similar result can be derived for the quantum 1/f fluctuations of the recombination rate r in the bulk of the p - region, the only difference being the presence of α_{Hnr} istead of α_{Hnd} in Eq. (60). The total noise is the given by Eq. (60) with α_{Hnd} replaced by the sum $\alpha_{Hnd} + \alpha_{Hnr}$

$$S_{Id}(f) = (\alpha_{Hnd} + \alpha_{Hnr})(eI/f\tau_n)\, F(a). \quad (60')$$

IX. Review of Quantum 1/f Noise in n±p Junctions

As we have seen in the Sec. II, quantum 1/f noise is a low-frequency fluctuation process present in elementary cross sections σ and process rates Γ, given by the fractional spectral density

$$S_\sigma(f)/\sigma 2 = S_\Gamma(f)/\Gamma^2 = (4\alpha/3\pi fN)(\Delta v/c)^2 \quad (62)$$

for conventional quantum 1/f noise which is applicable to small devices, and

$$S_\sigma(f)/\sigma^2 = S_\Gamma(f)/G^2 = (2\alpha/\pi fN) = \alpha_c/fN \quad (63)$$

for coherent state quantum 1/f noise which is applicable to large devices in which the energy of the carrier drift motion is predominantly magnetic, rather than kinetic. Here Δv is the velocity change of the carriers in the processes considered, $\alpha = 1/137$ is the fine structure constant, N the number of carriers simultaneously interrogating the cross sections or process rates, $\alpha_c = 4.6\ 10^{-3}$ the coherent quantum 1/f noise coefficient, and c the speed of light. The two forms of quantum 1/f noise are closely related infrared divergence phenomena which arise due to the interaction of electrons and soft photons.

Both in n^+p diodes[9] and metal-insulator-semiconductor (MIS) devices the current will be determined by cross sections s_i such as recombination and scattering cross sections (by phonons and lattice defects), as well as by other process rates (e.g., band to band and trap-assisted tunneling, particularly important in long-wavelength $Hg_{1-x}Cd_xTe$ detectors). The spectral density of the resulting quantum 1/f current fluctuations is therefore

$$S_I(f) = \Sigma_i\, [\partial I/\partial\sigma_i]^2 S_{\sigma i}(f) + \Sigma\, [\partial I/\partial\Gamma_i]^2 S_{\Gamma i}(f), \quad (64)$$

where the spectral densities on the right hand side will be given by Eqs. (62) or (63), depending on the size of the device.

The current density I of Eq. (64) contains a diffusion term I_d, a term I_r caused by recombination in the space charge region, a surface recombination term I_s, a tunneling term I_t and a photovoltaic term caused by the creation of electron hole pairs by photons:

$$I = I_d + I_r + I_s + I_t + q\eta\Phi$$

$$= qn_i\{(n_i/n_0)(D_n/\tau_n)^{1/2}(e^{qV/kT}-1) + (W/\tau)(e^{qV/2kT}-1) + s\} + I_t + q\eta\Phi. \quad (65)$$

Here n_i is the intrinsic concentration, n_0 the concentration of acceptors on the p side, D_n and τ_n the diffusion constant and lifetime of minority carriers on the p side, W the width of the depletion region, $t = \tau_{po} + \tau_{no}$ the Shockley-Hall-Read lifetime, V the applied voltage, s the surface recombination speed, η the quantum efficiency and Φ the incident flux of photons. With the exception of the last term, the terms in Eq. (65) are known as dark current components.

The first term in curly brackets in Eq. (65) gives the diffusion current density I_d, and yields a noise term[9] as shown in Eq. (60')

$$S_{Id}(f) = (\alpha_d + \alpha_r)(qI_d/f\tau_n)F(a);$$
$$\alpha_d = (4\alpha/3\pi)(h/m^*bc)^2\exp(-\theta/2T) \quad (66)$$

in Eq. (64) for small devices, and

$$S_{Id}(f) = \alpha_c(qI_d/f\tau_n)F(a) \qquad (67)$$

for large devices, with F(a) = -1/3 -1/2a +1/a^2 -(1/a^3)ln(1+a) and a = exp(qV/kT) -1. Here b is the lattice constant, m^* the effective mass of the electrons, and q the Debye temperature. These results are obtained by adding the scattering and recombination quantum 1/f cross section fluctuation spectra for all points on the p side of a long n^+p diode with the appropriate Green function weight $[\partial I/\partial\sigma_i]^2$, as prescribed by Eq. (64). The spatial dependence exp(-x/L_d) of this Green function is given by the diffusion equation applied to electrons on the p side, with L_d being the electron diffusion length. If the length w_p of the p side is shorter than L_d, we have to replace F(a) by F(a) - F[a exp(-w_p/L_d)] in Eqs. (66) - (67), and also to include the quantum 1/f noise of the recombination cross sections at the p contact.

Similarly we obtain[9]

$$S_{Ir} = \alpha_r qI_r \tanh(qV/2kT)/f\tau$$
$$= (4\alpha/3\pi c^2)[2q(V_{diff} - V) + 3kT]\{(m_n^*)^{1/2} + (m_p^*)^{1/2}\}^{-2}. \qquad (68)$$

The corresponding expression for the surface recombination current is given by Eq. (68) with half of the surface potential jump U added to the diffusion potential V_{diff}. The tunneling current I_t and its noise is the same as in MIS devices,and will be discussed below. In the final form of Eq. (64) the fine structure constant α appears as a general factor.

X. Quantum 1/f Noise in Squids

As we have seen in Sec. II and VI above, any cross section or process rate defined for electrically charged particles must fluctuate in time with a 1/f spectral density according to quantum electrodynamics, as a consequence of infrared-divergent coupling to low-frequency photons. This fundamental effect leads to quantum 1/f noise observed in many systems with a small number of carriers, and is also present in the cross sections and process rates which determine the resistance and tunneling rate in Josephson junctions, providing a lower limit of the observed 1/f noise.

In a Josephson junction the normal resistance R_n of the barrier is proportional to a scattering cross section or transition rate experienced by the electron in quasiparticle tunneling and by the Cooper pairs below the critical current I_c. Therefore

$$R_n^{-2}S_{Rn}(f) = (4\alpha/3\pi)\,[(\Delta v)^2/c^2Nf)$$
$$= (8\alpha/3\pi)\,(v_F^2/c^2Nf) = 4\ 10^{-14}/f\Omega, \qquad (69)$$

where we have approximated $(\Delta v)^2$ with $2v_F^2$, v_F being the Fermi velocity, and the number of carriers N simultaneously present in the barrier of volume Ω (in m^3) by 10^7, for barriers wider than 10^{-7}cm.

Assuming a linear relationship between the critical current I_c and $G_n=R_n^{-1}$, we obtain , from Eq. (69) for R_n, the spectral density of voltage fluctuations

$$S_V(f) = (4/f)\ 10^{-12}(T/3K)[R_S g(V)/(R_S+R_J)]^2$$
$$\times[I_c R_n(I^2/I_c^2 - 1)^{-1/2} + g(V)V]^2\Omega^{-1}, \qquad (70)$$

where $R_J(V)$ is the junction resistance, R_S the shunt resistance, and $g(V)=R_n/R_J$.

The noise caused in a SQUID by the source considered above can be obtained as the sum of the noise contributions from the two junctions.

The above quantum 1/f results of Eqs.(69) and (70) are in good quantitative agreement with the experimental data.

In conclusion, the fundamental quantum 1/f fluctuations of the cross sections and transition rates which determine the normal resistance have been evaluated in this Section for the case of a Josephson junction. Considering the velocity change in the quantum 1/f formula equal to twice the Fermi velocity and the concentration of carriers in the barrier $10^{19}cm^{-3}$, a spectral density of fractional fluctuations in the normal resistance of the barrier of 4 10^{-14}/f was obtained for a Josephson junction with a volume of the barrier of $10^{-12}cm^3$. These fluctuations are inversely proportional to the barrier volume and result in voltage fluctuations both directly and through the dependence of the critical current on the normal resistance, in good agreement with the experimental data.

XI. Quantum 1/f Effect in Frequency Standards

Frequency standards contain a main resonant mode which can be described as a harmonic oscillator with losses

$$dx/dt + \gamma dx/dt + \omega_o^2 x = F(t). \qquad (71)$$

The quantum 1/f fluctuations are present in the loss coefficient γ. They are given by an expression of the form

$$S_{\delta\gamma/\gamma}(f) = \Lambda/f, \qquad (72)$$

where Λ is a quantum 1/f coefficient characterizing the elementary loss process.

The resonance frequency is given by

$$\omega_r^2 = \omega_o^2 + \gamma^2. \quad (73)$$

The quantum 1/f fluctuations in the resonance frequency are given by $\omega_r\delta\omega_r = -2\gamma\delta\gamma$, or

$$\delta\omega_r/\omega_r = -2(\gamma/\omega_r)^2\delta\gamma/\gamma = -(1/2Q^2)(\delta\gamma/\gamma), \quad (74)$$

where $Q=\omega_r/2\gamma$ is the quality factor. Squaring, averaging and particularizing Eq. (74) for the unit frequency interval, we obtain the Q^{-4} law[21]

$$S_y(f) = <(\delta\omega_r/\omega_r)^2>_f = (1/4Q^4)<(\delta\gamma/\gamma)^2>_f = \Lambda/4fQ^4, \quad (75)$$

where y is the fractional frequency fluctuation $\delta\omega_r/\omega_r$. This law was found in approximative form empirically by Gagnepain and Uebersfeld for the case of quartz resonators. The coefficient Λ is calculated for an ideal quartz crystal as follows.

According to the general quantum 1/f formula in Eqs. (2) and (62), $\Gamma^{-2}S_\Gamma(f)=2\alpha A/f$, where $A=2(\Delta J/ec)^2/3\pi$ is the quantum 1/f effect in any physical process rate Γ. Setting $J=dP/dt=\dot{P}$ where **P** is the vector of the dipole moment of the quartz crystal, we obtain for the fluctuations in the rate Γ of phonon removal from the main resonator oscillation mode (by scattering on a phonon from any other mode of average frequency $<\omega>$) of the crystal, (or via a two-phonon-process at a crystal defect or impurity, involving a phonon of average frequency $<\omega'>$) the spectral density

$$S_\Gamma(f) = \Gamma^2 4\alpha(\Delta\dot{P})^2/3\pi e^2c^2, \quad (76)$$

where $(\Delta\dot{P})^2$ is the square of the dipole moment rate change associated with the process causing the removal of a phonon from the main oscillator mode. To calculate it, we write the energy W of the interacting resonator mode $<\omega>$ in the form

$$W = n\hbar<\omega> = 2(Nm/2)(dx/dt)^2$$
$$=(Nm/e^2)(e\,dx/dt)^2=(m/Ne^2)\varepsilon^2(\dot{P})^2; \quad (77)$$

The factor two includes the potential energy contribution. Here m is the reduced mass of the elementary oscillating dipoles, e their charge, ε a polarization costant, and N their number in the quartz crystal. Applying a variation $\Delta n=1$ we get

$$\Delta n/n = 2|\Delta\dot{P}|/|\dot{P}|, \text{ or } \Delta\dot{P}=\dot{P}/2n. \quad (78)$$

Solving Eq. (77) for $\dot{P}$ and substituting, we obtain

$$|\Delta\dot{P}| = (N\hbar<\omega>/n)^{1/2}(e/2\varepsilon) \quad (79)$$

Substituting $\Delta\dot{P}$ into Eq. (76), we get

$$\Gamma^{-2}S_\Gamma(f) = N\alpha\hbar<\omega>/3n\pi mc^2f\varepsilon^2 \equiv \Lambda/f. \quad (80)$$

This result is applicable to the fluctuations in the loss rate Γ of the quartz.

The corresponding resonance frequency fluctuations of the quartz resonator are given by

$$\omega^{-2}S_\omega(f) = (1/4Q)(\Lambda/f) = N\alpha\hbar<\omega>/12n\pi mc^2f\varepsilon^2Q^4, \quad (81)$$

where Q is the quality factor of the single-mode quartz resonator considered, and $<\omega>$ is not the circular frequency of the main resonator mode, ω_o, but rather the practically constant frequency of the average interacting phonon, considering both three-phonon and two-phonon processes. The corresponding $\Delta\dot{P}$ in the main resonator mode has to be also included in principle, but is negligible because of the very large number of phonons present in the main resonator mode.

Eq. (81) can be written in the form

$$S(f) = \beta V/fQ^4, \quad (82)$$

where, with an intermediary value $<\omega>=10^8/s$, with $n=kT/\hbar<\omega>$, T=300K and $kT=4\ 10^{14}$,

$$\beta = (N/V)\alpha\hbar<\omega>/12n\pi\varepsilon^2mc^2$$

$$=10^{22}(1/137)(10^{-27}10^8)^2/12kT\pi10^{-27}\,9\,10^{20} = 1. \quad (83)$$

The form of Eq. (81) shows that the level of 1/f frequency noise depends not only as Q^{-4} as previously proposed[21], but also on the oscillation frequency or the volume of the active region[17].

The case of other frequency standards is studied by the paper of Handel and Walls in this volume.

XII. Discussion

The derivations of conventional and coherent quantum 1/f noise in Sec. II and III correspond to different physical situations. These two situations have been discussed on the first page of the 1966 turbulence paper[12], at the beginning of this long journey which led us from the classical hydromagnetic or plasma turbulence to quantum 1/f noise and the general sufficient criterion. The discussion of these two situations was repeated identically[10] for the quantized form of our turbulence theory, i.e., for the two related quantum 1/f effects in 1985. It shows us that conventional quantum 1/f noise is observed in small samples, for which most of the drift

energy of the current carriers is included in the sum of their individual kinetic energies $mv^2/2$. For larger samples, and larger values[20] of the parameter s measuring this proportion numerically, most of the drift energy of the carriers is in their collective magnetic energy $LI^2/2$. The transition between the two situations is given by a physical interpolation formula[11], and is the focus a present research effort discussed by Handel and Zhang[20].

Our criterion shows how homogeneity provides the ingredient leading from nonlinearity to 1/f noise. Physically, the homogeneity is required both by the physical requirement of dimensional homogeneity of terms in the equations of physics, and by the invariance of the three-dimensional space with respect to rotations, i.e., by the isotropy of space, which requires x_1, x_2, and x_3 to enter in the same way into the basic laws of nature. In general, we conclude that the ubiquity of the 1/f spectrum is caused by the omnipresence of nonlinearities, no matter how small, and by the simultaneous requirement of rotational and Lorentz invariance which shape the world of classical and relativistic physics respectively. In general, we conclude that ontologically, i.e., from the construction of our world with quarks and leptons, quantum 1/f noise theory gives the cause of fundamental 1/f noise, while epistemologically, i.e., in the world of general notions, the combination of nonlinearity and homogeneity required by our general sufficient criterion is the ultimate cause of all fundamental 1/f noise, including the ontlogically primordial quantum 1/f noise as a special case. Mathematically, this happens in all fundamental 1/f spectra on the basis of the idempotence of $1/f^{1-\varepsilon}$ with respect to convolutions in the limit $\varepsilon \to 0$, with $\varepsilon = \alpha A$ in the case of quantum 1/f noise. In practice, however, the idempotent property of 1/f does not allow us to distinguish which systems will show 1/f fluctuations, while the general sufficient criterion, first presented at the Symposium on 1/f Nose and Chaos In Tokyo, March 1991, allows us to easily recognize the systems which generate 1/f spectra, if their mathematical definition is given in terms of a dynamical system of nonlinear integro-differential equations, or in simpler terms.

We note that our sufficient criterion explains the ubiquity of 1/f noise through a homogeneity which can be established sometimes even without knowing the exact form of the dynamical equation(s) governing a nonlinear system. The derivation of the criterion shows that it is obviously connected with (actually based on) the idempotent property of the 1/f spectrum with respect to the convolution operation. Therefore the 1/f spectrum corresponds to an accumulation point in Hilbert space, as was first demonstrated[12] directly in 1966 and 1971. Due to the divergence of the integral of 1/f at f=0, this author reformulated this accumulation point property in dimensional analysis terms before submitting his paper[19] for publication in 1980; in this form, the argument is more elegant and avoids the divergence at f=0.

The author acknowledges the support of the Air Force Office of Scientific Research and of the National Science Foundation.

REFERENCES

1. J.B. Johnson, "The Schottky-Effect in Low-Frequency Circuits", Phys. Rev. 26, 71-85 (1925).

2. W.H. Press, Comments Astrophys, Space Phys. 7, 103 (1978).

3. P.H. Handel: "1/f Noise - an 'Infrared' Phenomenon", Phys. Rev. Letters 34, 1492 - 1494 (1975); "Nature of 1/f Phase Noise", Phys. Rev. Letters 34, 1495-1498 (1975).

4. P.H. Handel: "Quantum Approach to 1/f Noise", Phys. Rev. 22A, p. 745 (1980).

5. P.H. Handel and D. Wolf: "Characteristic Functional of Quantum 1/f Noise", Phys. Rev. A26, 3727-30 (1982).

6. P.H. Handel: "The Quantum 1/f Effect and the General Nature of 1/f Noise", Archiv für Elektronik und Übertragungstechnik (AEÜ) 43, 261-270 (1989);

7. P.H. Handel: "Starting Points of the Quantum 1/f Noise Approach", Submitted to Physical Review B; "Fundamental Quantum 1/f Fluctuation of Physical Cross Sections and Process Rates", submitted to Phys. Rev. Letters.

8. T.S. Sherif and P.H. Handel: "Unified Treatment of Diffraction and 1/f Noise", Phys. Rev. A26, p.596-602, (1982).

9. A. van der Ziel, "Unified Presentation of 1/f Noise in Electronic Devices; Fundamental 1/f Noise Sources", Proc. IEEE 76, 233-258 (1988 review paper); "The Experimental Verification of Handel's Expressions for the Hooge Parameter", Solid-State Electronics 31, 1205-1209 (1988); "Semiclassical Derivation of Handel's Expression for the Hooge Parameter", J. Appl. Phys. 63, 2456-2455 (1988); 64, 903-906 (1988); A.N. Birbas et al., J. Appl. Phys. 64, 907-912 (1988); A. van der Ziel et. al., "Extensions of Handel's 1/f Noise Equations and their Semiclassical Theory", Phys. Rev. B 40, 1806-1809 (1989); M. Tacano, Proc. XI. Int. Conf. on Noise in Physical Systems and 1/f Fluctuations, T. Musha, S. Sato and Y. Mitsuaki Editors, Ohmsha Publ. Co., Tokyo 1991, pp. 167-170; M. Tacano, Proc. Fifth van der Ziel Conference "Quantum 1/f Noise and other Low Frequency Fluctuations" AIP Conference proceedings #282, P.H. Handel and A.L. Chung Editors, 1992; see also "Quantum 1/f Bibliography" by P.H. Handel, unpublished.

10. P.H. Handel, "Any Particle Represented by a Coherent State Exhibits 1/f Noise" in *"Noise in Physical Systems and 1/f Noise"*, (Proceedings of the VIIth International Conference on Noise in Physical Systems and 1/f Noise) edited by M. Savelli, G. Lecoy and J.P. Nougier (North - Holland, Amsterdam, 1983), p. 97.

11. P.H. Handel, "Coherent States Quantum 1/f Noise and the Quantum 1/f Effect" in *"Noise in Physical*

and 1/f Noise) edited by A. D'Amico and P. Mazzetti, Elsevier, New York, 1986, p. 469.

12. P.H. Handel: "Instabilities, Turbulence and Flicker-Noise in Semiconductors I, II and III", Zeitschrift für Naturforschung 21a, 561-593 (1966); P.H. Handel: "Turbulence Theory for the Current Carriers in Solids and a Theory of 1/f Noise", Phys. Rev. A3, 2066 (1971).

13. D. Zwanziger, Phys. Rev. D7, 1082 (1973); Phys. Rev. Lett. 30, 934 (1973); Phys. Rev. D11, 3481 and 3504 (1975); T.W.B. Kibble, Phys. Rev. 173, 1527; 174, 1882; 175,1624 (1968); J. Math. Phys. 9, 315 (1968).

14. Y. Zhang and P. H. Handel, Proc. Fifth van der Ziel Conference "Quantum 1/f Noise and other Low Frequency Fluctuations" AIP Conference proceedings #282, P.H. Handel and A.L. Chung Editors, 1992.

15. J.S. Gradshteiyn and I.M. Ryzhik, *"Table of Integrals, Series and Products"* Sec. 3.761, No. 9 and No. 7, Academic Press, New York 1965.

16. P.H. Handel, Proc. XI. Int. Conf. on Noise in Physical Systems and 1/f Fluctuations, T. Musha, S. Sato and Y. Mitsuaki Editors, Ohmsha Publ. Co., Tokyo 1991, pp. 151-157.

17. F.L. Walls, P.H. Handel, R. Besson and J.J. Gagnepain: "A New Model relating Resonator Volume to 1/f Noise in BAW Quartz Resonators", Proc. 46. Annual Frequency Control Symposium, pp.327-333, 1992.

18. T. Musha, G. Borbely and M. Shoji, Phys. Rev. lett. 64, 2394 (1990).

19. P.H. Handel, T. Sherif, A. van der Ziel, K.M. van Vliet and E.R. Chenette: "Towards a More General Understanding of 1/f Noise", submitted Physics Letters early in 1980. This unpublished manuscript received a wide distribution, in particular in Japan.

20. P.H. Handel and Y. Zhang, Proc. XII. Int. Conf. on Noise in Physical Systems and 1/f Fluctuations, P. Handel, and A. Chung Editors, Amer. Inst. of Phys. Conf. Proceedings #285, St. Louis, 1993, pp. 172-175.

21. P.H. Handel, "Nature of 1/f Frequency Fluctuations in Quartz Crystal Resonators", Solid State Electron. 22, 875-876 (1979).

22. H. Hellwig, "A Look Into the Crystal Ball, the Next 25 Years" in 25th Annual Precise Time and Time Interval Applications and Planning Meeting", NASA Conference Publ. 3267 (1993)

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PIEZOELECTRIC RESONATOR MATERIALS

Arthur Ballato and John G Gualtieri

US Army Research Laboratory
AMSRL-EP, Fort Monmouth, NJ 07703-5601, USA
(908) 544-4308; -3733 fax; a.ballato@ieee.org

Abstract

We introduce a new metric for judging material suitability for high frequency resonator use, and tabulate its value for a number of resonator materials. The new metric combines all the physical quantities associated with acoustic wave propagation in piezocrystals: mass density, dielectric permittivity, piezoelectric modulus, elastic stiffness, and viscosity. Use of this metric to rank materials yields some nonobvious and counterintuitive conclusions with implications for the directions the technology might take, and where the emphases for future progress might be placed.

Introduction

Piezoelectric materials function in a variety of electromechanical devices, such as resonators, filters, sensors, transducers, delay lines, and actuators. The material properties of primary importance vary with the application. In wideband filters, e.g., coupling coefficient considerations are paramount. Crystal resonators for frequency control and selection have been gauged by a variety of material or circuit measures over the years. To an extent, the differing measures pertain to the intended uses; few will debate the importance of capacitance ratio in VCXO crystals, or the role of resonator resistance in determining the insertion loss of a narrowband filter. On the other hand, none of the usual performance figures is entirely adequate for characterizing resonator materials for cellular radio, radar, and similar applications at the higher frequencies.

This paper describes the various performance metrics traditionally applied to resonators and resonator materials, and lists observed values for various piezoelectric crystals, both old and new.

It then introduces a new metric for judging material suitability for high frequency resonator use, and tabulates its value for these same materials. The new metric is a combination of all of the physical quantities associated with acoustic wave propagation in piezocrystals: mass density, dielectric permittivity, piezoelectric modulus, elastic stiffness and viscosity. Use of this characteristic number to rank resonator materials yields some nonobvious and somewhat counterintuitive conclusions that reflect on the directions the technology of these devices might take, and where the emphases for future progress might be placed.

Piezoresonators

The earliest bulk wave piezoelectric resonators were bars [1]. In order to go to the progressively higher frequencies required by applications, resonator development passed successively from use of length modes of bars and rods, through contour modes of plates, to thickness modes of plates, [1, 2] and presently to ring-supported, [3-8] stacked-filter, [9,10] and thin-film [11-16] configurations.

To focus the discussion, we consider here piezoelectric bulk wave resonators operating in thickness modes [17-21]. We first discuss quartz, but will subsequently be led to material independent conclusions. The discussion can be broadened, *mutatis mutandis*, to encompass SAW, [18, 21] SBAW, [22, 23] and other structures and wave types using piezoelectric, as well as piezomagnetic [24] materials.

In the early 1930s, typical frequencies for quartz plate resonators were 1 or 2 MHz. By the early 1940s, the upper frequency limit was 10 MHz; this was primarily dictated by manufacturing limitations. The limit that could be readily achieved in commercial practice remained about 10 MHz well into the 1950s. From the mid-1940s to the mid-1980s, the trend line for factory units showed a slope of about 70 years per decade, i.e., the upper frequency commercially manufacturable increased with a rate that would have led to a factor of ten increase over a time span of seventy years. For laboratory units, 30 MHz resonators could be made in the mid-1940s; the subsequent slope is 35 years per decade of frequency.

U.S. Government work not protected by U.S. copyright

For the past half-century and more, the push to higher frequencies was driven by practical applications; it continues today. During this time, a recurring question has been "What is the highest useful acoustic resonator frequency?" Until relatively recently the answer depended entirely on manufacturing technology and not *per se* on the material involved. But with the introduction of IDT-driven SAW, and ion-milled BAW resonators, the limiting value appeared to devolve to the propagation medium, with a limiting value around several GHz. If this limit were to hold, it would thus represent a thousand-fold increase over the frequencies of the earliest thickness-mode resonators.

We shall find that there is substantial reason to believe that the material-limited upper frequency for many piezoelectrics is considerably higher than this.

Acoustic Velocity

The orientation dependence of the propagation velocity of acoustic waves in crystals was known in the early 1930s [25]. For a given processing-limited minimum achievable plate thickness, it was recognized that use of the BT cut of quartz would give a frequency that was higher than that of the corresponding AT cut in the ratio of the frequency constants (half-velocities): $N_0(BT)/N_0(AT) \sim 2.53/1.66 \sim 1.53$. Acoustic velocity is one of the important attributes characterizing an acoustic medium; it depends primarily on the elastic stiffness and mass density.

Acoustic Loss

Another parameter of importance is that of acoustic loss. It is often quantified in terms of the Q of a vibrator. Starting in the 1950s, resonator designs and operating frequencies allowed one to separate out the losses due to the material and losses originating from external factors such as mounting and radiation into the ambient fluid surrounding the crystal [26-28]. It was recognized that a BT cut had a higher material-limited Q than an AT cut of the same frequency; indeed, it is even higher for plates of the same thickness.

The material-limited losses in quartz are often spoken of in terms of the constancy of the (Q•frequency) product. A related, and for many purposes, a better measure is the motional time constant τ_1 [17, 29-31]. This quantity is related to the acoustic viscosity, [32] and to the correlation of Q with infra-red spectra [33, 34].

The motional time constant τ_1 is another important attribute characterizing an acoustic medium, since it is a measure of loss; it depends primarily on acoustic viscosity and elastic stiffness.

Piezocoupling

The piezoelectric coupling factor k [31, 35] is a dimensionless measure of efficacy of electromechanical energy conversion in a crystal resonator. It is formed from the piezoelectric, elastic, and dielectric coefficients, and is bounded by $0 \le k < 1$.

The coupling factor is a third attribute of importance characterizing a piezoelectric acoustic medium, because it allows the resonator to be driven by an impressed voltage, and is a primary determinant of certain characteristic frequencies of interest.

Equivalent Network

The Butterworth-Van Dyke (BVD) equivalent network of a piezoelectric resonator [1, 2, 17-19, 36-39] is a simple, four-element, circuit used to characterize the mechanical vibrations of a crystal structure in electrical terms. The circuit element values [17, 18, 37] are related to the physical parameters of the medium and the geometry of the structure. For simplicity, we consider all material coefficients to be scalars, and the structure to be a thin plate of thickness t, with coextensive electrodes of area A on the major surfaces. Then, using conventional nomenclature for the material coefficients, circuit elements, frequencies, and harmonics, we can schematically outline the resultant concordance between physics and electrical engineering as follows:

Physics (ρ, ε, e, c, η) Geometry (t, A) EE (C_0, C_1, R_1, L_1)

$c^* = c + (e^2 / \varepsilon)$; $v^2 = c^* / \rho$; $k^2 = e^2 / (\varepsilon\, c^*)$;

$r^{(M)} = C_0/C_1^{(M)} = [\pi^2/8k^2] \bullet M^2$;

$\tau_1^{(M)} = R_1^{(M)} C_1^{(M)} = R_1 C_1 = \eta / c^* = 1/\omega_M Q_M$;

$\omega_M^2 = 1/ L_1^{(M)} C_1^{(M)}$; $f^{(M)} = M v /2t = \omega_M / 2\pi$

$C_0 = \varepsilon (A/t)$

$C_1^{(M)} = [8 e^2 / M^2 \pi^2 c^*] (A/t)$

$L_1^{(M)} = L_1 = [\rho / 8 e^2] (t^3/A)$

$R_1^{(M)} = [M^2 \pi^2 \eta / 8 e^2] (t/A)$

Network Functions

The BVD circuit yields the immittance functions, such as reactance, by simple network procedures. The nominal frequency of the resonator at harmonic M is given by $f^{(M)} = M v /2t$; this is often called the series resonance frequency, f_S , and is used for normalizing the network functions. Despite the simplicity of the BVD circuit, there are quite a few other distinct characteristic frequencies; these are defined from the zeros and extrema of the various network functions [36].

For example, setting the reactance function to zero defines the resonance and antiresonance frequencies, f_R and $f_A > f_R$; these are found from the roots of the quadratic $[(1 - x)^2 + x/Q^2] + (1/r)(1 - x) = 0$, where $\sqrt{x} = f/f_S$. Between f_R and f_A the reactance is positive, indicating that the crystal looks like an inductor; outside this range the reactance is capacitive. For most oscillator applications the crystal operates in the region of inductive reactance. It is seen from the quadratic that the f_R - f_A separation ("pole - zero" distance) depends jointly on Q and r.

Reference [36] details the various measures used to describe the region and extent of resonance. For our purposes here it is enough to recognize that resonance as usually construed ceases to exist when the condition $f_R = f_A$ is satisfied. At this point the crystal just ceases to achieve a positive reactance, and its utility for frequency control and selection purposes is limited. The point at which f_R and f_A coalesce to a single frequency is found by setting the discriminant of the quadratic equal to zero. This leads, in turn, to another quadratic on the square of the figure of merit, $\mathcal{M}$ = Q/r, as function of the capacitance ratio, r. By solving this we find $\mathcal{M}^2 = (2 + 1/r) + 2\sqrt{(1 + 1/r)}$; this is a weak function of r, and to a good approximation we get the criterion $\mathcal{M}$ = Q/r ≈ 2. For values of $\mathcal{M}$ of 2 or less, resonance ceases to exist [2]. Since Q decreases with absolute frequency and r increases with harmonic, $\mathcal{M}$ is diminished both by increasing the harmonic as well as the frequency of operation. The limiting condition $\mathcal{M}$ = 2 thus imposes a constraint jointly upon frequency (or equivalently, the thickness) and harmonic.

This is a very simple criterion, and we shall see that it leads to interesting conclusions. On the other hand, it is somewhat arbitrary, and is by no means applicable to all situations. For example, in high stability oscillators the quality factor Q, by itself, has an overriding importance with regard to noise behavior [40, 41].

Frequency Limits

We now use the criterion $\mathcal{M}$ = Q/r = 2 to infer the frequency limits imposed on simple thickness mode resonators by the intrinsic phenomenological material constants of piezoelectric acoustic media [42]. From the BVD relations given above we have:

$r^{(M)} = C_0/C_1^{(M)} = [\pi^2/8k^2] \bullet M^2$, so $r^{(M)}$ is quadratic in M.

$Q = (2\pi f \tau_1)^{-1} = (2\pi\tau_1)^{-1} \bullet (1/f)$;

$f = f^{(M)}$ = operating frequency at M^{th} harmonic.

Therefore, $\mathcal{M} = Q/r = [4k^2/\pi^3 \tau_1] \bullet (f M^2)^{-1}$.

The condition for $f_A = f_R$ is

$\mathcal{M} = 2 = [4k^2/\pi^3 \tau_1] \bullet (f M^2)^{-1}$, so

$f M^2 = [2k^2/\pi^3 \tau_1]$, therefore $f M^2$ is equal to a constant.

Using $f^{(M)} = M \bullet v/2t = (N_0/t) \bullet M$ with

$f M^2 = [2k^2/\pi^3 \tau_1]$ allows one to solve for M:

$M^3 = [2k^2/\pi^3 \tau_1 N_0] \bullet t$; hence

$M^3 = t / L_0$, where L_0 is a characteristic length [42].

$$L_0 = [\pi^3 \tau_1 N_0/2k^2]$$

By way of a concrete example, for AT-cut quartz, k = 8.80%, τ_1 = 11.8 fs, and N_0 = 1661 m/s [31, 32, 43]. Therefore, $L_0 \approx$ 39.2 nm, or roughly one hundred molecular spacings! L_0 is the thickness for which $\mathcal{M}$ = Q/r = 2 at the fundamental harmonic. The smallness of L_0 compared to the usual resonator thicknesses encountered in practice leads one to conclude that substantially higher frequencies are available as far the material constants are concerned; the difficulty arises from an inability to fabricate resonators to the requisite thinness so that they can be used at the fundamental harmonic.

The characteristic length L_0 can equivalently be expressed in terms of material (physics) constants:

$\tau_1 = \eta/c$; $N_0 = \frac{1}{2}\sqrt{(c/\rho)}$; $k^2 = e^2/\varepsilon^S c$; so

$L_0 = [\pi^3 \tau_1 N_0/2k^2] = (\pi^3/4) \bullet [\sqrt{(c/\rho)} \bullet \eta \bullet \varepsilon^S/e^2]$.

L_0 contains all relevant acoustic parameters: ρ, ε^S, e, c, and η.

$L_0 = L_0$(velocity, coupling, loss).

For a given plate thickness t, the maximum usable harmonic is
$M_{(max)} = [t/L_0]^{1/3}$, for $\mathcal{M}$ = Q/r = 2, and $[2t/\mathcal{M}L_0]^{1/3}$ in general.

Of course, M is to be rounded down to an odd integer.

At $M_{(max)}$, the operating frequency is:

$f = N_0/(L_0 \bullet t^2)^{1/3}$.

In general, for an arbitrary value of $\mathcal{M}$:

$f^3 = (N_0^3/L_0) \bullet (2/\mathcal{M}) \bullet (1/t^2)$, so a plot of log f versus log t has slope -2/3.

An Example

Consider three AT-cut quartz plates of differing thicknesses, operating at M = 1, 3, and 5, such that $\mathcal{M}$ = 2 in each case.

From $M_{(max)} = [t/L_0]^{1/3}$, we find that the thicknesses go as M^3 , and the corresponding fundamental frequencies are:

$f^{(1)} = (N_0/L_0) \bullet (1/M)^3$. The frequencies at the intended harmonics are:

$f^{(M)} = (N_0/L_0) \bullet (1/M)^2$.

With L_0 = 39.2 nm and N_0 = 1661 m/s, one obtains:

Plate	M	thickness	(μm)	f (Ghz)
#1	1	t_1 =	L_0 = 0.0392	$f^{(1)}$ = 42.4
#2	3	t_3 =	27 L_0 = 1.06	$f^{(3)}$ = 4.71
#3	5	t_5 =	125 L_0 = 4.90	$f^{(5)}$ = 1.70

If one desires to attain high frequencies, it is clear that one pays a severe penalty for using harmonics other than the fundamental. The issue is clearly settled on the basis of available technology for making thin films/membranes. Technology, not material, is the real limitation.

To be sure, there are certain material-related issues associated with crystal growth, etch pits, channels, impurities, surface finish [44, 45], and contamination. Manufacturing problems, such as achieving films with adequate parallelism [46-48] are, however, the main difficulty. Other issues, such as dynamic thermal compensation [49] must be considered as well. With suitable attention focused not on material limitations but primarily on overcoming fabrication shortcomings, the way will be clear for assuring a bright future for acoustic resonators at microwave and millimeter-wave frequencies.

Rotated Quartz Cuts

The availability of the full viscosity matrix for quartz [32] permits calculation of L_0 for all orientations. We use Bechmann's values [43] for the remaining quartz constants, and compute L_0 for rotated-Y-cuts by way of illustration. Thickness-excitation (i.e., electrodes on the major surfaces) for these orientations drives only the pure mode with particle motion along the digonal axis, and results in simple expressions well suited for an example. The calculation proceeds in stages as follows:

With $c\theta = \cos\theta$ and $s\theta = \sin\theta$,

- Rotated elastic stiffness:

$$c'_{66}(\theta) = c^E_{66}\, c^2\theta + c^E_{44}\, s^2\theta + 2\, c^E_{14}\, s\theta\, c\theta$$

- Rotated piezoelectric constant:

$$e'_{26}(\theta) = -\, c\theta\, [\, e_{11}\, c\theta + e_{14}\, s\theta\,]$$

- Rotated dielectric permittivity:

$$\varepsilon'_{22}(\theta) = \varepsilon^S_{11}\, c^2\theta + \varepsilon^S_{33}\, s^2\theta$$

- Piezoelectrically stiffened elastic stiffness:

$$c^*_{66}(\theta) = c'_{66}(\theta) + [e'_{26}(\theta)]^2 / \varepsilon'_{22}(\theta)$$

The piezoelectric contribution to the stiffening, $[c^*_{66}(\theta) - c'_{66}(\theta)]$, is plotted versus θ in Figure 1.

- Rotated frequency constant:

$$N_0(\theta) = \tfrac{1}{2}\sqrt{(c^*_{66}(\theta)/\rho)}$$

$N_0(\theta)$ is plotted versus θ in Figure 2.

- Rotated piezoelectric coupling coefficient squared:

$$[k'_{26}(\theta)]^2 = k^2(\theta) = [e'_{26}(\theta)]^2 / [c^*_{66}(\theta)\ \varepsilon'_{22}(\theta)]$$

$k'_{26}(\theta)$ is plotted versus θ in Figure 3, and $[k'_{26}(\theta)]^2$ is plotted versus θ in Figure 4.

- Rotated elastic viscosity:

$$\eta'_{66}(\theta) = \eta_{66}\, c^2\theta + \eta_{44}\, s^2\theta + 2\, \eta_{14}\, s\theta\, c\theta$$

- Rotated motional time constant:

$$\tau_1(\theta) = \eta'_{66}(\theta) / c^*_{66}(\theta)$$

$\tau_1(\theta)$ is plotted versus θ in Figure 5.

- Rotated characteristic length:

$L_0(\theta) = [\pi^3/2]\{\tau_1(\theta) N_0(\theta) / k^2(\theta)\}$

$L_0(\theta)$ and $1/L_0(\theta)$ are plotted versus θ in Figures 6 and 7

Other Materials

The examples and figures above pertain to quartz. This is only in part due to its ascendant importance as a frequency control material. The larger reason is that, for nearly all other piezoelectric acoustic materials, the elements of the viscosity matrix are unknown, and one is limited to patchy data in the form of Q values for specific cuts and modes [50-64]. The state of material science sophistication and accuracy of electrical parameter measurement have progressed to the stages where complete determinations of the viscosity tensor (including behavior with temperature) ought to be part of every data set.

In Figures 8 and 9 are shown log-log plots of operating frequency versus plate thickness for various materials and cuts, subject to the constraint $\mathcal{M} = 2$. Since $f^3 = (N_0^3/L_0) \bullet (1/t^2)$, the lines have slope -2/3. The lines are drawn by using M implicitly as a continuous variable. Indicated on each line are the physically meaningful, odd-integer, values of M in the range 1, 3, ..., 25. Above each curve is the region where $\mathcal{M} < 2$; the inequality is reversed for the region below each curve.

Figure 8 describes four non-temperature-compensated materials: AlN, ZnO, langasite, and lithium niobate. For the first three materials/cuts, the fundamental frequency is about 100 GHz, at thicknesses of 20 to 30 nm, or roughly 100 molecules thick. For lithium niobate, L_0 is only 1.3 angstrom; at such dimensions the continuum approximation breaks down, but the general conclusion is still valid: frequencies in excess of 100 GHz should be accessible for acoustic resonators if fabrication difficulties can be overcome.

Figure 9 is plotted for temperature compensated orientations of langasite, lithium tetraborate, and quartz. For these materials/cuts, the figure indicates that frequencies between 10 and 100 GHz are accessible, again providing that attention is focused on the manufacturing problems involved.

Table I lists frequency constants (N_0), motional time constants (τ_1), piezocoupling factors squared (k^2), and characteristic lengths (L_0) for some piezoelectric resonators.

Conclusions

We have introduced a new metric for gauging piezo materials for high frequency resonator application. It is a combination of velocity, loss, and coupling factor. Use of the measure leads one to conclude that piezo materials should be useful at frequencies well above X-band, and perhaps to 100 GHz and beyond. One concludes that much greater emphasis ought to be placed in the future on devising fabricating methods for making such resonator structures commercially manufacturable.

References

[01] W G Cady, Piezoelectricity, McGraw Hill, New York, 1946; Dover, New York, 1964.

[02] R A Heising, Quartz Crystals for Electric Circuits, Van Nostrand, New York, Chap. XII, pp. 394-410, 1946.

[03] G K Guttwein, A Ballato, and T J Lukaszek, "VHF-UHF piezoelectric resonators," US Patent 3,394,677, issued September 1972.

[04] M Berté, "Acoustic-bulk-wave resonators and filters operating in the fundamental mode at frequencies greater than 100 MHz," Electronics Lett., Vol. 13, No. 9, 28th April 1977, pp. 248-250.

[05] M Berté, "Acoustic bulk wave resonators and filters operating in the fundamental mode at frequencies greater than 100 MHz," Proc. 31st Ann. Freq. Contr. Symp., pp. 122-125, June 1977.

[06] L Bidart and J Chauvin, "Direct frequency crystal oscillators," Proc. 35th Ann. Freq. Contr. Symp., pp. 365-375, May 1981.

[07] M Nakazawa, T Lukaszek, and A Ballato, "Force- and acceleration-frequency effects in grooved and ring supported resonators," Proc 35th Annual Frequency Control Symp., pp. 71-91, May 1981.

[08] A Ballato, T J Lukaszek, and G J Iafrate, "Subtle effects in high-stability quartz resonators," Ferroelectrics, Vol. 43, Nos. 1/2, pp. 25-41, 1982.

[09] A Ballato and T Lukaszek, "Stacked-crystal filters," Proc. IEEE, Vol. 61, No. 10, pp. 1495-1496, October 1973.

[10] A Ballato, H L Bertoni, and T Tamir, "Systematic design of stacked-crystal filters by microwave network methods," IEEE Trans. Microwave Theory Tech., Vol. MTT-22, No. 1, pp. 14-25, January 1974.

[11] E A Gerber, T Lukaszek, and A Ballato, "Advances in microwave acoustic frequency sources," IEEE Trans. Microwave Theory Tech., Vol. MTT-34, No. 10, pp. 1002-1016, October 1986.

[12] K M Lakin, G R Kline, R S Ketcham, A R Landin, W A Burkland, K T McCarron, S D Braymen, and S G Burns, "Thin film resonator technology," Proc. 41st Ann. Freq. Contr. Symp., pp. 371-381, May 1987.

[13] S V Krishnaswamy, "Film bulk acoustic wave resonator technology," IEEE MTT-S Newsletter, No. 130, pp. 21-25, Fall 1991.

[14] K M Lakin, G R Kline, and K T McCarron, "High Q microwave acoustic resonators and filters," 1993 IEEE MTT-S International Microwave Symp. Digest, Vol. 3, pp. 1517-1520, June 1993.

[15] K M Lakin, G R Kline, and K T McCarron, "Overmoded high Q resonators for microwave oscillators," IEEE Internat. Freq. Contr. Symp. Proc., pp. 718-721, June 1993.

[16] R Stokes and J D Crawford, "X-Band thin film acoustic filters on GaAs," IEEE Trans. Microwave Theory Tech., Vol. 41, No. 6/7, pp. 1075-1080, June/July 1993.

[17] A Ballato, "Doubly rotated thickness mode plate vibrators," in Physical Acoustics: Principles and Methods, (W P Mason and R N Thurston, eds.), Vol. 13, Chap. 5. Academic Press, New York, 1977, pp. 115-181.

[18] E A Gerber and A Ballato, eds., Precision Frequency Control, Vol. 1 and 2, Academic Press, New York and Orlando, 1985.

[19] "IEEE Standard on Piezoelectricity," IEEE Standard 176-1987, IEEE, New York, 54 pp.

[20] J F Rosenbaum, Bulk Acoustic Wave Theory and Devices, Artech House, Boston and London, 1988.

[21] B A Auld, Acoustic Fields and Waves in Solids, 2nd edition, Vol. I and II, R. E. Krieger Pub. Co., Malabar, FL, 1990.

[22] A Ballato and T J Lukaszek, "Shallow bulk acoustic wave progress and prospects," IEEE Trans. Microwave Theory Tech., Vol. MTT-27, No. 12, pp. 1004-1012, December 1979.

[23] T J Lukaszek and A Ballato, "Microwave frequency sources employing shallow bulk acoustic wave devices," Microwave Journal, Vol. 28, No. 3, pp. 77-99, March 1985.

[24] "IEEE Standard on Magnetostrictive Materials: Piezomagnetic Nomenclature," IEEE Standard 319-1990, IEEE New York, 12 pp.

[25] R Bechmann, "Piezoelektrische Quarzoszillatoren mit beliebigem Temperaturkoeffizienten, insbesondere dem Wert Null," Naturwissenschaften, Vol. 21, No. 42, 20 October 1933, p. 752.

[26] H E Bömmel, W P Mason, and A W Warner, "Experimental evidence for dislocations in crystalline quartz," Phys. Rev., Vol. 99, pp. 1894-1896, 1955.

[27] A W Warner, "Design and performance of ultraprecise 2.5-mc quartz crystal units," Bell Syst. Tech. Jour., Vol. 39, pp. 1193-1217, September 1960.

[28] A Seed, "The internal friction of quartz as a function of strain amplitude and frequency," Brit. J. Appl. Phys., Vol. 16, pp. 87-91, 1965.

[29] G K Guttwein, T J Lukaszek, and A Ballato, "Practical consequences of modal parameter control in crystal resonators," Proc. 21st Ann. Freq. Contr. Symp., pp. 115-137, April 1967.

[30] G K Guttwein, A Ballato, and T J Lukaszek, "Design considerations for oscillator crystals," Proc. 22nd Annual Frequency Control Symp., pp. 67-88, April 1968.

[31] B Parzen, Design of Crystal and Other Harmonic Oscillators, Wiley and Sons, New York, 1985, Chapter 3 "Piezoelectric Resonators," pp. 66-122 and 432-436.

[32] J Lamb and J Richter, "Anisotropic acoustic attenuation with new measurements for quartz at room temperature," Proc. Roy. Soc. (London), Vol. 293A, pp. 479-492, 1966.

[33] B Sawyer, "Q capability indications from infrared absorption measurements for Na_2CO_3 process cultured quartz," IEEE Trans. Sonics Ultrason., Vol. 19, pp. 41-44, 1972.

[34] D B Fraser, D M Dodd, D W Rudd, and W J Carroll, "Using infrared to find the mechanical Q of α-quartz," Frequency, Vol. 4, No. 1, pp. 18-21, 1966.

[35] "IRE Standards on Piezoelectric Crystals: Determination of the Elastic, Piezoelectric, and Dielectric Constants - The Electromechanical Coupling Factor, 1958," Proc. IRE, Vol. 46, No. 4, April 1958, pp. 764-778. IEEE Standard 178.

[36] A Ballato, "Resonance in piezoelectric vibrators," Proc. IEEE, Vol. 58, No. 1, pp. 149-151, January 1970.

[37] A Ballato, "Frequency-temperature-load capacitance behavior of resonators for TCXO application," IEEE Trans. Sonics Ultrason., Vol. SU-25, No. 4, pp. 185-191, July 1978.

[38] J C Brice, "Crystals for quartz resonators," Revs. Mod. Phys., Vol. 57, pp. 105-146, 1985.

[39] A Ballato, "Equivalent networks for piezoelectric resonators and transducers," 4th European Frequency and Time Forum, Neuchâtel, March 1990. Published in: Proc. 5th European Frequency and Time Forum, Besançon, pp. 465-473, March 1991.

[40] E Hafner, "The effects of noise in oscillators," Proc. IEEE, Vol. 54, No. 2, February 1966, pp. 179-198.

[41] E Hafner, "Analysis and design of crystal oscillators, Part I," Technical Report ECOM-2474, May 1964, 47 pp., US Army Electronics Command, Fort Monmouth, NJ. Available from NTIS as document AD 611 200.

[42] A Ballato and J G Gualtieri, "Advances in high-Q piezoelectric resonator materials and devices," IEEE Trans. Ultrason., Ferro., & Freq. Control, in press.

[43] R Bechmann, "Elastic and piezoelectric constants of alpha-quartz," Phys. Rev., Vol. 110, No. 5, pp. 1060-1061, June 1958.

[44] J R Vig, J W LeBus, and R L Filler, "Chemically polished quartz," Proc. 31st Annual Frequency Control Symp., pp. 131-143, June 1977.

[45] R N Castellano, T R Meeker, R C Sundahl, and J C Jacobs, "The relationship between quartz surface morphology and the Q of high frequency resonators," Proc. 31st Annual Frequency Control Symp., pp. 126-130, June 1977.

[46] A Ballato, T Lukaszek, H Wasshausen, and E Chabak, "Design and fabrication of modern filter crystals," Proc. 20th Annual Frequency Control Symp., April 1966, pp. 131-160.

[47] E A Gerber, "Comments on unwanted responses in VHF crystals," Proc. 20th Annual Frequency Control Symp., April 1966, pp. 161-166.

[48] T Lukaszek, A Ballato, and H Wasshausen, "Additional considerations for VHF filter crystal design," Proc. IEEE, Vol. 54, No. 10, pp. 1466-1467, October 1966.

[49] A Ballato, "Static and dynamic behavior of quartz resonators," IEEE Trans. Sonics Ultrason., Vol. SU-26, No. 4, pp. 299-306, July 1979.

[50] Landolt-Börnstein, Numerical Data and Functional Relationships in Science and Technology, New Series, Group III: Crystal and Solid State Physics, Volumes III/1, 1966; III/2, 1969; III/3, 1969; III/9, 1974; III/11, 1979; III/16a, 1981; III/16b, 1982; and III/17a, 1982. Springer-Verlag, Berlin, New York.

[51] J G Gualtieri and A Ballato, "Advances in high-Q piezoelectric materials and devices," Proc. 15th Annual Piezoelectric Devices Conference, pp. 1-12, September 1993. Electronic Industries Association.

[52] J G Gualtieri, J A Kosinski, and A Ballato, "Piezoelectric materials for acoustic wave applications," IEEE Trans. Ultrason., Ferro., and Freq. Control, Vol. 41, No. 1, pp. 53-59, January 1994.

[53] J G Gualtieri, J A Kosinski, and A Ballato, "Piezoelectric material constants for surface wave research," Proc. 7th European Frequency and Time Forum, Neuchâtel, pp. 231-234, March 1993.

[54] J G Gualtieri, J A Kosinski, and A Ballato, "Piezoelectric materials for SAW applications," IEEE Ultrason. Symp. Proc., pp. 403-412, October 1992.

[55] A Ballato and T Lukaszek, "Microwave acoustic material properties," Microwave Journal, Vol. 33, No. 5, pp. 105-114, May 1990.

[56] A A Kaminskii, I M Silvestrova, S E Sarkisov, and G A Denisenko, "Investigation of trigonal $(La_{1-x}Nd_x)Ga_5SiO_{14}$, (II spectral laser and electromechanical properties)," Physica Status Solidi, Vol. 80, 1983, pp. 607-620.

[57] A B Ilyaev, B S Umarov, L A Shabanova, and M F Dubovik, "Temperature dependence of electromechanical properties of LGS crystals," Physica Status Solidi A, Vol. 98, December 1986, pp. K109-K114.

[58] I M Silvestrova, Yu V Pisarevskii, B Mill, and A A Kaminskii, "Acoustical and electromechanical properties of piezoelectric crystals with a trigonal Ca-gallogermanate structure," Sov. Phys. Dokl., Vol. 30, No. 5, May 1992, pp. 402-403.

[59] J A Kosinski, A Ballato, and Y Lu, "Measured properties of doubly rotated dilithium tetraborate ($Li_2B_4O_7$) resonators and transducers," IEEE Trans. Ultrason., Ferro., and Freq. Control, Vol. 40, No. 2, March 1993, pp. 154-161.

[60] A Ballato, J Kosinski, and T Lukaszek, "Lithium tetraborate transducers," IEEE Trans. Ultrason., Ferro., and Freq. Control, Vol. 38, No. 1, January 1991, pp. 62-66.

[61] D H Turnbull, M D Sherar, and F S Foster, "Determination of electromechanical coupling coefficients in transducer materials with high mechanical losses," IEEE Ultrason. Symp. Proc., October 1988, pp. 631-634.

[62] E Philippot, A Ibanez, A Goiffon, A Zarka, J Schwartzel, and J Détaint, "New approach of crystal growth and characterization of a quartz and berlinite isomorph: $GaPO_4$," Proc. 1992 IEEE Frequency Control Symp., May 1992, pp. 744-752.

[63] B H T Chai, M L Shand, E Buehler, and M A Gilleo, "Experimental data on the piezoelectric properties of berlinite," IEEE Ultras. Symp. Proc., September 1979, pp. 577-583.

[64] J van Randeraat, ed., Piezoelectric Ceramics, N V Philips' Gloeilampenfabrieken, Eindhoven, The Netherlands, June 1968, 118 pp.

[65] Properties of Lithium Niobate, London and New York: INSPEC, IEE 1989, EMIS Data reviews series No. 5, esp. pp. 32, 83 and 123.

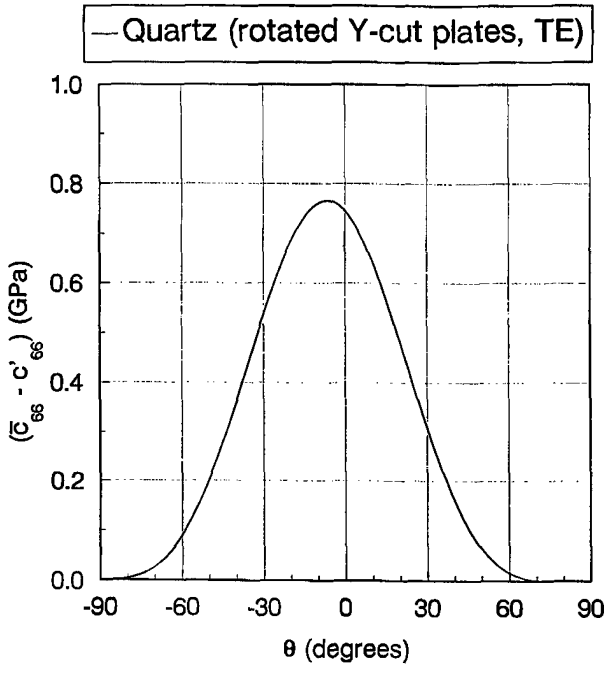


FIGURE 1. Piezoelectric contribution to the elastic stiffening. Rotated-Y-cuts of quartz.

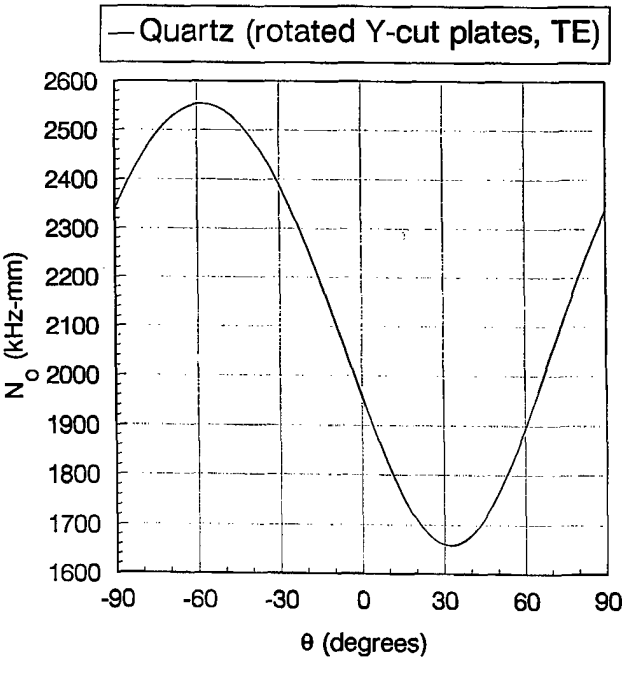


FIGURE 2. Frequency constant. Rotated-Y-cuts of quartz.

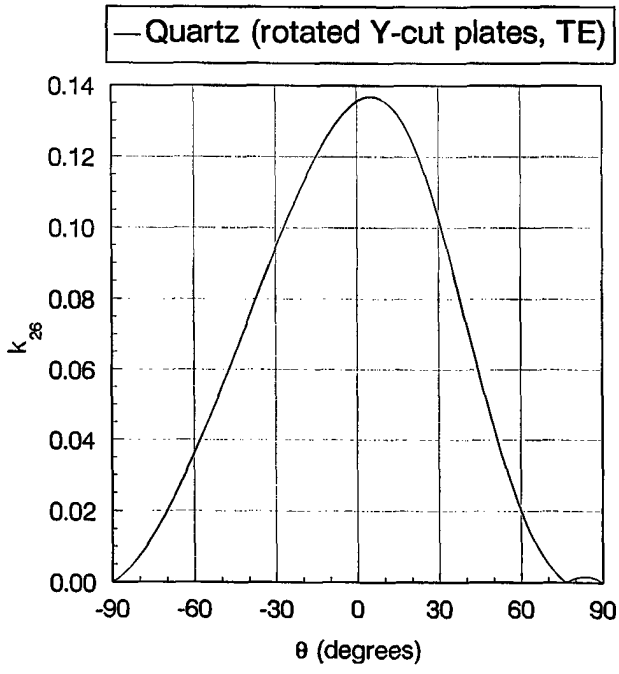


FIGURE 3. Piezoelectric coupling factor. Rotated-Y-cuts of quartz.

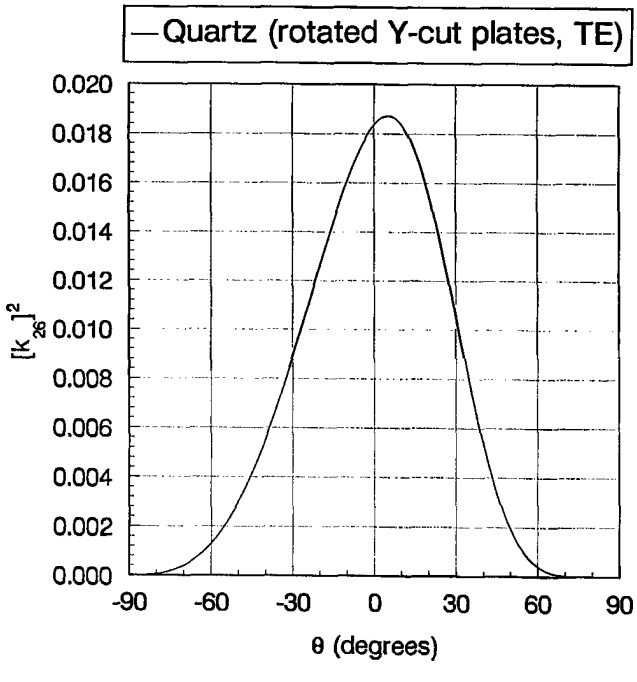


FIGURE 4. Piezoelectric coupling factor squared. Rotated-Y-cuts of quartz.

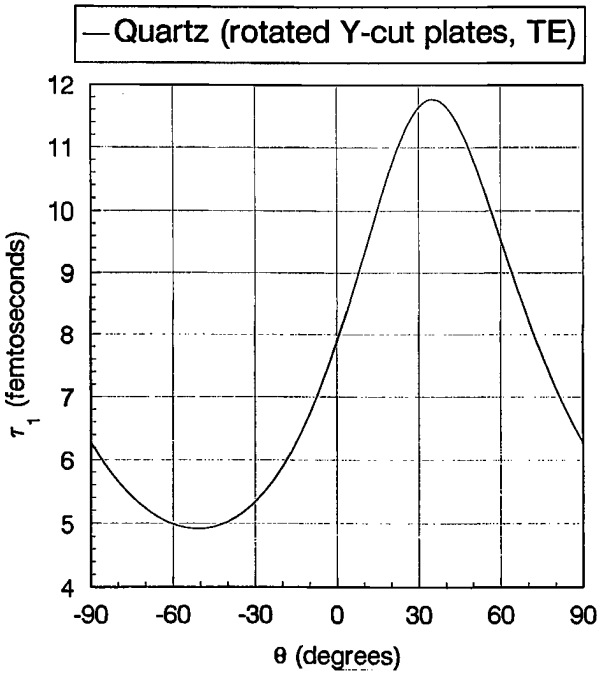


FIGURE 5. Motional time constant. Rotated-Y-cuts of quartz.

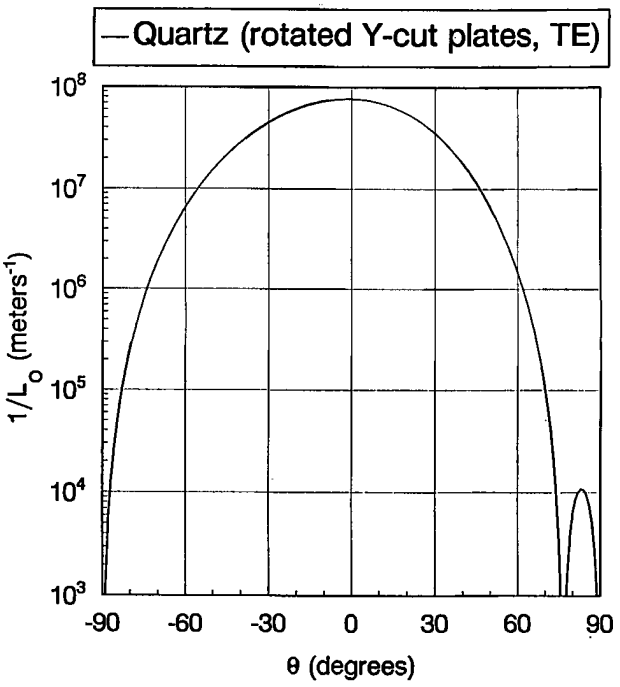


FIGURE 7. Reciprocal characteristic length. Rotated-Y-cuts of quartz.

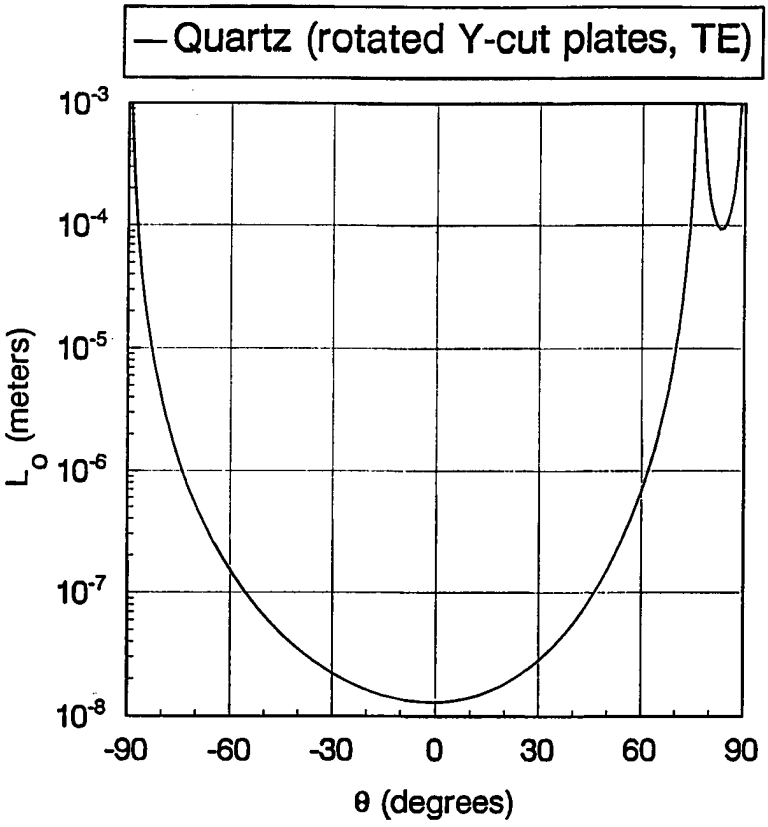


FIGURE 6. Characteristic length. Rotated-Y-cuts of quartz.

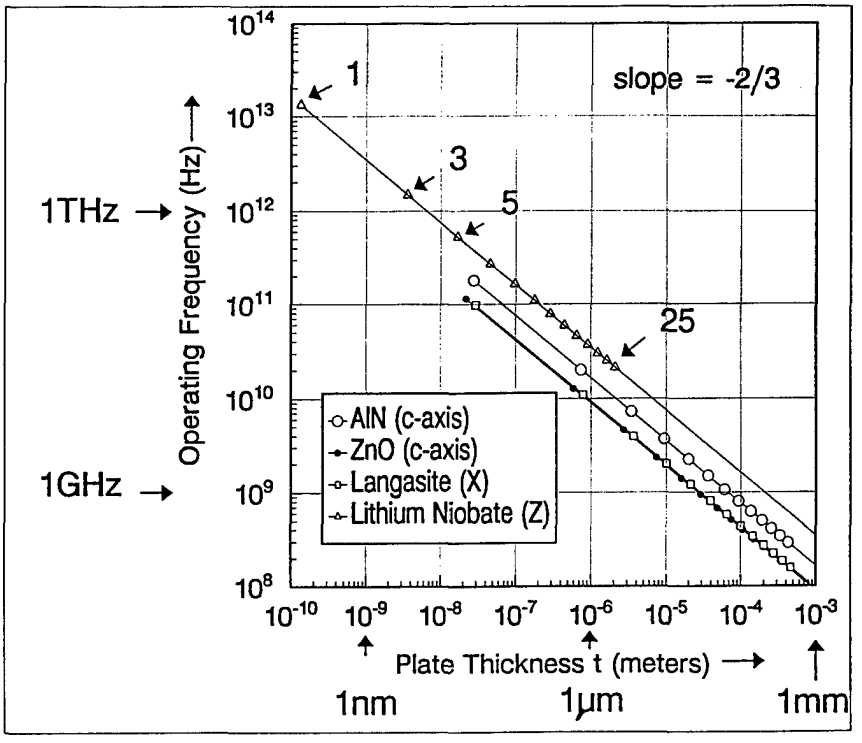


FIGURE 8. Operating frequency versus plate thickness for nontemperature compensated materials/cuts.

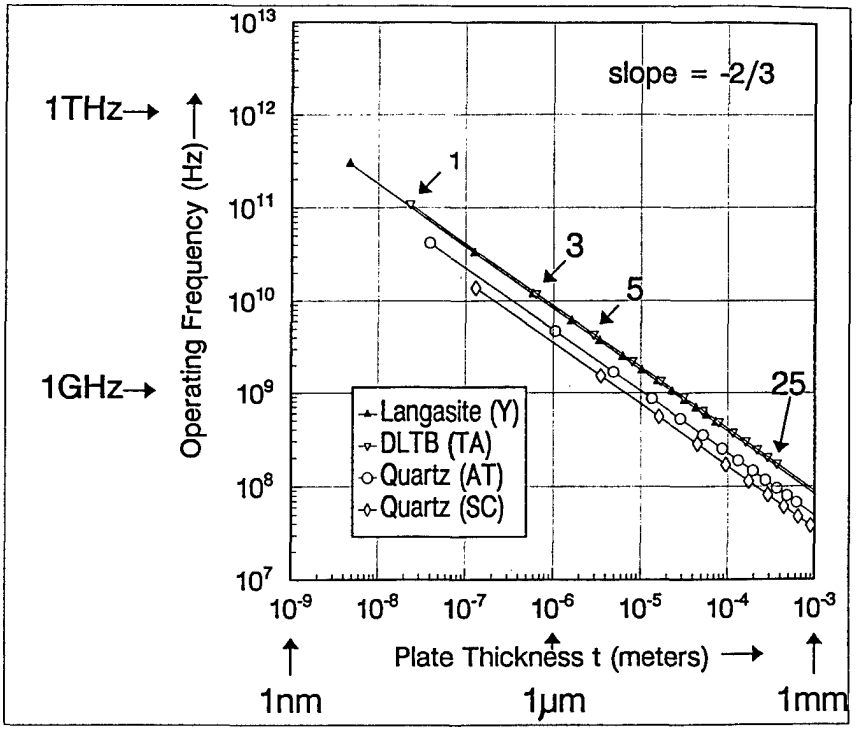


FIGURE 9. Operating frequency versus plate thickness for temperature compensated materials/cuts.

TABLE I. MOTIONAL TIME CONSTANTS AND CHARACTERISTIC LENGTHS FOR SOME PIEZOELECTRIC RESONATORS

Type/Cut	N_0 (m/s)	Q	τ_1 (ps)	k^2 (%)	L_0 (μm)	Reference
Bulk Resonators						
Quartz (AT)	1661		0.0118	0.77	0.0392	[31]
Quartz (BT)	2536		0.0049	0.32	0.0620	[31]
Quartz (SC)	1797		0.0117	0.25	0.1309	[31]
α-$AlPO_4$ (Y)	1813	3500	6.27	5.4	3.26	[63]
$GaPO_4$ (Y,21°)	1258	5000	6.37	2.0	6.20	[62]
$GaPO_4$ (Y,15°)	1261	12800	2.49	2.6	1.87	[62]
$GaPO_4$(Y,12.5°)	1270	10100	3.15	2.6	2.39	[62]
$LiNbO_3$ (Z)	1787		0.0029	60.7	0.0001	[65]
$La_3Ga_5SiO_{14}$ (Y)	1384		0.0063	2.89	0.0047	[56,57 58]
$La_3Ga_5SiO_{14}$ (X)[+]	2892		0.0042	0.64	0.0294	[56,57 58]
$Li_2B_4O_7$ (TA)[*]	2550	450	0.0078	1.32	0.0234	[59, 60]
Thin-Film Resonators						
ZnO (c-axis) (edge-spt'd)	2500[′]	7000	0.045	8	0.0218	[12]
AlN (c-axis) (edge-spt'd)	5000[′]	5000	0.032	9	0.0276	[12]
PVDF (TE)	1025	12	128	1.44	141	[61]
PZT/epoxy (TE)	1950	37	878	32.5	81.7	[61]
PZT 5A (TE)	2225	135	236	22.09	36.8	[61]
PZT/epoxy (LE)	1060	25	17	0.64	43.6	[61]
PZT 5A (LE)	1450	40	11	9	2.75	[61]
PZT (radial)	2300	300	53[$]	27	7.00	[64]

[+] Measured for the a-mode.
[*] Measured for the b-mode.
[′] Assuming a thickness of 5 μm.
[$] Assuming a frequency of 10MHz.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

MEASUREMENTS OF ACOUSTIC WAVE ATTENUATION IN $La_3Ga_5SiO_{14}$ USING HBAR TECHNIQUE

G.D.Mansfeld

Institute of Radioengineering and Electronics RAS
Mokhovaya 11, Moscow, 103907, Russia

The results of the investigation of the attenuation of bulk acoustic waves in one of the perspective material for acoustoelectronic - langasite are presented. The data of propagation losses in principal crystalline directions are obtained using recently suggested and developed method of HBAR acoustic spectroscopy.

Introduction

Langasite ($La_3Ga_5SiO_{14}$) is one of the perspective crystalline material for various applications in acoustoelectron and piezodielectric devices [1]. Most of physical properties of this material are already investigated [2,3] . This material seemed to have less acoustic losses than that of quartz [4] , but no systematic data about the attenuation were not known at the beginning of this work. The main goal of this report is to present recently obtained experimental data concerning acoustic losses in X, Y and Z principal crystalline directions in langasite. The samples were cut in a shape of thin flat parallel face plates which are usually used in bulk acoustic wave resonators. The most important parameter of these devices - quality factor depends on real losses in such structures. So it seems to be resonable to investigate acoustic losses using recently suggested and developed method of high overtone bulk acoustic wave resonator spectroscopy [5,6] . The total losses in structure consist of intrinsic (Akhieser and lattice) losses and losses connected with diffraction, nonparallelism of faces of the plate, roughness of their surfaces and other parasitic factors. In order to separate these losses the experiments at room and at liquid helium temperatures were carried out. Then intrinsic losses were found as a difference between these data. The physical nature and degree of contribution of parasitic factors in total losses are briefly discussed.

Experiment

Description of the method

The idea of the method of HBAR spectroscopy is based on frequency measurements of peculiarities of the phase versus frequency dependenses of multifrequency composite resonator structure shown on fig.1.

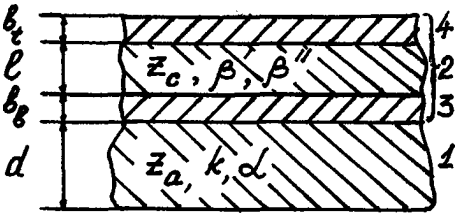


Fig.1. Schematic representation of the resonator structure: 1 - investigated layer, 2 - piezoelectric film, 3,4 - electrodes, k and β' - wave vectors, β'' and α - attenuation constants, z_c and z_a - acoustic impedances.

The structure consists of investigated layer (1) - in our case langasite plate and electromechanical transducer (2) made of piezoelectric with metallic contacts (3), (4). Such a structure has the frequency responses of impedance modulus and electromagnetic wave reflection coefficient modulus and phase as shown on fig.2 a,b,c correspondingly. As it was rigorously shown in [5,6] the difference between positions of peaks on phase vs frequency characteristics Δf_n (for f_n resonant peak) is directly connected with coefficient of attenuation. For high loss materials when it is possible to neglect losses connected with reflections and transducer layers (in piezoelectric and contacts)

$$\alpha = \pi \Delta f_n \ [s^{-1}]$$

0-7803-1945-1/94 $3.00 © 1994 IEEE

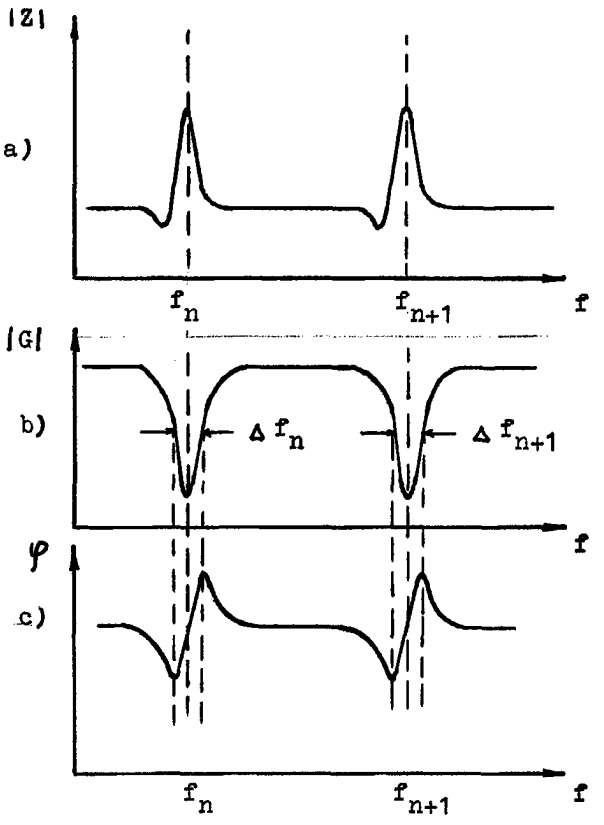


Fig.2. a) Modulus of electric impedance of the resonator structure as a function of frequency; b) and c) Absolute value and phase of electromagnetic wave reflection coefficient as a function of frequency.

In low loss material it is necessary to take into account some of these neglected factors

$$\alpha = \pi \Delta f_n - \beta'' \frac{\ell}{d} v_s - \Phi \frac{v_s}{Kd} \quad [s^{-1}]$$

β'' - attenuation constant in piezoelectric material used in transducer, ℓ - thickness of transducer, d - thickness of investigated layer, v_s - sound velocity in this layer, Φ - total losses connected with reflections from faces and nonparallelizm per one travel.

$$K = \gamma (\sec kd)^2 / [1 + \gamma^2 (\tan kd)^2]$$

γ is the ratio of the acoustic impedances of layer and transducer materials. The sound velocity may be found using measured frequency difference between two resonant peaks f_{n+1} and f_n, using the formula from [5]. In composite resonator structure resonant peaks are usually observed in a very wide frequency band up to few GHz. Distance between neighbour resonances depends on thickness of structure and is of an order of few or tens MHz. So with the same composite resonator structure one can measure frequency dependence of attenuation coefficient in a wide frequency band. So and it is very important the only frequency measurements are necessary to perform in order to obtain data on acoustic attenuation in material in wide frequency band.

The usage of the HBAR method is very informative for investigation of additional losses arising due to nonideal character of the reflections of acoustic waves from real surfaces. In contrast with echo-pulse method the amplitude and phase of measured signal in a much more degree governed by multiple reflections of acoustic waves between surfaces.

Samples

Samples were cut in the form of flat parallel faces plates with dimensions approximately equal: width - 2 mm, length - 3 mm, thickness - 0.2 - 0.8 mm. In the best samples the faces were parallel with angle less then 8". Plate orientation with respect to axis was better then 0.3°. Faces were optically polished (14 class). Density of dislocations as it was found experimentally in used material was not grater then 10^2. Conventional methods were used to fabricate electromechanical ZnO transducers. Aperture of transducers was varied from sample to sample from 200 to 600 μm correpondingly thickness of ZnO film in samples varied from 4 to 8 μm.

Procedure of measurements

Samples were mounted in a special holder for measurement of electromagnetic wave reflection coefficient from structure at room and liquid helium temperatures. A conventional type network analyzer was used for measurement of peculiarities of phase vs frequency characteristics. Special electromagnetic wave transformer was used to mismatch resonator structure and 50 Ohm transmission line in order to avoid a mistake connected with strong coupling between resonator and line in accordance with [6].

At first the frequency dependence of total losses were measured at room temperature and then at liquid helium temperature. The difference between these data (after substracting losses in material of transducer) was considered as intrinsic losses in material.

Results of the experiments

The experimentally found frequency dependence of attenuation coefficient of longitudinal acoustic waves in Z direction are presented on fig.3.

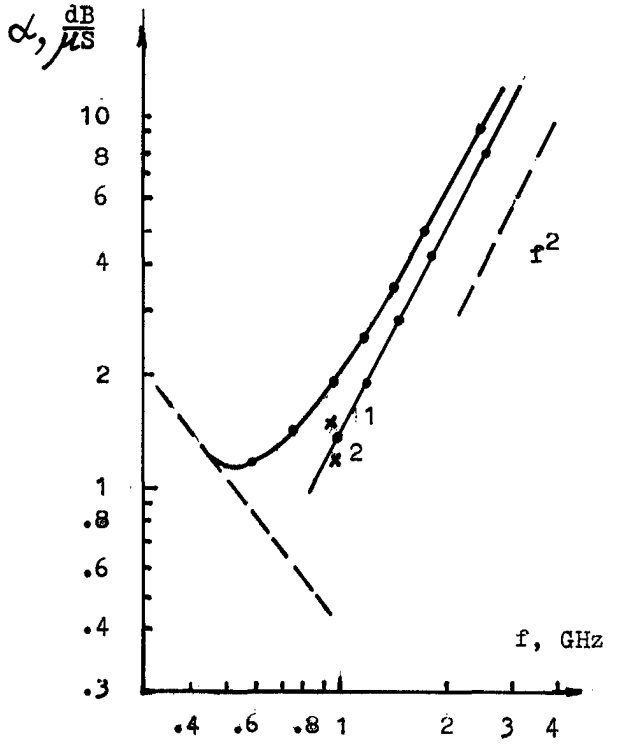


Fig.3. Frequency dependences of attenuation of longitudinal acoustic waves in Z direction. Upper curve - total losses, dashed curve - calculated diffraction losses. Lower curve - losses in material. Points 1, 2 - data for attenuation of longitudinal acoustic waves in quartz from [8] and [7].

The upper curve is the total measured losses, the dashed one is a diffraction losses calculated from [7]. Intrinsic material losses are represented by lower curve with a slope corresponding to Akhiezer's losses (f^2). Points 1 and 2 correspond to data for quartz [8,7]. As it follows from our experiments for longitudinally polarized acoustic waves in Z direction the losses in langasite are of the same order as that of quartz for the same direction.

For longitudinal wave in X direction the almost ideal f^2 dependence is also observed (see fig.4). The measured losses are almost coincide (within the limits of experimental error) with the data of [4] obtained by echo-pulse method (point 1). The losses of longitudinal wave in X

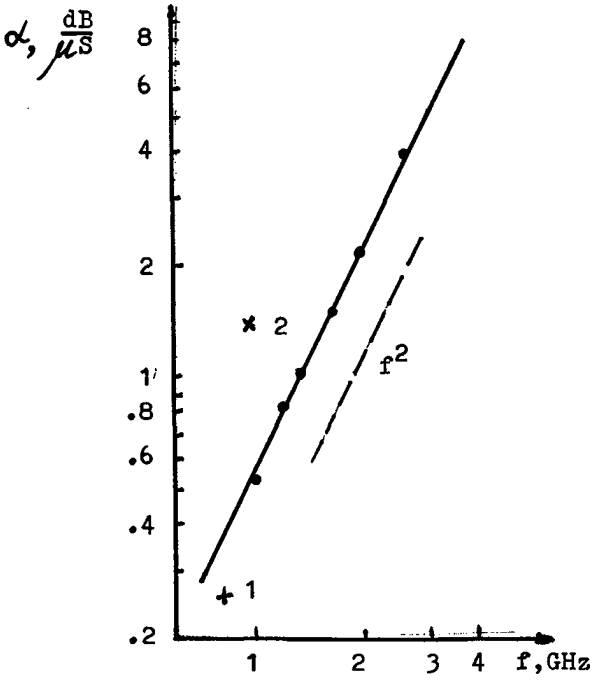


Fig.4. Frequency dependence of attenuation of longitudinal acoustic waves in X direction. 1 - data from echo-pulse measurements 4 ; 2 - data for quartz.

direction are noticeably less then that for X direction of quartz [8] (point 2).

Experimental data for Y direction are presented on fig.5,6 correspondingly for longitudinal and slow transverse waves.

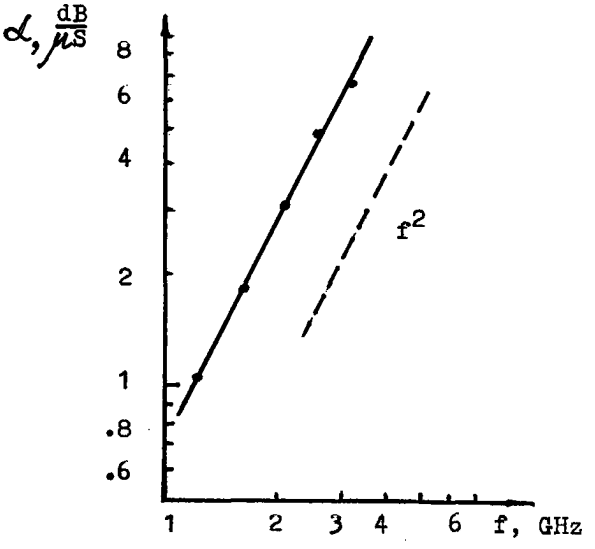


Fig.5. Frequency dependence of attenuation of quazilongitudinal acoustic waves in Y direction.

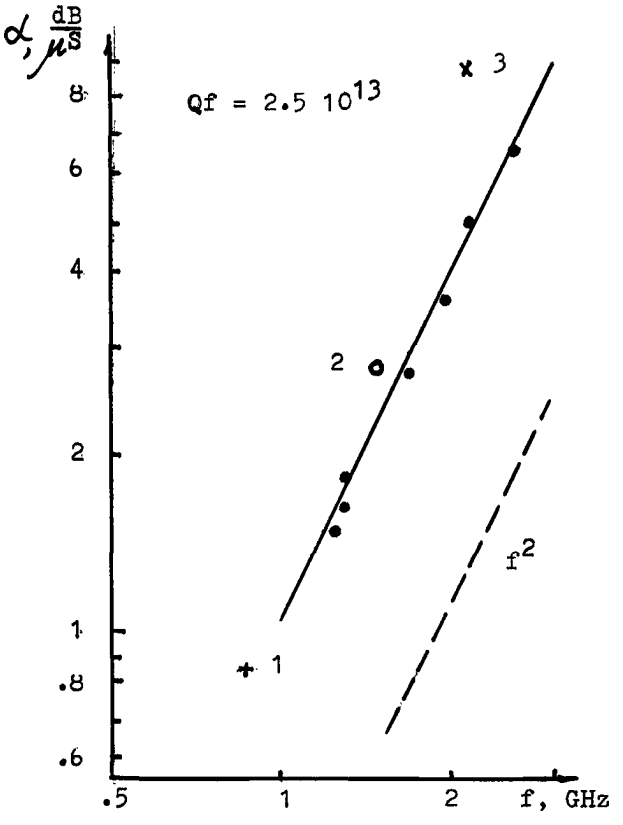


Fig.6. Frequency dependence of attenuation of slow shear acoustic waves in Y direction. 1 - data from echo-pulse measurements [4] ; 2,3 - our data for shear waves in BT and AT cuts of quartz.

In both cases square dependences of acoustic losses as a function of frequency were observed. The data for slow transverse waves (which are thermostable and hence interesting for practice) are in agreement with the echo-pulse data [4] and are somewhat less then our own data for BT (point 2) and AT (point 3) cuts of quartz.

The experimental results for 1 GHz point are presented in Table.

Wave direction and polarization	Velocity km/S	Attenuation dB/(μS·GHz^2)
X-longitud.	5.78±.01	0.52±.04
Z-longitud.	6.75±.01	1.30±.04
Y-quasilong.	5.79±.01	0.75±.05
Y-slow shear	2.77±.01	1.05±.05

Discussion of nature of additional losses

As it was found such factors as roughness and nonparallelism of faces led to higher total losses in structure and decrease of the Q-factor of resonator. The roughness of surfaces was experimentally measured using atomic force microscope technique.It was found that this part of additional losses became essential-mean square root deviation from ideal surfaces was of an order of 10 nm on frequencies higher than 1 GHz. These additional losses also square dependent on frequency and could be estimated experimentally only from comparison of room and liquid helium temperature data. It is evident that this kind of additional losses can easily be avoided by proper surface polishing.

The special numerical calculations were performed to estimate losses of energy arising from nonparallelism of surfaces. Usually these losses are almost frequency independent in used frequency band. When the angle between faces has the order of few tens of arc seconds the additional losses may become comparable with intrinsic losses in material and Q-factor of resonator may be drastically decreased. This conclusion was drawn from the experiments with many samples.

Diffraction effects also can increase the total losses in the structure. They can be taken into account using well developed formalism and in many cases be successfully avoided.

Conclusions

As it follows from obtained data the attenuation in LGS crystals is smaller or in some cases is of the same order as that for the same directions of quartz. The attenuation coefficients for measured directions of propagation and polarizations have the square dependences on frequency and hence are in agreement at least qualitatively with Akhiezer's theory. The used method let to evaluate additional losses arising from nonparallelism of faces and roughness.

Acknowledgements

Author is grateful to V.V.Bezdelkin ("Fonon" Institute), S.A.Sakharov ("Fomos" Company) for their help in obtaining high quality LGS crystals, I.M.Kotelyansky, E.S.Kucheryavaya, A.D.Freik (IRE RAS) for preparing samples and experimental measurements, C.Tellier (Ecole Nationale Supéreure de Mécanique et des Microelectroniques, France) for her help in AFM surface characterization.

References

1 J.A.Kosinski, "New acoustoelectronic materials", in Proceedings of the 6th Conference"Acoustoelectronics 93", Varna, Bulgaria, September 19 - 25, 1993, pp. 61 - 74.

2 J.G.Gualtieri, J.A.Kisinski, A.Ballato, "Piezoelectric material constants for surface wave research" in Proceedings of the Forum EFTF 93, Neuchatel, Switzerland, March 16 - 18, 1993, pp. 231 -234.

3 A.A.Kaminskii, I.M.Silvestrova, S.E. Sarkisov, G.A.Denisenko, "The investigation of trigonal $(La_{1-x}Nd_x)_3Ga_5SiO_{14}$ crystals, Phys.Stat.Sol.(a), vol.80, pp. 607 - 620, 1983.

4 A.A.Kaminskii et al., "Pure and Nd^{3+} -doped $Ca_3Ga_2Ge_4O_{14}$ and $Sr_3Ga_2Ge_4O_{14}$", Phys.Stat.Sol. (a), vol.86, pp. 345 - 362, 1984.

5 G.D.Mansfeld, A.D.Freik, B.N.Krutov, "High overtone BAW resonator as a tool for acoustic characterization of thin films", in Proceedings of the 6th Conference "Acoustoelectronics 93", Varna, Bulgaria, September 19 - 25, 1993, pp. 20 - 30.

6 G.D.Mansfeld, A.D.Freik, B.N.Krutov, "A high overtone BAW resonator as a tool for acoustic characterization of thin films and layers" , in Proceedings of the Forum EFTF 93, Neuchatel, Switzerland, March 16 - 18, 1993, pp. 261 - 266.

7 A.I.Morozov et al., Piezoelectric Transducers For Radioelectron Devices. Moscow: Radio i sviaz, 1981.

8 B.A.Auld, Acoustic Fields and waves in Solids. New-York, London, Sydney, Toronto: John Willey Interscience Publ., 1973, vol.1, p.96.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

STRESS-INDUCED OPTICAL ACTIVITY IN PIEZOELECTRIC CRYSTALS AND INTERNAL STRESSES METHOD OF CONTROL IN QUARTZ, LANGASITE, LITHIUM NIOBATE CRYSTALS

V.S. Naumov, I.I. Kalashnikova, * S.S. Pashkov

S & R Institute "Poljus", *J.S.C. "Phonon", Moscow, Russia

The optical activity change in quartz and langasite crystals and the appearance of gyration effect (rotation of the light polarization plane) in lithium niobate crystals under external load and internal stresses influence (stress-induced gyrotropy) have been investigated. The investigations have shown that internal stresses in crystals affect the gyrotropy in the same way, as the external load do. The possibility of internal stresses in crystals determination by gyration change has been discovered. Based on investigations made a method of internal stresses in these crystals has been developed.

In anysotropical crystal materials internal stresses can induce the change of anysotropical properties. These stresses can be in consequence of mechanical strains and inhomogeneity thermal actions which arise during crystal growth and mechanical treatment.

The cracking of crystals, sections and plates have taken place under mechanical treatment and other operations at a plants where piezoelectric devices have been manufactured. For example, the rejects during cutting process can attain 40% in stressed crystals. Therefore the problem of stresses' control in crystals is of considerable interest.

We have investigated the influence of external load and internal stresses on optical activity (rotation of the light polarization plane) in piezoelectric crystals: alpha-quartz, langasite, lithium niobate and discovered the appearance of stress-induced optical activity in these crystals.

Quartz and langasite crystals possess natural optical activity, lithium niobate crystals have not that. The optical activity in lithium niobate crystals appears under mechanical stress.

Induced optical activity is the appearance or change in optical activity (gyration) under the influence of external conditions: magnetic or electric fields [1], stresses, etc. Stress-induced gyrotropy is caused by the change in propagation direction velocities of anticlockwise and clockwise circularly polarized waves in crystals and their corresponding refraction coefficients n_l and n_r under mechanical stress action.

The investigated samples having a form of a rectangular parallelepiped were placed into a press equipped with a load sensor (pressure sensor). For stress-induced gyrotropy observation a linearly polarized laser light with the wave-length of 628 nm was passed along the optical axis of crystal which was placed between crossed polarizers. The load was applied along the crystallographic **X**- and **Y**-directions. The gyration change was determined by angular position of the polarizer-analyzer in comparison with the original position. The minimum transmission of the polarizer-crystal-analyzer was registrated by means the photodetector with the indicator.

The specific gyration (gyration angle at the unite of the crystal length) under mechanical stresses action ρ_3 can be represented by expansion into a series by limiting quadratic terms only:

$$\rho_3 = \rho_3^0 + \gamma_{3ij}\sigma_{ij} + \beta_{3ijlm}\sigma_{ij}\sigma_{lm} \quad (1)$$

where: ρ_3^0 - is specific gyration with no stresses, determined by g_{33} components of the axial tensor of rank two; at λ = 628 nm in quartz ρ_3^0 = 18,8 degree/mm, in langasite ρ_3^0 = 2,65 degree/mm, in lithium niobate ρ_3^0 = 0;
σ_{ij} ; σ_{lm} - the components of stress tensor;
γ_{3ij} ; β_{3ijlm} - the stress-induced gyration coefficients, which are in common case the components of the tensor rank four or six respectively.

In fig.1,2,3 experimental dependences of the change of gyration angle (at the unit of length), along the optical axis in quartz, lithium niobate and langasite crystals respectively under uniaxial stresses along axis <X> and <Y> influence are represented at the logarithmic scale.

The measuring results (fig.1, fig.2) have shown that effect of induced gyrotropy under uniaxial stress along X-axis in quartz and lithium niobate crystals manifests stronger than under uniaxial stress along Y-axis.

0-7803-1945-1/94 $3.00 © 1994 IEEE

TABLE
Stress-induced gyration coefficients

Crystal	β_{311}	β_{322}	γ_{31}	γ_{32}
	degree·mm^3· kgf^{-2}		degree·mm · kgf^{-1}	
quartz	2,8	0.92	-	-
lithium niobate	27	20	-	-
langasite	-	-	0,47	0,66

The change of specific gyration quadratically depends on load value (fig.1), the tangent of slope angle is equal to 2. This allowed to find the quadratic coefficients β_{311} and β_{322} (1) for this crystals. The linear coefficients γ_{31} and γ_{32} (1) can be considered as zero. The values of these coefficients are given in the table.

The rotation direction in quartz crystals is related with morphological enantiomorphism. The crystals with right morphological properties have the right gyration, the crystals with left morphological properties - the left one.

The anticlockwise rotation of the light polarization plane in the light beam propagation direction is taken positive (as for right-handed quartz).

The curves (fig.1,2) have shown that uniaxial stress along X- and Y-axis for right-handed quartz and lithium niobate causes gyrotropy of opposite sign (left and right). The values of quadratic coefficients are 10 - 20 times more in lithium niobate crystals than that in quartz crystals.

The measuring results (fig.3) have shown that effect of induced gyrotropy under uniaxial stress along Y-axis in langasite crystals manifests stronger than under uniaxial stress along X-axis. The change of specific gyration versus the stress value is a linear dependence (fig.3) in this stress region that allowed to find the linear coefficients γ_{31} and γ_{32} (1) (see tab.). The quadratic coefficients β_{311} and β_{322} for langasite crystal can be considered as zero in this stress region. Stress-induced gyration for langasite occurs in clockwise direction that reduces the natural specific gyration ρ_3^e (1).

All langasite crystals were investigated have had right natural gyration (as right-handed quartz). The value of natural gyration of langasite crystals is in 6-7 times less than this of quartz crystals.

The langasite crystals belongs to the space group of symmetry P321 - D_3^2 [2].

Sometimes in literature the symmetry of langasite crystals is believed the same as symmetry of alpha-quartz crystals due to this crystals belong to common point group 32. However the space group of symmetry of alpha-quartz is $P3_12$ - D_3^4 or $P3_22$ -D_3^5, but of langasite - P321 - D_3^2.

There were not found the langasite crystals with an orderly arrangement of Ga and Si ions that could offer to decrease of the space group of symmetry to P3 (as for α-SiO_2) [2].

From the crystal-chemical point of view the space group P321 does not assume the existence of morphological enantiomorphic forms.

The change of the crystal symmetry and in consequence the change of the symmetry of the dielectric tensor ε_{ij} under stress influence is a main reason of the appearance or change its optical activity. These optical properties can be explained from macroscopic consideration of the dielectric tensor ε_{ij} with a spatial dispersion [3].

The increase of load in quartz crystal (more than 1,0 kgf/mm^2) don't bring to change the form of curves. The increase of load in lithium niobate crystal (more than 0,15 kgf/mm^2) results in change and complication of the curve form. This complication can be caused by the another factors influence such as the movement of the domain boundaries (reconstructure of the crystal domain structure), by the effect of piezoelectricity and its electric fields.

The increase of load (more than 1,0 kgf/mm^2) in langasite crystals results in weaker change of the curves form than in lithium niobate crystals.

The different compressibility of these crystals along X-and Y-axis can give the different values of stress-induced coefficients and distinct dependence of the specific gyration versus stress.

The possibility of internal stresses in crystals determination by gyration change has been studied. The investigations have shown that internal stresses in

crystals affect the gyrotropy in the same way, as the external loads do. Based on the investigations made a method of internal stresses in quartz, langasite and lithium niobate crystals has been developed. This method is based on the registration of the gyration change along the optical crystal axis, what enables to evaluate the internal stresses. The observation of stress-induced gyrotropy along the optical axis is preferred because it is not masked by optical birefringence.

This method allows to determine the sign of stresses (tension or compression). There is compression along optical axis of quartz crystals if the additional rotation is carried out in clockwise direction about the original analyzer position (for nonstressed crystal) for right-handed crystal and in anticlockwise direction for left-handed crystal. There is tension along optical axis when the additional rotation is carried out in anticlockwise direction for right-handed quartz crystal and in clockwise direction for left-handed quartz crystal in the light beam propagation direction.

As for as its sensitivity to small mechanical stresses (to 10^{-3} kgf/mm^2) the method proposed can be comparable only with the interferometry methods, however, it is considerably simpler in realization and interpretation of the results. Moreover this method allows to control stresses in the samples with lapped surfaces and does not required high precision in the plan-parallelism of samples as in the case of the interferometry method.

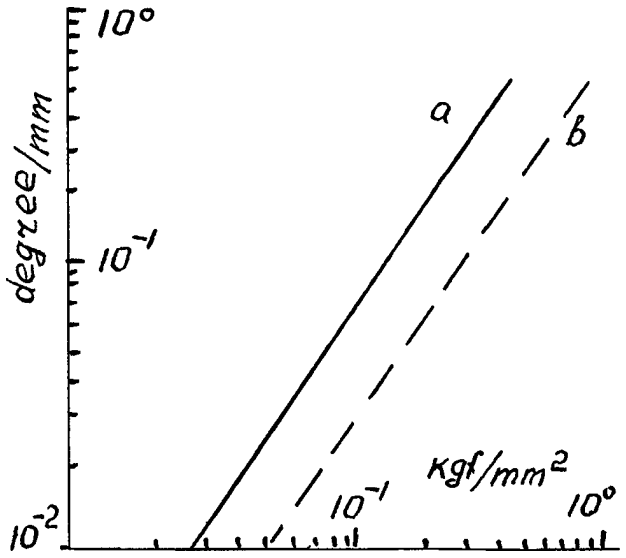


Fig.1. The change of specific gyration versus uniaxial stress in quartz crystal.
a - along X-axis, b - along Y-axis.

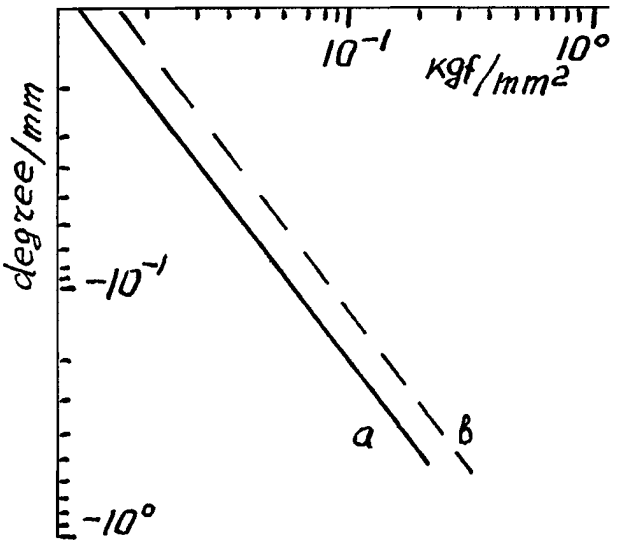


Fig.2. The stress-induced specific gyration versus uniaxial stress in lithium niobate crystal.
a - along X-axis, b - along Y-axis.

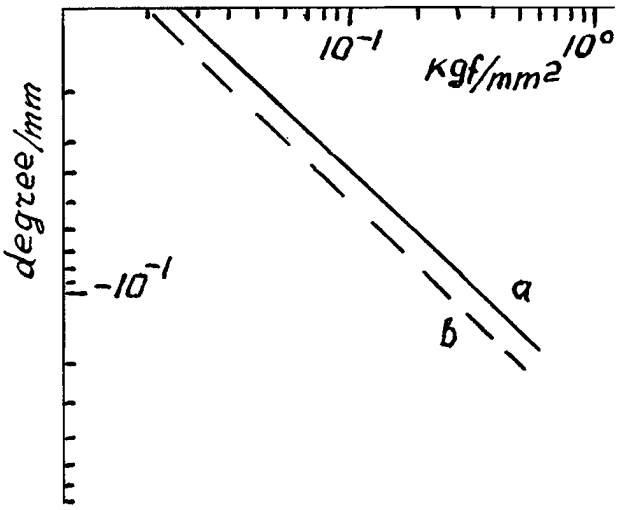


Fig.3. The change of specific gyration versus uniaxial siress in langasite crystal.
a - along X-axis, b - along Y-axis.

REFERENCES

[1] I.S.Zheludev. The optical activity in crystals under electric field influence. Sov. Uspehi of Phys. Nauk, v.120, № 4, p.p.702-703, 1976.

[2] B.V.Mill, A.V.Butashin, G.G.Hodzhabagjan, E.L.Belokoneva, N.V.Benlov. Modificated rare earth gallats with Ga Ga Ge O structure. Sov. "Dokladi Acad. Nauk", 1982, v.264, № 6 p.p.1385-1389, 1982.

[3] V.M.Agranovich, V.L.Ginsburg. Crystal optics with space dispersion and the theory of exitons. Moscow, Nauka. 1979.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

LANGASITE ($La_3Ga_5SiO_{14}$) AN OPTICAL PIEZOELECTRIC:GROWTH AND PROPERTIES

M.F.Dubovik, I.A.Andreyev, Yu.S.Shmaly

Institute for Single Crystals of Acad.Sci. of Ukraine,
60 Lenin Av., 310001,Kharkov,Ukraine

Abstract

Langasite ($La_3Ga_5SiO_{14}$) is a new effective piezoelectric with $k \simeq 0.17$ and a structure analogous to that of quartz(32 symmetry class). The single crystals with a diameter of about 60 mm and a mass reaching 1 kg were grown by the modified Czochralski method in direction $[2\bar{1}\bar{1}0](x)$. The monophase raw material was prepared by the solid phase method. Determined were the main parameters of the crystals' quality and the functional characteristics of thermostable langasite elements. The basic characteristics of these elements were found to compare favourably with those of quartz. In particular, their resonance interval are 2-3 times as large, the dynamic inductance and dynamic resistance for VAW decrease by 6-18 and 2-6 times, respectively, the squared electromechanical coefficient grow by 2-3 times, the losses by SAW propagation drop by 2 times, the dimensions of LGSO elements being smaller by 20-30 %.

Introduction

Crystalline quartz which is a weak piezoelectric cannot comply with the stringent requirements of modern piezo- and acoustoelectronics. Lithium tantalate ($LiTaO_3$) and berlinite ($AlPO_4$) the effective piezoelectrics with zero temperature frequency coefficient (TFC) are not promising due to their complicated growth methods and high cost.

Lantanum gallium silicate $La_3Ga_5SiO_{14}$, i.e. langasite (LGSO) is a new effective piezoelectric, a structural analog of α-quartz (32 symmetry class). For the first time it was obtained in the USSR as an active laser medium /1,2/. As shown in our investigations /3,4/, it has cuts with zero TFC.

Presented here are the results of studying the process of raw material synthesis, the improvement of crystal growth procedure as well as the main physical and technical parameters of LGSO crystals in view of their possible practical use.

Raw material synthesis

As shown in /1/ LGSO is the only triple compound in La_2O_3-Ga_2O_3-SiO_2 oxide system which melts congruently ($T_m \simeq 1743$ K) and has no phase transitions at temperatures higher than room one. However in the first series of experiments on the growth of LGSO by the Czochralski method the deviations from the stoichiometric composition and the presence of blocks were observed in the crystals. To some extent this was conditioned by the presence of unreacted components in the raw material and their interaction with the crucible material.

The raw material used in our experiments was prepared from the stoichiometric mixture of the high-purity components La_2O_3, Ga_2O_3 and SiO_2. The temperature and time conditions were chosen on the base of differential thermal analysis (DTA), X-ray diffraction analysis and the measurement of the resistance R of tabletted LGSO and the mixtures of some of its components.

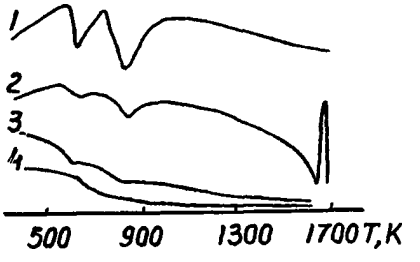


Fig.1. Derivatographic curves for La_2O_3 and $La_3Ga_5SiO_{14}$ (LGSO): DTA La_2O_3 (1); DTA LGSO (2); weight loss curve LGSO (3); weight loss curve La_2O_3 (4).

The DTA curves (Fig.1) are characterized by the presence of two endothermal effects at 633 and 803 K, one endothermal effect at 663 and one exothermal effect at 1683 K. The first two effects seem to be caused by the dehydration of lanthanum oxide, the third and the fourth effects are conditioned by the formation of the liquid and LGSO phase, respectively.

As seen from the expansion-shrinkage curves (Fig.2), the size of the samples sharply changes starting from 1373 K and this increment reaches 9 % at 1573 K. Within 623-823 K interval a change in the

0-7803-1945-1/94 $3.00 © 1994 IEEE

samples' resistance is also observed.

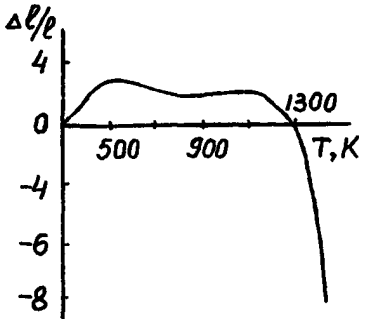


Fig.2. Expansion-shrinkage curve of LGSO.

The behaviour of the R(T) and expansion-shrinkage curves points to the fact that at temperatures up to 1373 K no evident interaction takes place even in the surface-adjacent layers of the samples. The investigation of mechanism solid phase interaction by the method of diffusive annealing, as well as the study of mass transfer by inert (platinum) mark method show that Ga^{3+} and Si^{4+} ions are the most "mobile" reactants.

The performed investigations allowed to optimize the synthesis of the monophase raw material with the deviation of the main phase content from the preset value not exceeding 0.1 mass %, the contents of the controlled Al, Cr, Fe, Mn, Cu, Ni, Pb, Mg impurities being not larger than 10^{-5} - 10^{-4} mass %.

Crystal growth

The congruence of melting permitted to grow the crystals by the well-known Czochralski method. However the anisotropy of their structure imposes certain requirements to the orientation of the seed and thermal conditions.

The growth direction along the third order axis $C_3(Z)$ is the most preferable /1,2/. Nevertheless, in view of practical application the growth is to be realized along the piezoelectric effect axes [2$\bar{1}$10] (x) and [10$\bar{1}$0] (y). Taking into accounts the crystal growth anisotropy ($V_z \simeq 2V_y < V_x$ /1,2/) and consequently possible cracking of the crystals, it is necessary to ensure the axial symmetry of thermal field in the process of both crystal growth and cooling. In the case under consideration this is achieved by means of a special design crystallization unit /5/. The unit consisted of a platinum crucible shaped as cylinder with a diameter d_c almost equal to its height $H_c \simeq 100$ mm, a cylindrical platinum screening heater which were heated by radio frequency currents ($f \simeq 8$ kHz), as well as heat insulating corundum screens. The unit also comprised coaxial cylindrical ceramic rings with symmetric holes placed between the crucible and the heater. During the crystal growth the cross-section of the rings' holes were varied to create certain temperature gradients in the thermal zone.

The monophase character of the raw material allowed to grow the crystals in air on x-oriented seeds using the modern apparatuses "Kristall-3M" and "Skif-5" produced in Ukraine. An automated system for controlling the main parameters (i.e. the crystal's diameter, pulling rate, rotation frequency as well as electrical power supplied to inductance coil) maintained their stability to an accuracy of about 0.1 %. This permitted to obtain rather homogeneous crystals with a diameter of 60-70 mm and a mass reaching 1 kg (Fig.3).



Fig.3. LGSO crystal and oriented samples.

The as-grown crystals were annealed in the crystallization unit in a special furnace at $T \simeq 1623$ K.

Crystal quality

The defects usually observed in the crystals grown by the Czochralski method, i.e. volume defects localized in near the crystal axis and transverse growth lamination were not pronounced in the majority of cases. The dislocation density determined from the member of etch pits in the plane z (0001)z-cut, did not exceed 10^3 cm^{-2}, the anomalous biaxiality angle $2V \leqslant 90^1$, the changes of the ordinary and extraordinary refractive indices were the following: $\delta n_o \leqslant 4 \cdot 10^{-5}$ cm^{-1}, $\delta n_e \leqslant 7.4 \cdot 10^{-4}$ cm^{-1}. No other phases were revealed within the limits of X-ray diffraction analysis sensitivity (0.1 mass %). However present in some parts of the crystals were microblocks with a disorientation of 3-5 angular min.

The crystals were amber coloured. Their absorption spectra were characterized by the presence of a faint wide line in 450 nm wavelength region (Fig.4). The d.c. resistance $R_z \geqslant 10^{13}$ Ohm·cm, the tangent of dielectric loss angle $tg\delta_z \leqslant 3.6 \cdot 10^{-4}$. The high quality of LGSO crystals manifest itself in a number of expe-

rimental samples (the plates of x- and z-cuts).

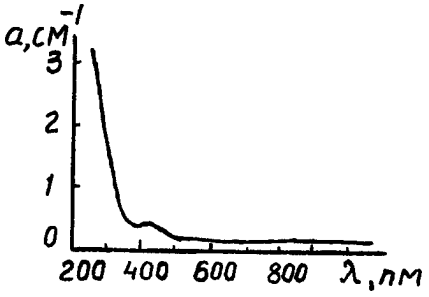


Fig.4. Absorption spectrum of LGSO.

The crystals under consideration are rather thermostable and nonhydroscopic. As shown by experiments performed under standard conditions, the oriented and polished samples of LGSO crystals measuring (5-10)·10·(15-20) mm^3 withstand vibration with f = 100-200 Hz under 5-20 g (g= 9.8 m/s^2) acceleration during 24 hours, impact load under 40 g and 500 g acceleration (1200 and 3 impacts, respectively), the action of temperatures ranging from 233 to 373 K and moisture (98 %) at 313 K during 10 days.

Main properties of langasite crystals and articles based on them

Since LGSO and quartz belong to the same symmetry class, the similarity of their main physical properties should be expected, piezoelectric characteristics being the most important. As shown in /2/ the piezoelectric moduli of LGSO d_{ij} and e_{ij} and the elastic modulus c_{ij} exceed those of quartz by 2-6.3 and 2.2-15 times respectively, the attenuation of ultrasound in LGSO at f 900 MHz is lower by 2-5 times.

The aim of our investigations was to study the thermal stability of the main LGSO parameters and to search for cuts with TCF=0 for volume and surface acoustic waves (VAW and SAW, respectively).

The temperature coefficients of electromechanical constants c_{ij}, e_{ij}, the dielectric permeability ε_{ij}, the thermal coefficient α_{ij} and the acoustic wave velocity at temperatures ranging from 196 to 403 K and frequencies f varying from 50 to 10 MHz were measured for the first time /6,7/.

Calculated were the values of the rotation angles φ and γ as well as the coefficients of electromechanical coupling factor K corresponding to the case of zero thermal velocity coefficients (TVC) (yxbl/γ/φ) cuts /6/. In particularly, for a thickness-shift mode (c) K_{max} = 0.17 at φ = -12.7°, γ =0° for 1 order of TVC and φ =-0.9°, γ =0° for 2 order of TVC.

Experimental samples of VAW resonators different types of vibrations were made /7/. So, the main parameters of the thickness-shift resonators: the resonance interval $(f_a-f_r)/f_r$ = 0.5 %, the capacitence ratio C_o/C_k=100, the dynamic inductance L = 0.026 H, the dynamic resistance R = 5 Ohm, the frequence variation $\Delta f/f$ = $-3\cdot10^{-4}$ in a interval of the temperatures T=213-353 K for the frequency f=3000kHz. For the longitudinal, countour, bending resonators these parameters are equal: $(f_a-f_r)/f_r$=0.83; 0.25; 0.37, C_o/C_k = 60; 200; 130, L = 3.5; 1.2; 13 H, R = 100; 75; 750 Ohm, $\Delta f/f$= -3.5; -3.5; $-3.2\cdot10^{-4}$ for f=128; 512; 123 kHz, respectively.

The main peculiarities of the properties of LGSO resonators are the following: the TFC of all the modes are parabolas with C = $-(50-70)\cdot10^{-9}/K^2$; the temperature of TFC extrema (T_o) may be easily controlled by changing the angle of the cuts and the geometry of piezoelectric elements within a wide temperature range; the requirements to the orientation accuracy of LGSO cuts are less stringent as against quartz; TFC for bent and longitudinal modes with the same T_o are realized in the cut xys/φ, the angle φ being unchanged; on passing from thickness-shift mode to the third overtone T_o shifts by 323 K, the coefficient C decreases from $-60\cdot10^{-9}/K^2$ to $-50\cdot10^{-9}/K^2$; in LGSO piezoelectric effect is more prononounced as against quartz and provides the improvement of all the parameters of piezoelectric elements, in particular, the resonance interval $(f_a-f_r)f_r$ increase by 2 or 3 times, the dynamic inductance L and the dynamic resistance R_o decrease by 6-18 and 2-6 times, respectively; the lower sound velocity in LGSO allows to reduce the dimensions of piezoelectric elements by 20-30 % in comparison with those of quartz and lithium tantalate elements; the absence of ferroelectric properties in LGSO increases the temperature and time stability of dynamic and static parameters by 5 times; the Q-factor of LGSO piezoelectric elements $Q \geqslant 10^4$, their electric strength is 4 times as high in comparison with lithium tantalate elements; the introduction of some additions (e.g. Al_2O_3 or TiO_2) improves the main parameters of piezoelectric elements, e.g. for LGSO:Al and LGSO:Ti longitudinal vibration resonators the value of L decreases by 5-7 % in 125-130 kHz range, Q increases by 12-18 % as against the corresponding parameters of the resonators made of pure LGSO crystals.

LGSO piezoelectric resonators meant for SAW also compare favourably with quartz resonators. As shown in /8/, the squared electromechanical coupling coefficient (K^2) for SAW may exceed that of quartz by

10 times. So, for x-cut $(2\bar{1}\bar{1}0)$ the Eulerian angles $\theta=90°, \varphi=0°, \alpha=-6°$ and K^2 = 0.477, the phase SAW velocity on free surface V_f = 2382.8 m/s; for y-cut $(10\bar{1}0)$ $\theta = \varphi = 90°, \alpha=0°$ and K^2 =0.355, V_f = 2287.82 m/s.

Manufactured were thermostable SAW delay lines on LGSO substrates (y'x and x'y cuts). For these samples we obtained the two- and three-fold increase of K^2, the losses by the propagation of SAW with f=88.8 MHz being half as much in comparison with those for the ST-cut of quartz (Table 1) /9/.

Another important effect revealed in LGSO crystals is a strong eletrostatic interaction observed while studying the resonance frequency of longitudinal mode excited in the plates of xys/α cuts and depending on the direction of the static electric field $E_=$. The frequency was found to be a linear function of $E_=$ and temperature within the whole range of the angles α (Fig.5,6)/10/.

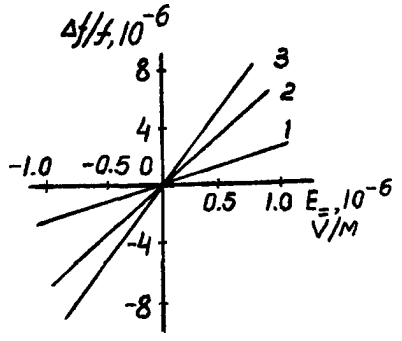


Fig.5. Frequency variation for various LGSO cuts depending on external controlling field: 1 - 0°; 2 - 10°; 3 - 15°.

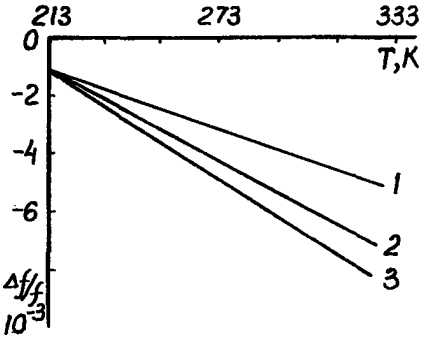


Fig.6. Frequency variation of various LGSO cuts as a function of temperature: 1 - 0°; 2 - 10°; 3 - 15°.

Calculated were the efficiency of such an interaction $Q^2 = \gamma^2/c\varepsilon^2$, the frequency control coefficient $\gamma = 1/f \cdot \Delta f/\Delta E$ and the corresponding temperature coefficients $T_f = 1/f \cdot \Delta f/\Delta T$. The obtained values compared with those for other crystals are presented in Table 2.

As seen from the Table 2, LGSO crystals are promising for the effective control of elastic vibration frequency or sound velocity. They also may be used for the parametric amplification of elastic waves especially at frequencies larger than 100 MHz (in this case sound attenuation coefficient is a factor of 2-5 lower in comparison with that for quartz).

Conclusion

Langasite is a promising material for piezo- and acoustoelectronics owing to the presence of a pronounced piezoelectric effect and first order zero TFC, low sound attenuation, strong electroacoustic interaction, temperature and time stability of its parameters.

A wide practical use of langasite is defined by a rather simple method of growing large crystals of different orientation, the possibility to manufacture various articles from this material as well as its lower cost as against lithium tantalate or lithium niobate.

The growth technique for langasite single crystals, the method of their modification by introducing aluminium and titanium additions, as well as the technique of manufacturing piezoelecric elements for various types of vibrations are protected by a number of patents beginning in 1984.

Table 1
Main parameters of SAW substrates made of LGSO and other materials

Substrate	V*, m/s	K^2	δ**, $B \cdot cm^{-1} \cdot GHz^2$
ST-SiO_2	3160	0.15	1.7
y'x-LGSO	2790	0.35	0.8
x'y-LGSO	2790	0.45	0.8
xz-$Li_2B_4O_7$	3510	1.0	-
yz-$LiTaO_3$	3250	1.1	0.25
yz-$LiNbO_3$	3490	4.8	0.3
yx-$AlPO_4$	2750	0.25	48

*SAW velocity
** δ sound attenuation

Table 2

Electrostatic interaction parameters for LGSO and other types of crystals

Crystal	T_f, 10^{-6} K^{-1}	γ, 10^{-12} m/V	Q^2, 10^{-12} m/J
LGSO	60	200	10
SiO_2	25	15	0.4
$LiTaO_3$	35	60	0.3
$LiNbO_3$	95	80	0.7
$Bi_{12}GeO_{20}$	140	130	80

References

/1/. A.A.Kaminski, G.G.Khodzhabagyan, A.F.Konstantinova et al, Investigation of trigonal $(La_{1-x}Nd_x)_3Ga_5SiO_{14}$ crystals, Phys.stat.sol.(a),1983,vol.80,pp.387-398.

/2/. A.A.Kaminski, B.V.Mill, I.M.Silverstova et al, Nonlinear active material $(La_{1-x}Nd_x)_3Ga_5SiO_{14}$, Izv. AN SSSR, ser. fiz.,1983,vol.47,pp.1903-1908.

/3/. I.A.Andreyev, M.F.Dubovik, Langasite ($La_3Ga_5SiO_{14}$), a material with zero temperature coefficient of elastic vibration frequency, Pis'ma v ZhTF,1984, vol.10,pp.487-491.

/4/. A.B.Ilyaev, B.S.Umarov, M.F.Dubovik et al, Temperature dependence of electrochemical properties of LGS crystals,Phys. stat.sol.(a),1986,vol.98,pp.K109-K114.

/5/. M.F.Dubovik, Growth and properties of some oxide single crystals, Neor.materialy. Acad.Sci. of Russia, 1992, vol.28,pp.1999-2006.

/6/. I.A.Andreyev, M.F.Dubovik, Langasite a promising material for thermostable decies for VAW and SAW, in Proceedings of III All-Union Conf. "Problems of obtaining and use of ferro-and piezoelectrical materials", Moscow, NIITEKHim publ., 1987, p.143.

/7/. I.A.Andreyev, M.F.Dubovik, Modified langasite, a base for HF piezoelectronic devices, in Proceedings of XIV All-Union Conf. on acoustoelectronics and physical acoustics of solids, Kishinev, 1989, Part II, p.67.

/8/. A.A.Blistanov, V.S.Bondarenko, A.A.Perelomova et al, Anisotropy of the parameters of the surface acoustic waves in new crystal $La_3Ga_5SiO_{14}$, in Proceedings of II All Union Conf. "Problems of obtaining and use of ferro- and piezoelectrical materials", Moscow, NIITEKHim publ., 1984, Part.II, p.266.

/9/. I.A.Andreyev, M.F.Dubovik, Main parameters of HF devices for VAW and SAW based on langasite, in Proceedings of XIV All-Union Conf. on acoustoelectronics and physical acoustics of solids, Kishinev,1989,Part II, p.69.

/10/. I.A.Andreyev, M.F.Dubovik, Strong electroacoustic interaction in langasite crystals, in Proceedings of XIII All-Union Conf. on acoustoelectronics and quantum acoustics, Kiev, 1986, Part II, pp.143-144.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

STUDY OF GALLIUM PHOSPHATE AND LANGASITE CRYSTALS AND RESONATORS BY X RAY TOPOGRAPHY

B.CAPELLE*, A.ZARKA*, J.DETAINT**, J.SCHWARTZEL**, A.IBANEZ***, E.PHILIPPOT***, J.P.DENIS****

*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
***L.P.C.M.S. Université de Montpellier, CNRS-URA D0407, case 03, 34095 Montpellier France
****CRISMATEC. 2 rue des Essarts, ZI des Mayencins, 38610 Gières France

Abstract

The gallium phosphate ($GaPO_4$) is, as berlinite, a close crystallographic analogue of quartz with the advantage that it has no α-β phase transition. It was shown that this material has very attractive properties for the applications to piezoelectric devices, due to its large coupling coefficients and to the existence of compensated cuts [1,2]. Furthermore, it can be used up to extremely high temperatures (900°C). To obtain much larger crystals, epitaxial growth of $GaPO_4$ on berlinite plates with several orientations was performed [3].

The langasite crystals (LGS: $La_3SiGa_5O_{14}$) were obtained by the Czochralsky method. This new material has probably one compensated cut with a large coupling coefficient and a very reduced angular sensitivity.

Using the synchrotron radiation delivered by the DCI storage ring at the LURE (Orsay, France), we have studied by the X-ray topography technique the crystalline perfection of new $GaPO_4$ samples and different langasite crystals of different sources. For this study we have used traverse and section topography .

Introduction

The gallium phosphate and the langasite belong to the crystal class 32 as quartz and berlinite. the gallium phosphate is a close crystallographic analogue of quartz with the advantage that it has no α-β phase transition and then it can be used up to high temperature (900°c). This material being obtained by the hydrothermal growth method, seeds are necessary. To obtain rapidly much larger crystals, epitaxial growth on berlinite plates was performed. With seeds cut in the X zone of such crystals a new growth in phosphoric acid gives large crystals.

The langasite crystals were obtained by the Czochralsky method, and then there is no problem of seed.

For the study of these two materials we have used traverse and section topography with the synchrotron radiation at Lure in Orsay.

Properties of materials

The most important properties related to the crystal growth of these materials are summarised in the table 1 with those of quartz and berlinite.

Quartz and its analogous, berlinite and phosphate gallium, have phase transitions below their melting point. They display non negligible solubility in both alkaline and acid aqueous solutions at elevated temperature and high

Material	Phase Transition	Solvent	Growth Method	Temperature (°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.

0-7803-1945-1/94 $3.00 © 1994 IEEE

pressure. So that the natural choose 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 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.

The figures 1a and b show the typical quality observed for quartz and berlinite. In a the topograph is related to a Y plate, of about 1 mm thick, of synthetic quartz grown by SICN. The observations that can be made in the topograph are representative of the most frequently encountered crystalline quality. The dislocation density is about 10 per 1 mm^2, and is nearly constant from the seed to the end of the Z growth sector. Most of the dislocations are the prolongation of those existing in the seed, very few are created at the surface of the seed or due to inclusions in the Z growth sector. Quartz crystals having a much reduced dislocation density, a few per cm^2 and even zero dislocation, can be grown but, to our knowledge, there is presently no steady production of such crystals. The topograph in the figure 1b is relative to an AT plate of berlinite grown at SICN. It represents also the usual present quality of berlinite crystal. Some fluctuations can be observed at the growth restart, but the dislocation density is of the same order as for the quartz sample. Some surface defects due to the polishing process can be observed together with contrasts due to the mounting clips and to the edges of the electrodes deposited on the plate.

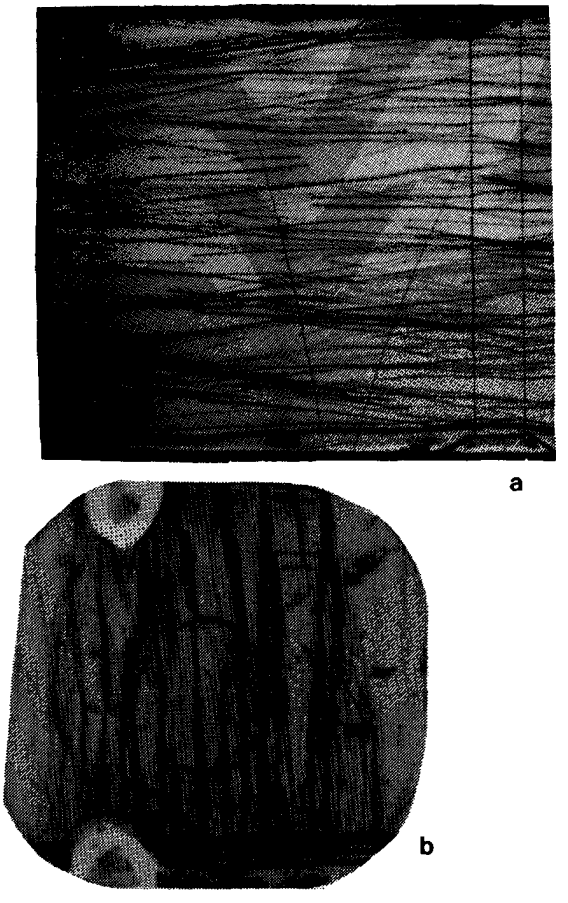


Figure 1-This figure shows the typical quality observed for quartz in **a** and berlinite in **b** grown by SICN.

Gallium phosphate

The gallium phosphate crystal in figure 2b, obtained from a Z berlinite seed in sulphuric acid and VTG method, has only the growth rate along X significant. The figures 2a and 2c show two

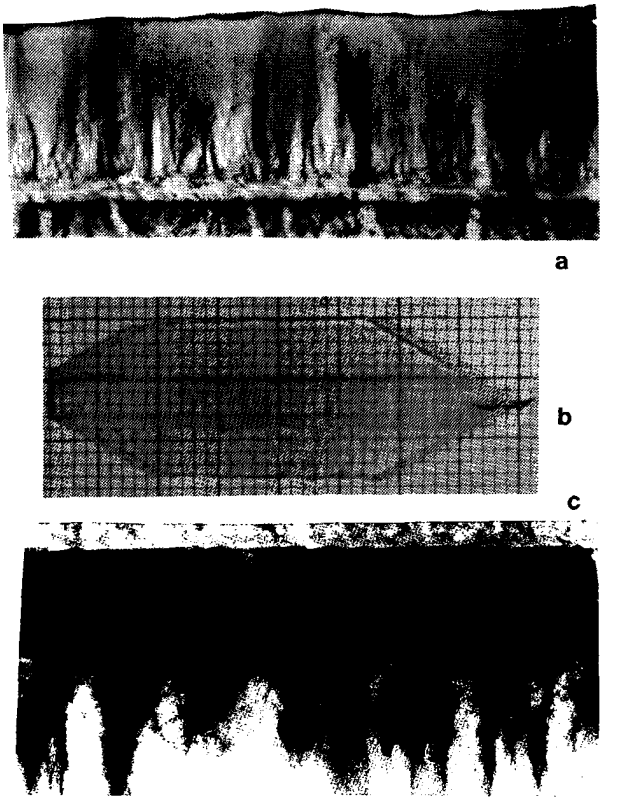


Figure 2-This gallium phosphate crystal in **b**, obtained from a z berlinite seed in sulphuric acid and VTG method, has only the growth rate along X significant. In **a** and **c** two traverse topographs of this crystal.

traverse topographs of this crystal, in a the top of the berlinite seed and the X zone appear, in c the berlinite seed and a part of the $\bar{X}$ zone. In a, strains, which decrease when the distance to the seed increases, appear at the interface between the berlinite seed and the gallium phosphate.
To analyse this epitaxial growth, section topographs have been made with a fine vertical slit.

An example of a section topograph obtained with this crystal is shown in the figure 3. The image of the seed appears very well on the left, the images of the two gallium phosphate layers are not separated because the strong strains give over large images. On the top the gallium phosphate of the X zone becomes better progressively far from the seed. At the interface between the seed and the X zone no image appears showing that there is an amorphous zone. Alone the epitaxial gallium phosphate seems to be unbroken.

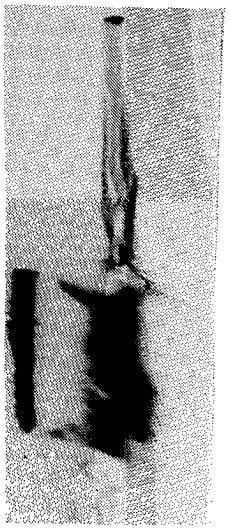


Figure 3-An example of section topographs obtained with the crystal of the previous figure. The image of the seed appears very well on the left, the images of the two gallium phosphate layers are not separated because the strong strains give over large images.

The three sections topogoraphs of another crystal, in the figure 4, have been recorded simultaneously with different diffraction vectors in the same Laue pattern. In each of these topographs three zones are visible. The central zone is composed of the image of the berlinite seed which appears as a black line and the images of the layers of gallium phosphate which

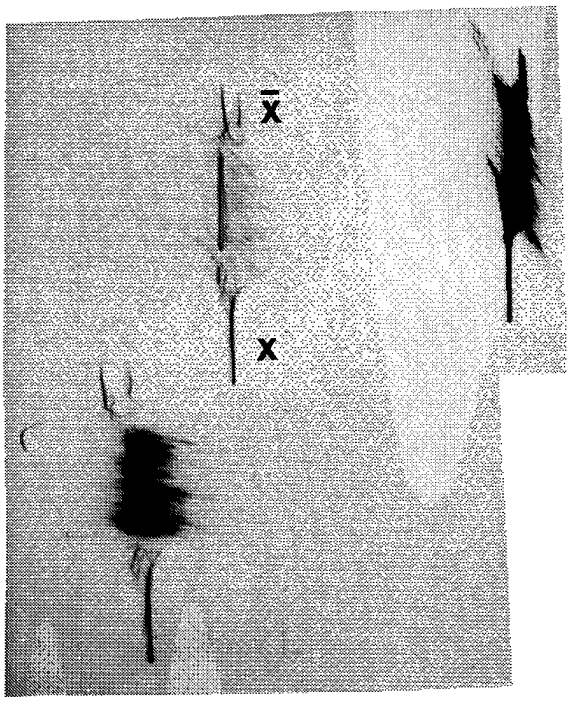


Figure 4-These three sections topographs of another crystal have been recorded simultaneously with different diffraction vectors in the same Laue pattern. The $\bar{X}$ zone is double, the two layers of the gallium phosphate are separated and give two crystals which grow parallel.

are very perturbed. The two other zones below and above the previous one correspond respectively to the X and $\bar{X}$ zones. The X zone is well defined and shows a good crystalline quality of the gallium phosphate. However the interface between this zone and the seed is very disturbed. The $\bar{X}$ zone is very interesting. This zone is double, the two layers of the gallium phosphate which are in position of epitaxy, at each side of the seed, are separated in $\bar{X}$ zone and give two crystals which grow parallel. In the X zone, it may be observed that at the start, the growth had a double direction. Progressively, the two parts joint together and give rise to a single X zone.

This example shows that the growth of the gallium phosphate is obtained from the epitaxial layers on the berlinite seed which develop to give the X and $\bar{X}$ zones instead of a direct growth on the lower and the upper parts of the seed. This result is very interesting and shows how to obtain large good seeds of gallium phosphate.

To have a significant growth rate along Z we continue the growth from the crystal in phosphoric acid and composite gradient method. This method, with nutrient in the colder side of the glass autoclave, allows to obtain larger crystals than the slow heating method. The fluid circulation, from cold to hot zone due to the horizontal temperature gradient, is done at constant temperature or coupled to a slow heating rate of 0,25 to 1°C/day. Then we can cut gallium phosphate X seeds for a growth in sulphuric acid and VTG method or Z seeds for a new growth in phosphoric acid and the SHT method or the composite gradient method.

The two topographs in the figure 5 show two examples of growth from this type of seed. On the top a Z slice of a crystal obtained from a X seed of gallium phosphate and on the bottom a X slice of a crystal obtained from a Z seed. The quality of these crystals is quite good. It is possible to see some individual dislocations.

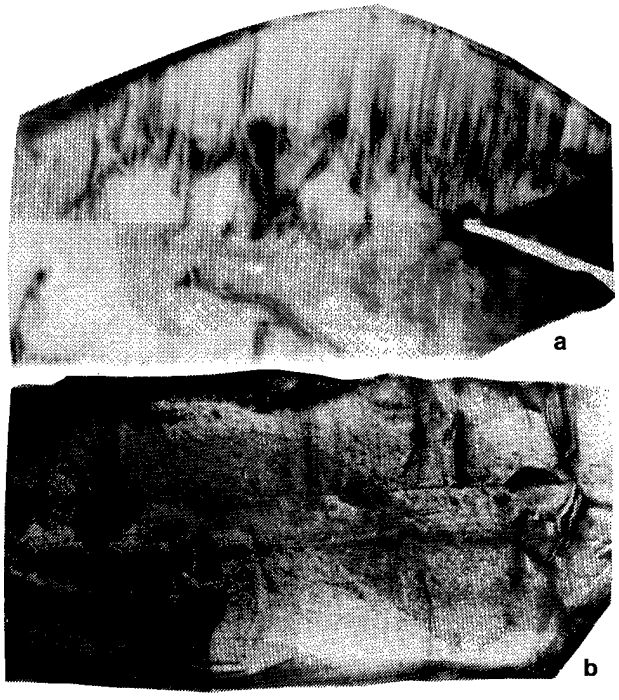


Figure 5-a/ Topograph of a Z slice of a crystal obtained from a X seed of gallium phosphate. **b/** Topograph of a X slice of a crystal obtained from a Z seed of gallium phosphate.

Langasite

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.

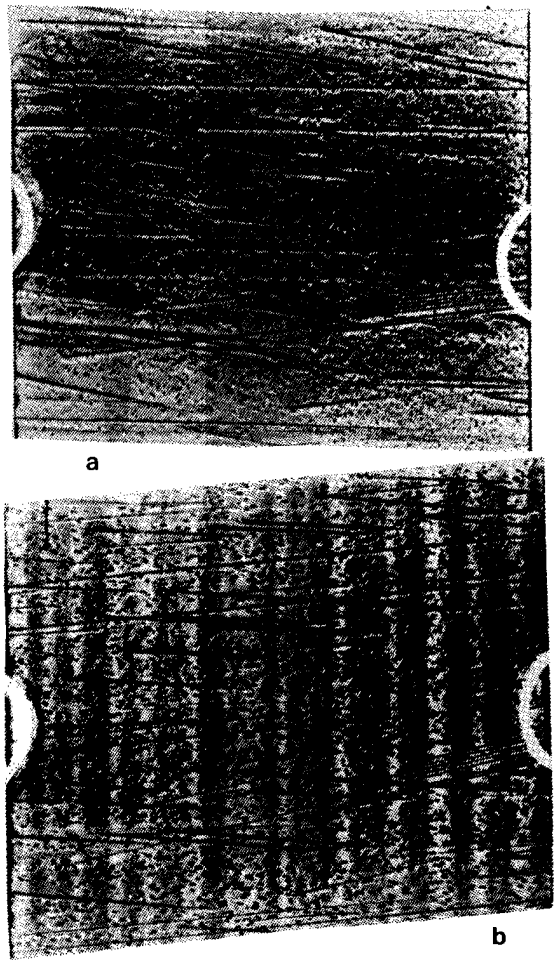


Figure 6-Two topographs of a langasite plane resonator of about 0,5 mm thick. The surface of this Y plate is about 1 cm^2 and the electrode diameter is equal to 6 mm. These two topographs are obtained simultaneously in the same Laue pattern with two perpendicular diffraction vectors.

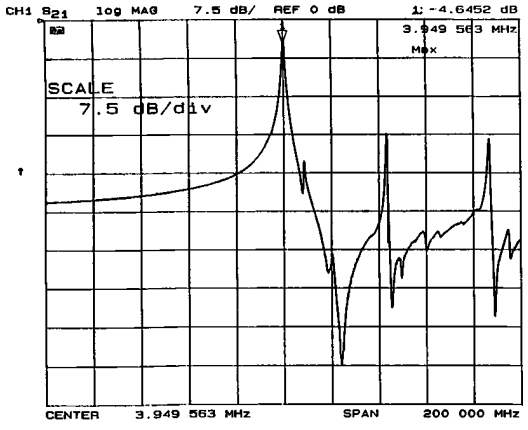


Figure 7-Response curve of the crystal of the previous figure.

The figure 6 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. 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 doping. The large bands alternatively black and white which appear in the topograph in the figure 6a are growth bands. They are perpendicular to the pulling axis and do not appear on the other topograph because the diffraction vector is perpendicular to their displacement vector. Indeed, generally the displacement vector of the growth bands is perpendicular to the bands.

The response curve of this crystal is represented in the figure 7. The frequency of the fundamental mode is equal to 3.950 MHz. The mode which appears in the middle of the right side has been identified like a flexure mode, we shall see it in the figure 8. The two next modes are two antisymmetrical enharmonic modes, the last is a symmetrical enharmonic one. The first antisymmetrical enharmonic mode appears before the antiresonance of the fundamental mode, this fact is a characteristic of materials with a high electromechanical coupling.

The topographs in the figure 8 show the fundamental vibrating mode (Fig.8a and b) and the flexure vibrating mode (Fig.8c and d). The

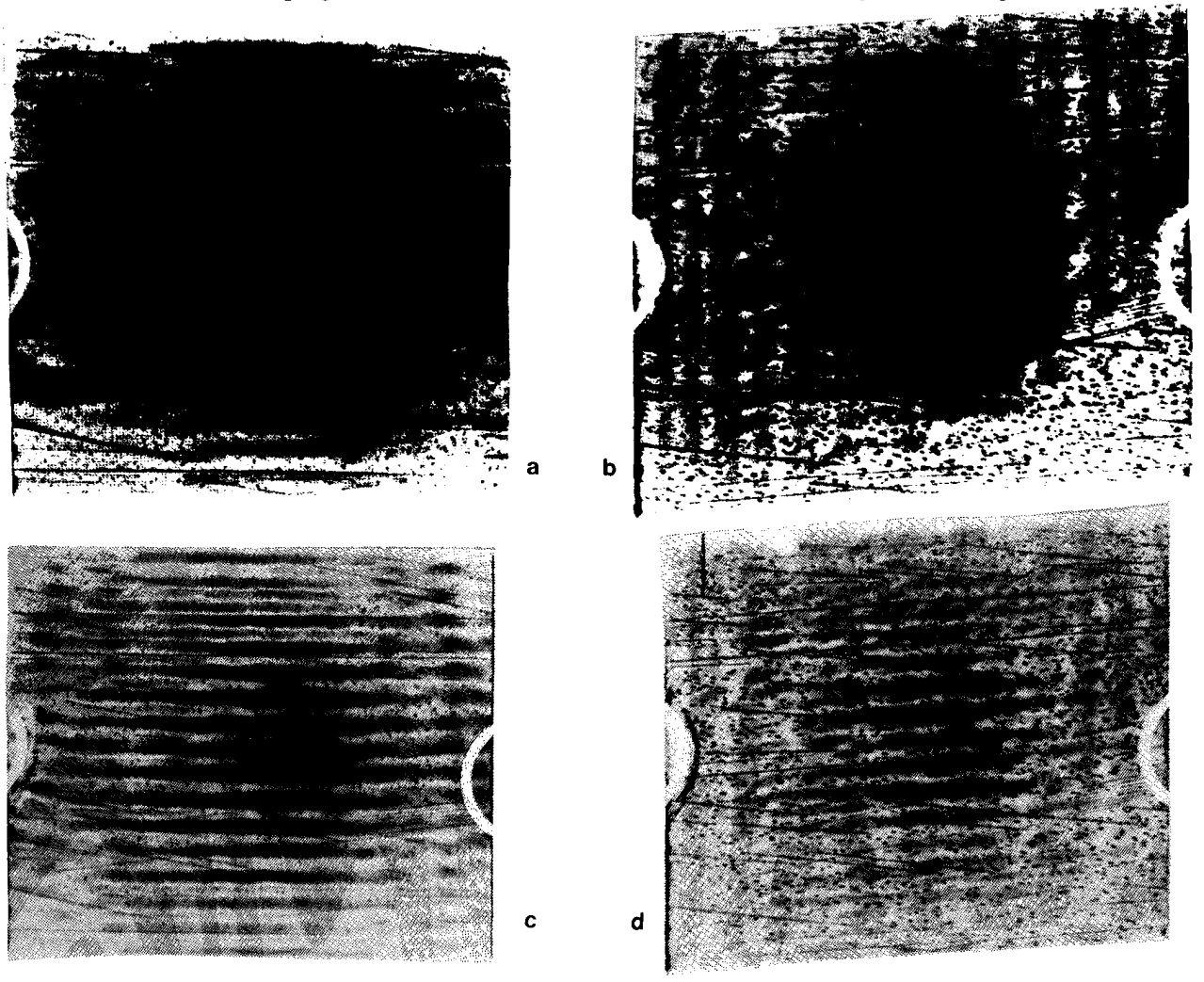


Figure 8-These topographs show the fundamental vibrating mode (u_1 component in **a** and u_3 component in **b**) and the flexure vibrating mode (u_1 component in **c** and u_3 component in **d**).

topographs of the same line are obtained simultaneously with two perpendicular diffraction vectors and reveal the u_1 component (Fig.8a and c) and the quasi pure u_3 component (Fig.8b and d). With the flexure mode appears the fundamental mode by coupling because they are very close together.

In the figure 9, the topographs of the same line are also obtained simultaneously and reveal the u_1 component (Fig.9a and c) and the u_3 component (Fig.9b and d) of the two antisymmetrical enharmonic modes. The mode (Fig.9a) with a nodal line parallel to the x direction is the first antisymmetrical enharmonic one and it appears coupled with the flexure mode.

These two antisymmetrical enharmonic modes are due to a small defect of the parallelism of the plate. We can see that because the two parts of the mode in the figure 9c, in particular, are not equal.

The last topographs, figure 10, of this crystal show the symmetric enharmonic mode, the u_1 component in the figure 10a and the u_3 component in the figure 10b. This mode is coupled with a plate mode which is not the previous flexure mode. The plate mode interacts with the growth bands which divide it into fractions. The unusual shape of the symmetric enharmonic mode is due to the fact that it extends to the edges of the plate.

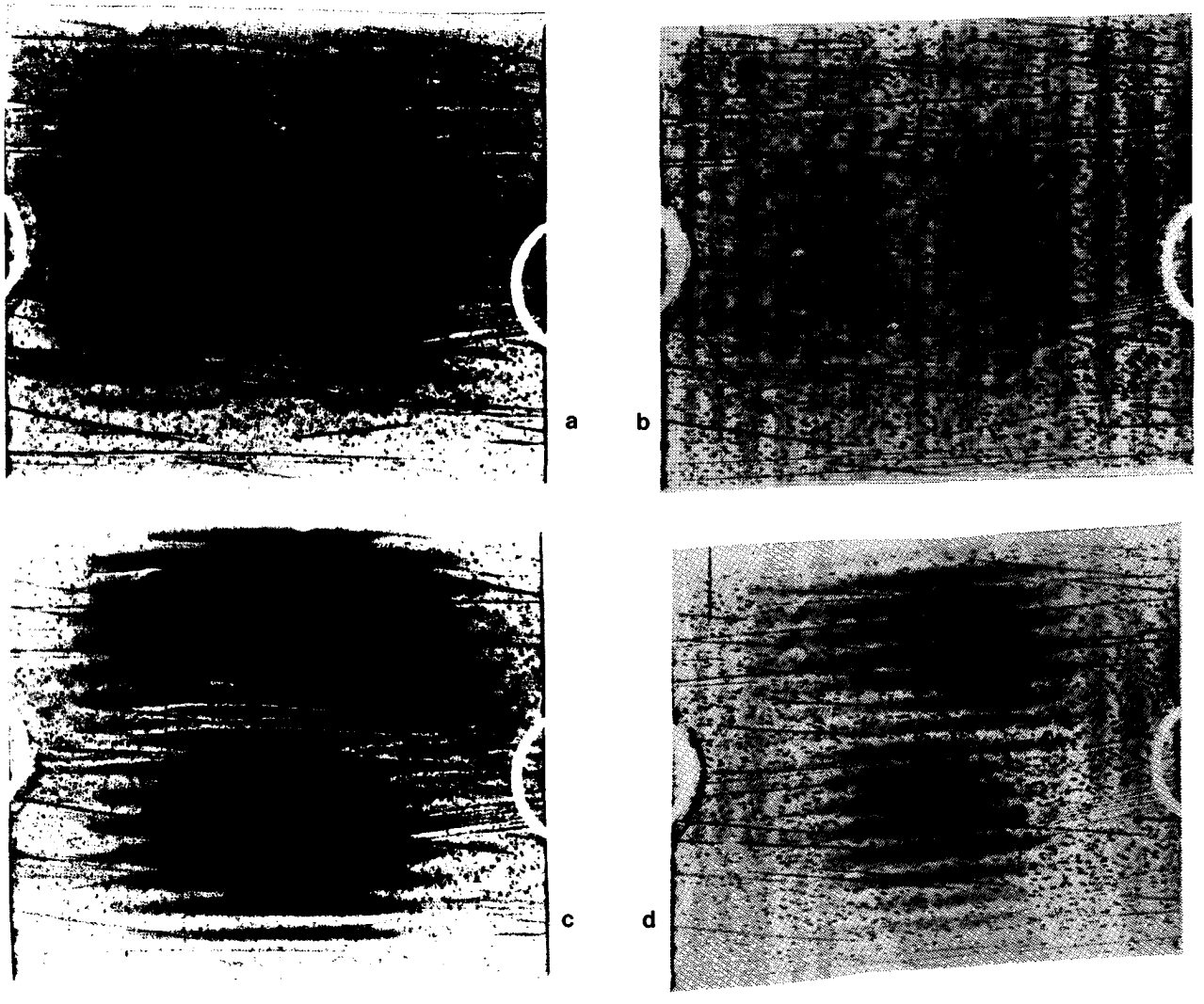


Figure 9-Topographs of the two antisymetrical anharmonic modes. **a** and **c/** the u_1 component. **b** and **d/** the u_3 component.

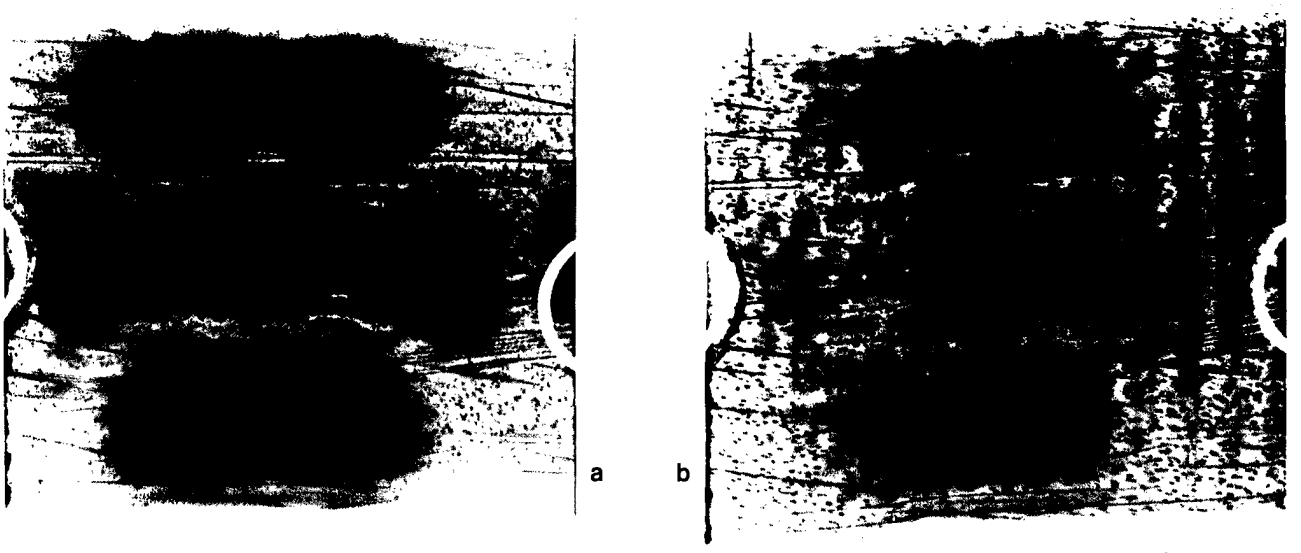


Figure 10-Topographs of the symmetric anharmonic mode, the u_1 component in **a** and the u_3 component in **b**.

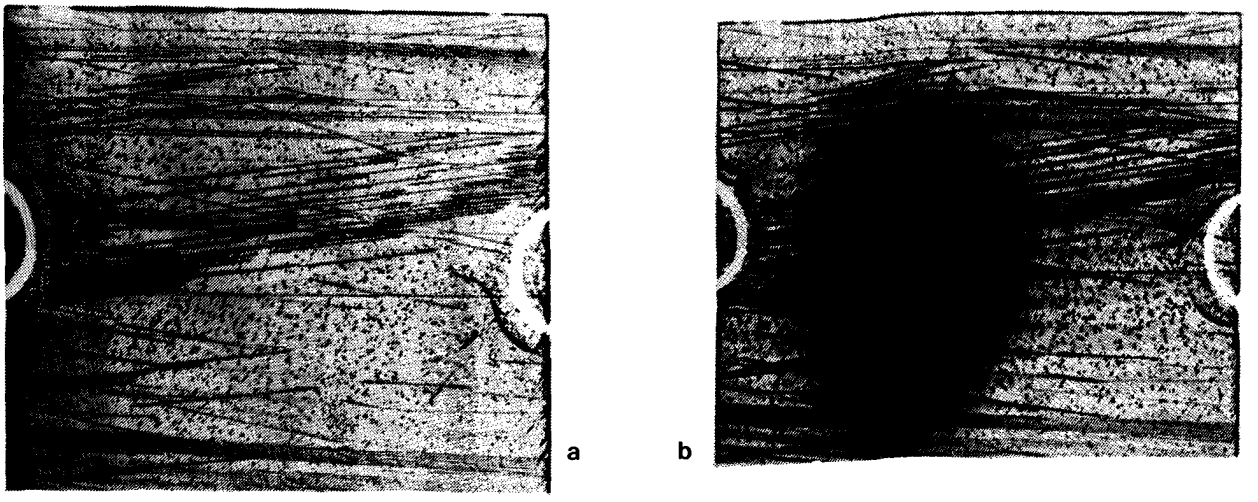


Figure 11-Another example of langasite crystal. **a/** a topograph of the crystal without vibration. **b/** a topograph of the fundamental mode. It appears diffuse because the numerous parallel dislocations guide the energy.

Another example of langasite crystal is presented in the figure 11. In the figure 11a a topograph of the crystal without vibration is obtained with the diffraction vector which does not show the growth bands. We can see the same density of dislocations that in the previous sample but with an inhomogeneous distribution. In the figure 11b the fundamental mode appears diffuse because, in particular, the numerous parallel dislocations guide the energy like in the quartz resonators.

Topographs of two another samples are represented in the figure 12. The first crystal (Fig.12a) presents many dislocations with unusual contrasts. It is necessary to make further investigations to explain these images. The second one (Fig.12b) has also many dislocations but with usual contrasts. The pulling direction is vertical, parallel to the dislocations. We have studied the acoustic modes with this sample which is a Y cut plano convex 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 13 presents the third overtone of this resonator. The excitation level is enough high

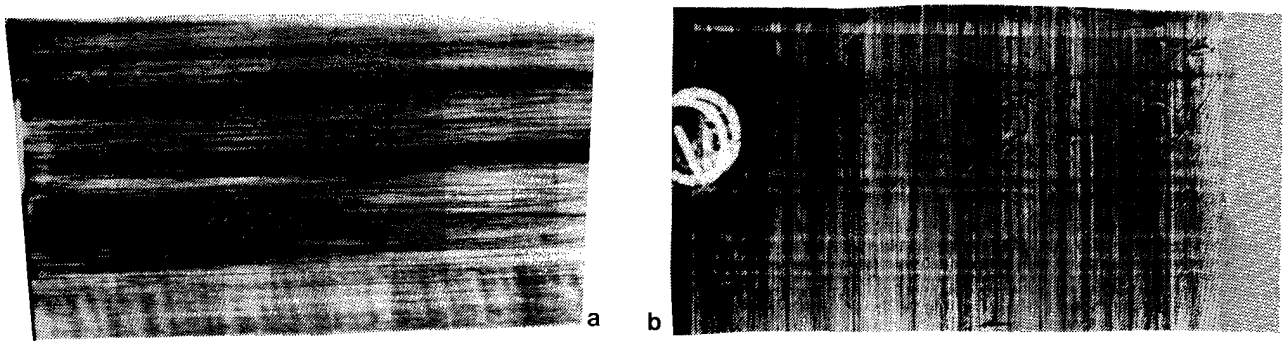


Figure 12-Topographs of two another langasite samples.

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.

The fifth overtone (Fig.14) 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.

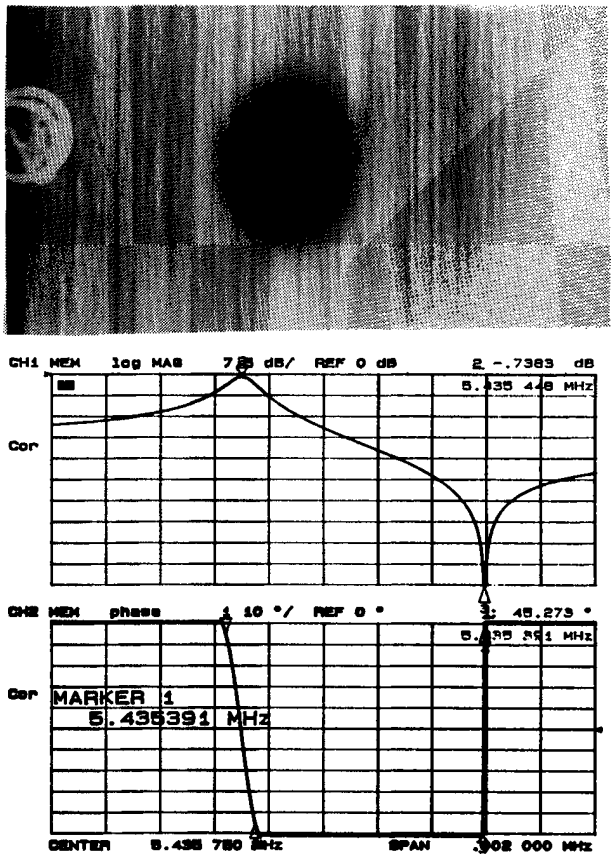


Figure 13-Topograph of the third overtone of the resonator represented in the figure 12b. The response curve shows a low transmission losses. We can remark that the phase slope is steep at the resonance but still more at the antiresonance.

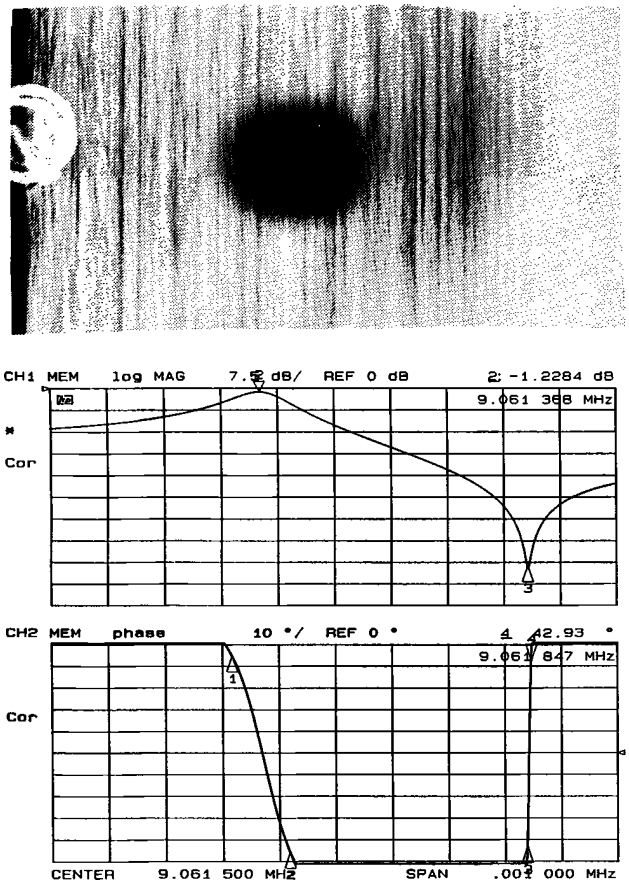


Figure 14-Topograph of the fifth overtone.The response curve shows a Q factor equal to that of the third overtone but at a frequency higher.

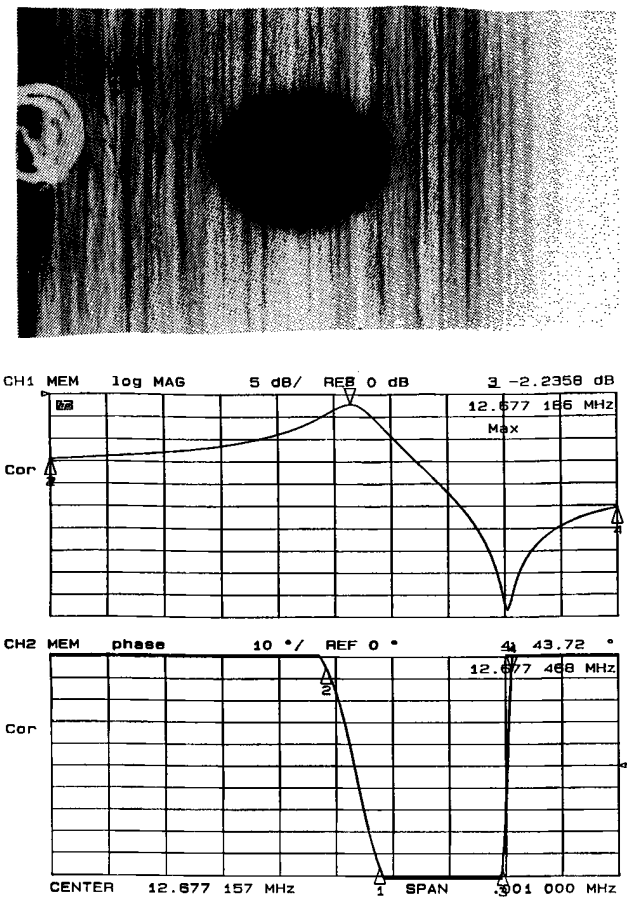


Figure 15-This topograph, of the seventh overtone, is a stroboscopic topograph. The vibration is synchronous with the pulsed synchrotron radiation, but there are no progressive components.

For the seventh overtone shown in the figure 15, 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 synchronous with the pulsed synchrotron radiation, but there are no progressive components.

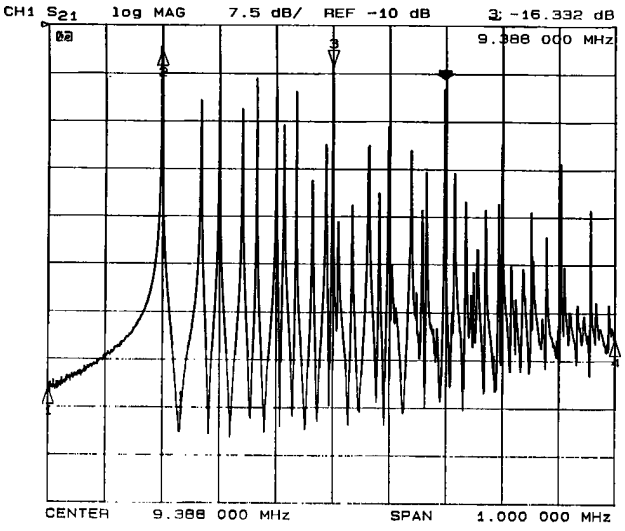
On the table 2 there are the principal parameters of the resonator showed in the figure 12b and some values for the overtones 3, 5 and 7. Electrically these three modes are good with a Q factor, calculated for 5 MHz with the usual hypothesis of 1/f dependence, respectively equal to 0.6, 1.2 and 1.5 millions. These values are very good especially for our first langasite resonator with parameters which were not totally optimised.

Thickness excitation
Y cut 2h=0.756mm Rc=200mm
R=1.04% 2Re=8mm Co=16.41pF

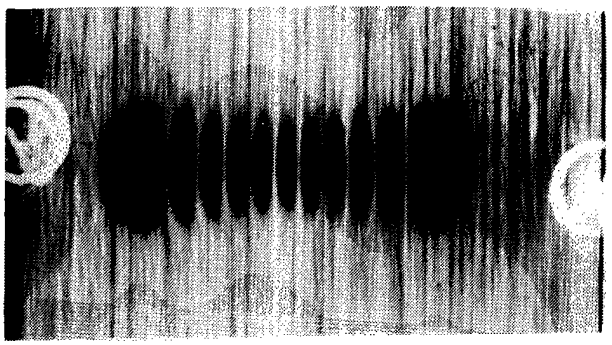
Ov.	$Fr_{(MHz)}$	$Q_{(Fr)}$	$Q_{(5MHZ)}$	$L_{(H)}$	$R_{(Ohm)}$
3	5.435	0.607 *1.248	0.660 *1.357	0.157	8.87
5	9.061	0.681 *1.084	1.234 *1.965	0.181	15.2
7	12.677	0.609 *0.750	1.545 *1.903	0.224	29.3

(*) Q factors of antiresonance (all Q in 10**6)

Table 2-Principal parameters of the resonator showed in the figure 12b and some values for the overtones 3, 5 and 7.



a



b

Figure 16-Response curve of the crystal represented in the figure 12b. It shows the existence of a very great number of anharmonics with an electric response no attenuated, for instance, the topograph, in **b**, of a mode with many nodal lines.

The response curve (Fig.16) shows the existence of a very great number of anharmonics with an electric response no attenuated, for instance, the topograph, in the figure 16b, of a mode with many nodal lines. The great number of non 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, miniaturised resonators.

CONCLUSION

In conclusion this investigation has demonstrated the feasibility of gallium phosphate epitaxy on berlinite seeds. Therefore the topography permits to explain this kind of growth mechanism.

We have seen, with the topography technique, different behaviours of the langasite non foreseeable. These observations allow to further optimise the parameters of the resonators in particular for the high overtones. Indeed the calculation does not give for instance the direction of the anisotropy.

References :

[1] P.Krempl, J.Stadler, W.Wallnöfer, W.Ellmeyer and R.Sclic; Proc. 5th European Frequency and Time Forum, Besançon (1991) p.143.

[2] J.Detaint, J.Schwartzel, C.Joly, B.Capelle, A.Zarka, Y.Zheng, Y.Toudic, E.Philippot; Proc. 6th European Frequency and Time Forum, Noordwijk (1992) p.383-388.

[3] E.Philippot, A.Ibanez, A.Goiffon, B.Capelle, A.Zarka, J.Schwartzel, J.Detaint; Proc. 6th European Frequency and Time Forum, Noordwijk (1992) p.223-227.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Bulk Wave Propagation and Energy Trapping in the New Thermally Compensated Materials with Trigonal Symmetry.

J.Détaint*, J.Schwartzel*, A.Zarka**, B.Capelle**, J.P. Denis***, E.Philippot****.

* FRANCE TELECOM CNET PAB/BAG 196, Avenue H. Ravera, B.P. 107; 92225 Bagneux France.
** L.M.C.P. Universités Paris VI et Paris VII, CNRS. 4, Place Jussieu 75252 Paris Cedex 05 France.
***CRISMATEC Zone Industrielle de Mayencin; 2, rue des Essarts 38610 Gières France.
****L.P.M.S. Université du Languedoc, CNRS. Place E.Bataillon 34095 Montpellier Cedex 05 France.

Abstract: In this communication, we compare the properties of the new piezo-electric materials that belong to crystal class 32: the aluminium phosphate, the gallium phosphate and langasite with those of quartz. In a first part, a comparison of the main piezo-electric properties of these crystals is made. Computed and experimental results concerning the Y rotated cuts and particularly the compensated cuts are given. The most important facts observed are the increase of the coupling coefficient from quartz, to berlinite, langasite and gallium phosphate, and the simultaneous shift of the angular position of the most interesting compensated cuts towards the angles of maximum coupling coefficient. New determinations of the angular positions of the corresponding compensated cuts of gallium phosphate and doped langasite are also reported; the former one displays the largest coupling coefficient and an outstanding thermal stability.

In the second part, the energy trapping properties of these four materials are compared. We first consider the fundamental and the 3rd overtone of the piezo-electric (shear) modes. For the latter, the conventional energy trapping behaviour disappears for angular regions whose sizes increase from quartz to gallium phosphate. Computed and experimental results relative to the trapped modes of plane resonators using the different overtones of Y-cut langasite are then reported. A comparison of the mode shapes and of the electrical properties of filter type resonators using the four materials of class 32 is then made. It indicates that filters having more than 4 times the bandwidths obtained with quartz can be made using the new materials. Extremely high Q factors and the ability to withstand large excitation levels were observed in langasite resonators making use of energy trapping by the geometry (plano-convexe). They appears as very promising for low phase noise applications.

We conclude to the important interest of these materials which possesses an unique set of complementary properties, a high crystalline perfection, superior thermal behaviours, and different coupling coefficients adapted to an extremely broad range of filtering and frequency generation applications.

1. INTRODUCTION:

The evolution of electronic systems towards higher frequencies and baud rates has led to the interest to find new thermally compensated piezo-electric materials permitting to obtain filters with larger bandwidths and oscillators with larger shifts or larger frequency stability (mostly short term) than quartz. This evolution conducts to favour the materials that allow the obtainment of very high frequency devices by some intrinsic property or some particular ability to withstand technological operations leading to higher operating frequencies. Another evolution of the electronics equipments lead to research an important reduction of the dimensions of the devices [1][2].

In the recent years, it was successively observed that berlinite ($AlPO_4$), lithium tetraborate ($Li_2B_4O_7$), gallium phosphate ($GaPO_4$), and langasite ($La_3Ga_5SiO_{14}$), can fulfil these conditions while equalling or surpassing quartz for other very important properties such as the crystalline perfection, the acoustic losses. Most of these new materials belong to the crystal class of quartz. Two of them are close structural analogues of quartz.

Berlinite was the first analogue of quartz to be synthesized [3][4][5][6]. Important progresses in the growth methods and in the knowledge of this material were obtained in the past ten years [7][8][9][10][11]. Since few years, crystals with a good crystalline perfection and a sufficient size are available to permit precise evaluation of devices [10][12] and industrial prototypes. It was recently establish that bulk wave devices using this material have outstanding thermal stabilities and shifts or bandwidths twice those of quartz.

The second quartz analogue, for which crystals having significative dimensions and perfection were obtained was gallium phosphate. The growth of such crystals, the measurements of several sets of constants [14][16], and the first evaluations of devices [16][17][19][20] were reported in the recent years. Two major advances in the development of high quality crystals were reported more recently: the growth by lateral epitaxy on berlinite seeds [18][19][21] and the growth using the direct solubility at high temperature [22].

A very impressive amount of results concerning a lanthanum silico-gallate ($La_4Si_3GaO_{14}$ Langasite or LGS), and several of its modifications (substitution of one or several of the elements of the formula by other elements) or structural analogs has been obtained in Russia since the beginning of the eighties [23][24]. Very recently several important advances were reported: the obtainment of crystals having large dimensions [25][26][27][28][30], the measurement of several sets of constants and of physical properties [25][29][31] and the studies of devices [25][29][30][31]. Langasite and its analogs belong also to the crystal class of quartz.

Many domains of applications of piezo-electric devices are now requiring new characteristics for these devices either to enhance the achievable system performances or simply to allows the implementation of new systems. The most important among them are most probably, the recent and the future land mobile and space digital radiocommunication systems. They are now requiring filtering devices with much larger bandwidth at increased centre frequencies so that new resonators having larger coupling coefficient together with better thermal stabilities are required. For these applications, a drastic reduction of the volume of the devices is also wanted to allow to decrease the size of the equipments (hand-held terminal). Devices using the new materials considered above, are among the best candidates to fulfil such requirements.

Equally, for some of these radio-communication systems, for the high baud rates optical fibre transmission systems and for radars an important advance of the frequency stabilities of the oscillators (often short term, or of their phase noise) may be required. The obtainment of larger values of the shifts for the VCXO together with enhanced spectral characteristics will also be necessary in many cases due to the larger frequency bands allocated to the new systems and to the larger bandwidth of their signals.

0-7803-1945-1/94 $3.00 © 1994 IEEE

Table 1: Properties of SAW device using the new materials.

MATERIALS FOR I.F. FILTERS (MOBILE GSM, DCS1800, UMTS). SAW FILTERS B.W.# 270 & 1080kHz				
Material	k^2 (%)	Relative shift for 100°C	Shift (kHz) at 225MHz	Max.Band W. At 225MHz
Quartz ** (ST cut)	0.14 dv/v=.058*	90 10^{-6}	20	***
$AlPO_4$ (ST cut)	dv/v=.245*	lower than for quartz ?	?	***
$GaPO_4$ (ST cut)	dv/v=.146*	105 10^{-6} or smaller?	28	***
L.G.S. & analogs	0.45 **	quite small?	?	***
$Li_2B_4O_7$	>0.8 ? **	.50-1.0 10^{-3}	112-225.	***
$TaliO_3$	1.3 -1.5 **	2.400 10^{-3}	540.	*** >15MHz
$LiNbO_3$ (coupe Y)	4,8 -5.57 **	7.2-9.4 10^{-3}	1900.	*** >25MHz

*** Wallnöfer et al. [16], **Large wafers available**
*****max. bandwidth and insertion losses are not independent they are very dependant on the kind of design used.**

Table 2: Properties of BAW devices using the new materials.

MATERIALS FOR I.F. FILTERS (MOBILE GSM, DCS1800, UMTS). B.A.W. FILTERS, B.W.=270 & 1O80 kHz)				
Material	k (%)	$\Delta f/f$ -25° to +75°	Shift (kHz) at 225MHz	Max.B.W. at 225MHz
Quartz (AT cut)	8.	12.10^{-6}.	2,7 (OK)	500.kHz
$AlPO_4$ (AT cut)	>11	<20.10^{-6}	4.5 (OK)	1100.kHz
$GaPO_4$ (AT cut)	>16 (>18?)	<20.10^{-6}	4.5 (OK)	2300.kHz ov.3 OK for 270kHz.
L.G.S. & analogs **	15 to 25 ?	100 ? to 150.10^{-6}	22.5?-33.7 (OK ?)	1900.kHz ov.3 OK for 270kHz ?
$Li_2B_4O_7$ **	23	300 ? to 600.10^{-6}	too much	4800.kHz ov.3 OK for 270kHz.
TaliO3 **(Xcut)	45	400.10^{-6}	too much	>16. MHz

*** Max. bandwidth=$4k^2/n^2pi^2$ (somehow conservative).**
****Large wafers available.**

An example of the new requirements that can be fulfilled by devices using the new materials, can be found considering the intermediate frequency filters for the present and future European radio-communication systems. Schematically, most of these systems (GSM, DCS1800, DECT, UMTS....) make use or will probably use two kinds of channel width (one of the order of 200kHz, the other of about 4 times this value) [1]. The centre frequencies of these filters are being increased from typically 70-150 Mhz for the GSM to probably more than 300MHz for the 1.8 and 2GHz systems. For the lowest value of the channel width, at the very high centre frequencies, it may be very useful to find devices having a better thermal stability than that presently obtained, so that new devices using more thermally stable waves or materials, or both, can be the possible solutions. For the largest value of the channel width, the use of a new material can be very helpful to obtain larger coupling coefficients, a lower impedance level and reduced dimensions. The new possibilities given by the use of new material in SAW and BAW filters are summarized in **tables 1 and 2.** A more detailed discussion of the interest of using new materials in the frequency generation and filtering devices for the future radiocommunication systems, can be found in reference [1].

In this paper we report investigations made to further precise and compare the piezo-electric properties of the new materials of class 32 and the most important characteristics of devices that can be obtained with them for applications to filtering and frequency generation in the telecommunication equipments.

2. BULK WAVE PROPAGATION AND EXPERIMENTAL RESONATORS PROPERTIES.

2.1 Computed results:

A comparison of the piezo-electric properties of the bulk waves propagating in these four materials was made using computed results concerning the plane piezo-electric (elastic) waves propagating in the Y rotated cuts. These computations

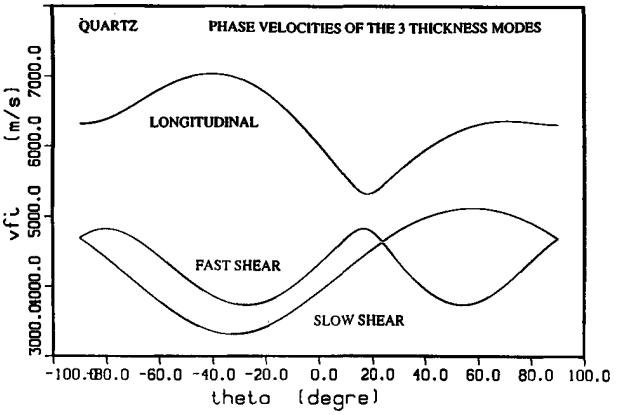


Figure 1a: Velocities of the thickness modes in the Y rotated cuts of quartz.

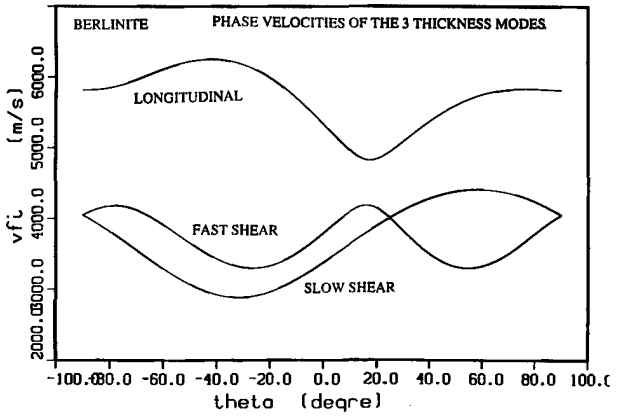


Figure 1b: Velocities of the thickness modes in the Y rotated cuts of berlinite.

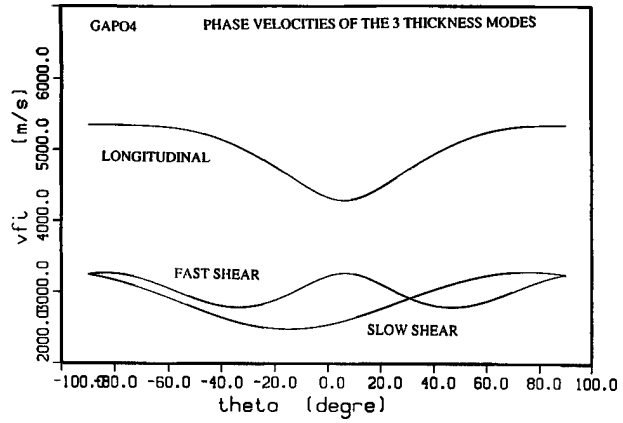


Figure 1c: Velocities of the thickness modes in the Y rotated cuts of gallium phosphate.

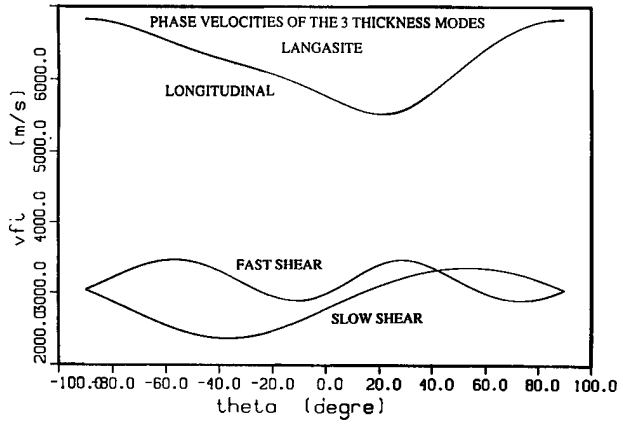


Figure 1d: Velocities of the thickness modes in the Y rotated cuts of LGS.

were made using the most recent constants published for each material [34][35][16][29].In the Y-rotated plates, one two fold axis of class 32, is always contained in the plane of the plate which keeps a monoclinic symmetry. For plates of such symmetry, only one, among the 3 one dimensional modes is piezo-electrically excited by an electric field normal to the surface. Depending on the rotation angle, the piezo-electrically active mode is either the slow or the fast shear. The longitudinal mode is never excited with this field configuration but it can be of interest in U.H.F. composite resonators or in resonators with lateral field excitation (see below §3).

Three quantities which determinate the most important properties of the devices will be considered hereafter: the phase velocities (or the related frequency constants NFa=V/2), the coupling coefficients and the temperature coefficients of the resonance frequencies.

In figures 1 a,b,c,d, the variations of the velocities of the plane waves existing in the Y-rotated cuts of these 4 materials are represented as a function of the rotation angle (theta). We can observe a similar angular behaviour for quartz and berlinite but with some how reduced velocities for the latter material. For $GaPO_4$, the extreme values of the velocities for each mode, appear at quite different angle than for quartz and berlinite. For this material, the velocities are now quite smaller either for the longitudinal mode or for the shear ones. The shear velocities of LGS are of the same order of magnitude as for gallium phosphate but the longitudinal mode is much faster. The angular variations observed for LGS present some similarities with those computed for quartz and berlinite.

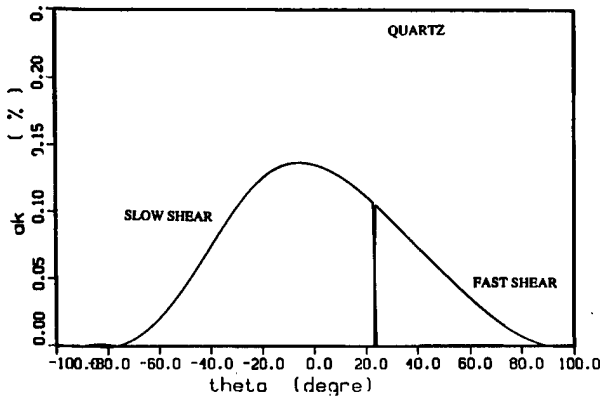


Figure 2a: Coupling coefficient of the piezo-electric shear mode of the Y rotated cuts of quartz.

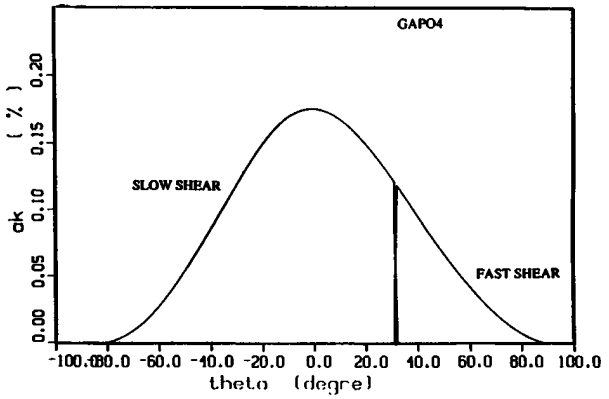


Figure 2c: Coupling coefficient of the piezo-electric shear mode of the Y rotated cuts of gallium phosphate.

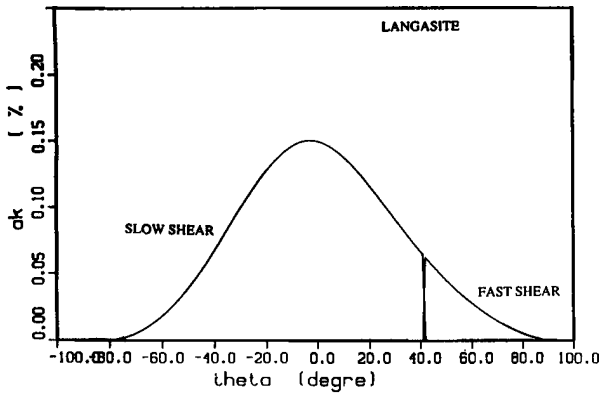


Figure 2d: Coupling coefficient of the piezo-electric shear mode of the Y rotated cuts of LGS.

In figure 2a,c,d we compare the coupling coefficients of quartz, gallium phosphate and langasite. For berlinite, no curve is given since the published sets of constants leads either to much too optimistic or to rather pessimistic results. The exact values for this material would most probably give a curve situated between those of quartz and gallium phosphate. For all the materials considered, the angular variations are very similar. This is mostly the consequence of the fact that, in all cases, one of the two independent piezo-electric

constants is dominant compared to the other and thus that the relevant rotated constants e'_{26} have essentially the same angular variations. We must notice that the coupling coefficient increases in a significant manner from quartz to gallium phosphate. We must also emphasize the fact that for most of the resonators properties, it is the square of the coupling coefficient which must be considered. Then, the differences between the materials become much more important.

The larger coupling coefficient observed for the new materials of class 32 allows to use overtones of higher ranks than with quartz, and thus is a very favourable factor for applications at higher frequencies.

On figures 3a,b,c,d, we compare the variations of the first order temperature coefficient of the resonance frequencies (RFTC) of the three modes. Again, the angular variations of the FTC are very similar for quartz and berlinite. Again slightly reduced values are observed for berlinite. For these materials two Y rotated compensated cuts exist. We have demonstrated that the Y-33° cut of berlinite has a 3rd order behaviour very similar to that observed for the AT cut of quartz. The AT cut of berlinite display a somehow reduced angular sensitivity as compared to the AT cut of quartz. The experiments have also shown that the FTC measured with recent high purity berlinite crystals are, for many orientations, lower than the computed values.

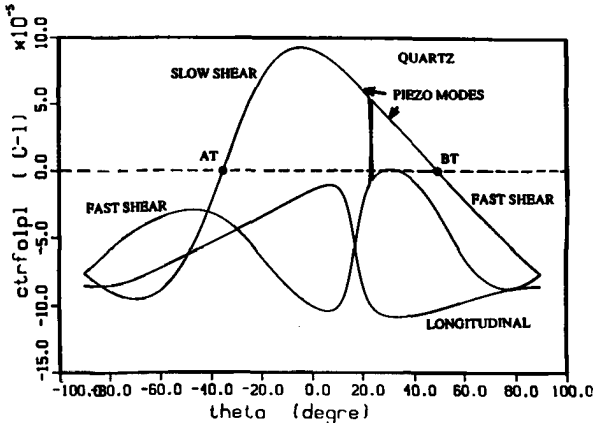


Figure 3a: FTC of the thickness modes in the Y rotated cuts of quartz.

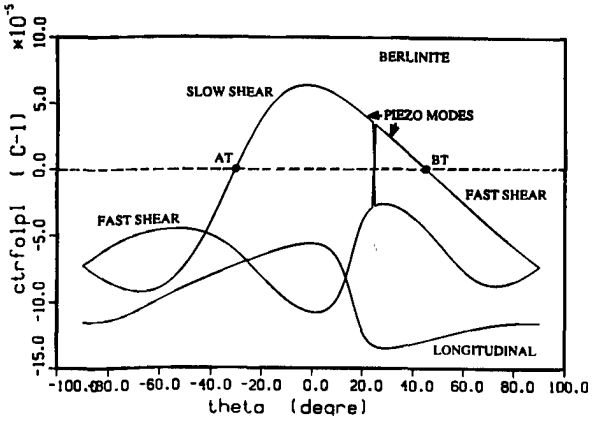


Figure 3b: FTC of the thickness modes in the Y rotated cuts of berlinite.

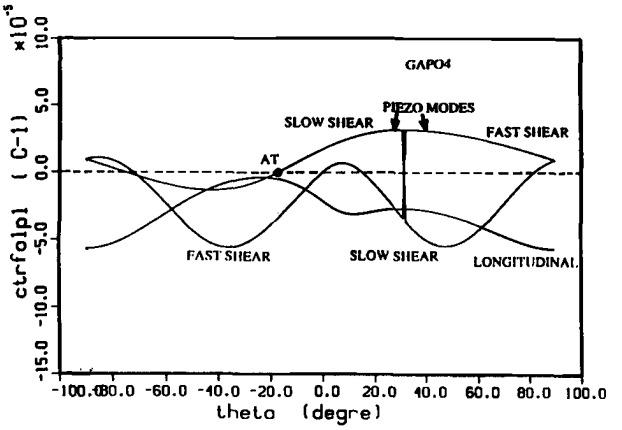


Figure 3c: FTC of the thickness modes in the Y rotated cuts of gallium phosphate.

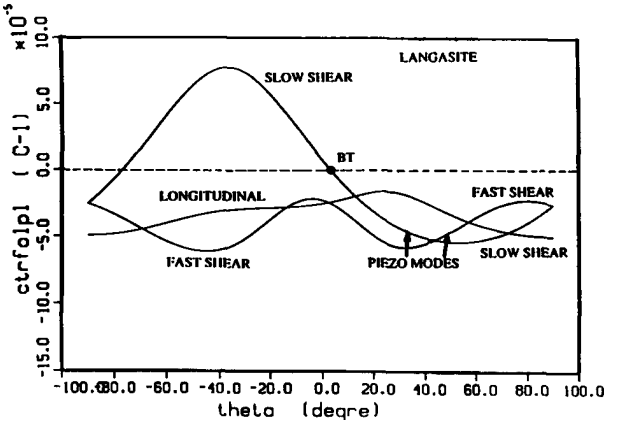


Figure 3d: FTC of the thickness modes in the Y rotated cuts of LGS.

For gallium phosphate, the curves calculated using the last published constants [16] [36] display a very reduced amplitude of variation around zero. This is a first indication of the fact that gallium phosphate has a much greater thermal stability. Two Y rotated compensated cuts appear to exist. Only one of them displays a large coupling coefficient for a thickness excitation. The angular sensitivity of this cut appears to be much lower than for the AT cuts of quartz and berlinite (reduced value of the slope of the curve at FTC=0).

The angular variations of the FTC of the Y rotated cuts of LGS are very different of the preceding ones. Two compensated cuts exist but again, one only corresponds to a large coupling coefficient for thickness excitation. This cut has also a lower angular sensitivity than the AT cut of quartz.

For several of these materials no accurate values of the temperature coefficients of the piezo-electric constants are known. Calculations made supposing that they are of the same order of magnitude as for quartz, have indicated that slightly different results can be found for the temperature coefficients of the resonance frequency, particularly in the angular regions of large coupling coefficients.

As suggested in several recent publications (see for example reference [29] and [30]), the numerous possibilities of substitutions of La, Si and Ga atoms of the LGS formula by other elements can permit to vary greatly the angular position and the properties of the compensated cuts.

An important observation is to be made about the figures 2 and 3. From quartz to berlinite, gallium phosphate and langasite, the angular position of the most interesting compensated cut shifts towards the small values of theta (in absolute values) and hence towards the region of maximal coupling coefficients. This fact is one of the most important reasons of the large interest of the non-quartz materials of class 32.

2.2 Experimental results:

Many experiments were made to find the compensated cut of berlinite [12][14]. Recently, similar experiments concerning gallium phosphate and LGS energy trapping plane resonators were made using crystal grown respectively at the Montpellier University and by CRISMATEC.

In table 3, the experimental results concerning the most interesting Y rotated compensated cuts of berlinite gallium phosphate and langasite are summarized and compared to similar quantities for AT quartz. For all the new materials, the experimental angular positions of the compensated cuts are slightly different (some degrees) of the computed values. For berlinite and gallium phosphate this results of the progress of the quality of the materials which has made that the materials used for the present measurements have a better quality than those used some time ago to measure the material constants used in the calculations. For LGS, this results probably of the use for these measurements of Nd doped crystals. These

Table 3: Comparison of experimental results (main compensated cuts of crystals of class 32).

Material	THETA (degree)	k**2	Nfr (kHz.mm)	Q (at 5MHz)	FTC2 (10-9°C-2)	FTC3 (10-11 °C-3)
Quartz AT	-35.25	.65%	1650	2.7 e6	.0 (25°C)	9-10 (25°C)
AlPO4 AT	-33.02	1.3%	1471	8.5 e5	.0 (62°C)	8-12 (62°C)
GaPO4 AT	-13.1	2.8%	1270	7.5 e4	#.0 (40°C)	4-16 (40°C)
LGSY	#.0	2.2%	1381	8.5 e5	-63. (25°C)	-2.6 (25°C)

doped crystals have shown extremely interesting properties (the cut compensated at room temperature is the Y cut, and they have a favourable energy trapping behaviour (see §3)).

The agreement between calculated and experimental results is quite good for the velocities (or the frequency constants). However, as in previous measurements [7][10], we have observed that the frequency constants and almost the coupling coefficients measured for the best samples of berlinite and of gallium phosphate are higher than the computed values (see also below). The berlinite material is now stabilized at a high level of quality, so that a new determination of the constants and of their temperature coefficients can be very useful.

The experimental thermal behaviour observed near the compensated cuts of berlinite, gallium phosphate and langasite are represented in figures 4a,b,c. For berlinite, the thermal variations of the resonance frequency near the AT cut are very similar to those of quartz but with a larger inflexion temperature. The value of the third order temperature coefficient of the resonance frequency is given in table 3.

The gallium phosphate crystals used for this study display a reduced concentration of OH impurity (of the order of 100-150ppm) and the experience acquired with berlinite leads us to believe that the thermal behaviour observed is already quite close of the intrinsic behaviour of highly pure $GaPO_4$.

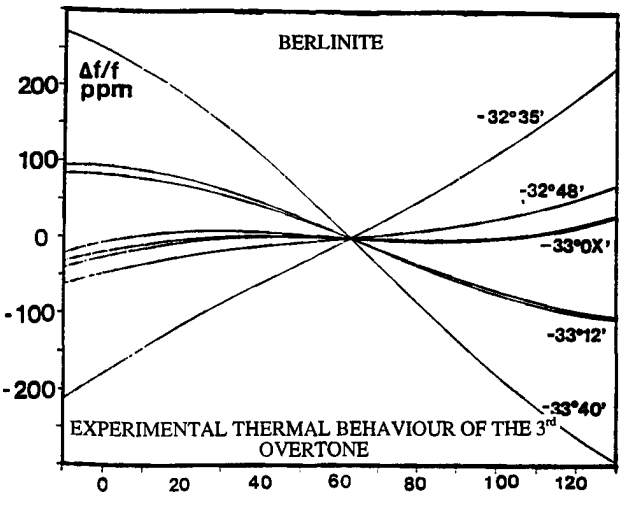


Figure 4a: Thermal behaviour near the AT cut of berlinite.

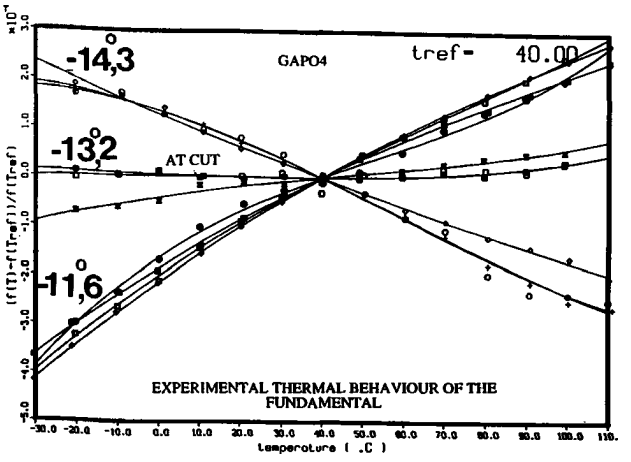


Figure 4b: Thermal behaviour near the AT cut of gallium phosphate.

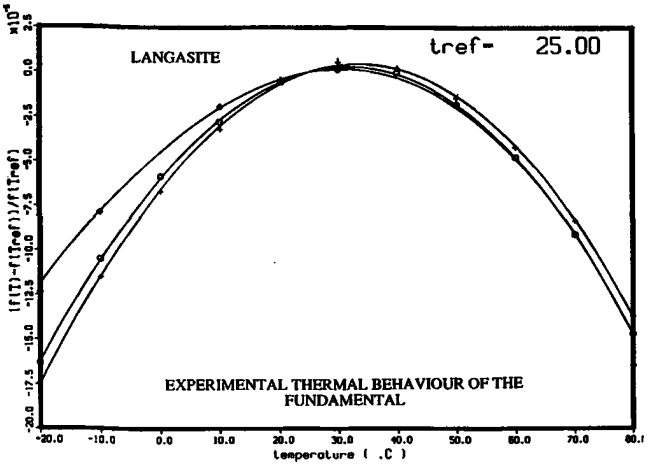


Figure 4c: Thermal behaviour for the Y cut of LGS (Nd doped crystals).

The present results indicate a third order behaviour with a cancellation of the second order temperature coefficient in the vicinity of 40°C. In spite of some dispersions, nearly all the examined samples having an orientation near to theta=-13°15' have displayed an extremely good thermal stability (of the order of 20 ppm in the temperature range [-20,+80°] for the best samples).

For LGS we have observed parabolic thermal variations similar to those previously indicated by Russian authors [25]. The measured value of the second order temperature coefficient is 63. 10^{-9} °C^{-2}.

During the experiments many cases of strong coupling of the fundamental shear mode with plate modes (mostly flexure) were observed by x-ray topography [33][46] for gallium phosphate and LGS resonators. Many of the dispersions observed in the measured temperature coefficients of the resonance frequency may result of such couplings which occur more frequently when rectangular plates are used.

The responses curves of resonators, typical of the observations made with AT berlinite, AT gallium phosphate and Y cut LGS are represented in figures 5a to 5c. The "effective" coupling coefficient displayed on these figures and in table 3 are the values extracted from the resonance-antiresonance relative difference using the formula $\Delta f/f = 4k^2/n\pi^2$ without any correction for the parasitic capacitances of the resonators. For berlinite and for gallium phosphate filter type resonators, the plate to case capacitances have values ranging from one quarter to half the static capacitance, so that they lead to much under-estimated values of the coupling coefficients. For example, for the AT gallium phosphate resonator of figure 5b, which was made using elliptical electrode respecting very nearly the in-plane anisotropy (see §3 below), the values of the parasitic capacitances were found to be C_{p1}#C_{p2}#0.948pF, the "true" static capacitance was: Co#2.295pF while the "effective" static capacitance was $C'_o=Co+C_{p1}.C_{p2}/(C_{p1}+C_{p2})$=2.770pF. The observed antiresonance frequency was 5.569680MHz it corresponds to the capacitance C'_o, the corrected antiresonance frequency (corresponding to C_o) is of about 5.5832MHz, while the resonance frequency is, in first approximation, independent of the stray capacitances (5.505905MHz). The coupling coefficient corresponding to this value, calculated using the preceding formula is k=18.4%. When it is calculated using the exact expression for a one dimensional mode a very similar value is found (k=18.35%). These values are much greater than those predicted by the calculations.

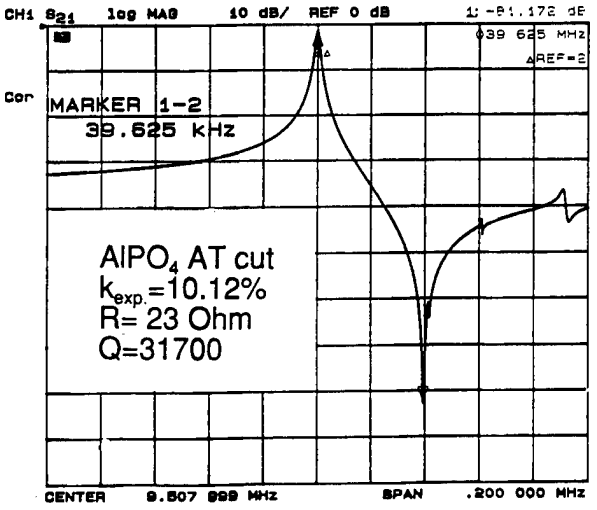


Figure 5a: Response curve for an AT cut berlinite plane resonator.

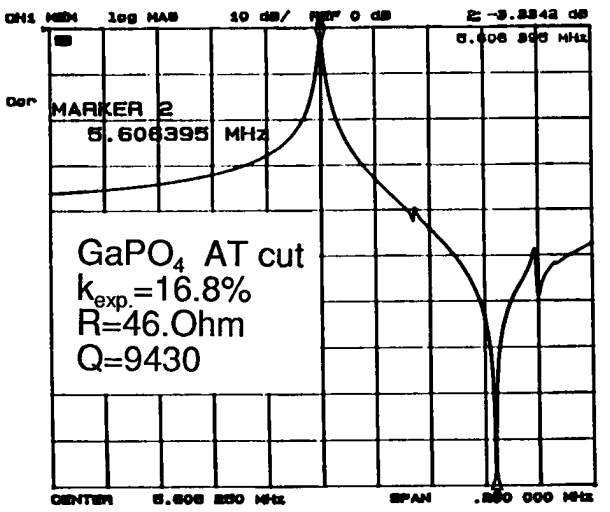


Figure 5b: Response curve for an AT cut Gallium Phosphate plane resonator.

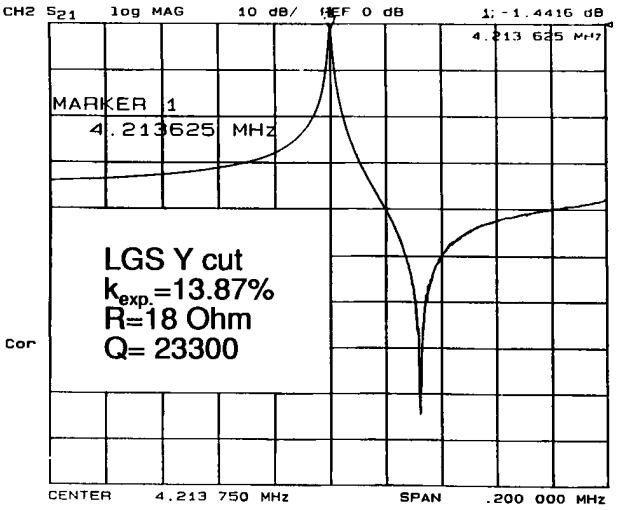


Figure 5c: Response curve for an Y cut LGS plane resonator.

3. ENERGY TRAPPING PROPERTIES.

3.1 Computed results:

3.1.1 γ coefficients for the piezo-electric modes of class 32 crystals: A comparison of the energy trapping properties of the four materials was made using models of the plane resonators [37][38][39] based upon the equations given by H.F. Tiersten and co-authors [40][41][42]. On figure 6 we have represented the angular variations of the square root of the ratio of the two first coefficients of the equation governing the lateral variations of the main displacement (u_1) of the fundamental mode in plane resonators. The solutions of this equation, are very dependant on this ratio which, when it is real, is one of the most important factor of the lateral variations of the mode. When it is imaginary no conventional energy trapping can occur. For the Y rotated cuts of class 32 materials, the M_n coefficients can be real positive or negative numbers or complex numbers. For the plano-convexe resona-

tors, in which the energy trapping results of the contouring, the lateral anisotropy of the mode is governed by $\sqrt{\gamma} = \sqrt{M_n/P_n}$ [40] [42][43][44].

For the fundamental mode, as we can observe in figure 6, γ is always a real number, greater than unity, which means that the vibration modes excited by circular electrodes are elongated in the direction of the two fold axis x. This lateral anisotropy increases from AT quartz to AT berlinite and AT gallium phosphate while keeping quite moderate values. It is very large for the BT cut of these materials and for the compensated cut of langasite. The angular variations of γ are very similar for quartz, berlinite, and, surprisingly, for LGS.

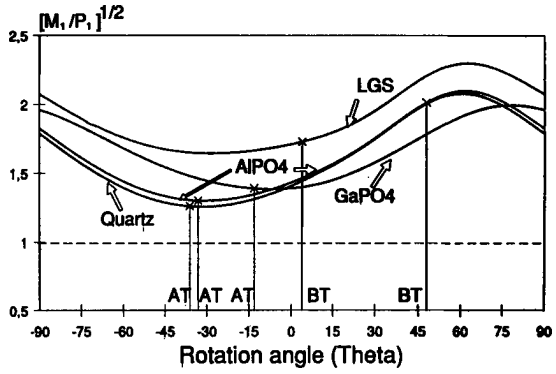


Figure 6: γ coefficients for the fundamental piezo-electric mode of Y rotated cuts of class 32 crystals.

Several characteristics of plane resonators are very dependant on the mode shape and hence on the value of γ. Among them are the values of the motional elements of the equivalent scheme which depend of a quantity that is very approximately the fraction of total acoustic energy confined under the electrodes. For a resonator having round electrodes of a given mass loading, a value of γ very different from the unity leads to a mode shape elongated in one direction and to a value of the inductance which is greater than for $\gamma \neq 1$. We have previously demonstrated that the use of elliptical electrodes having an axis ratio equal to γ minimizes the inductance and optimizes the anharmonic spectra (for given values of the electrode area and mass loading). This technique will have a large interest for several of the new materials. The lateral anisotropy is also an important factor of several non linear properties of the resonators [force-frequency effects, acceleration sensitivity,....).

The values of the γ ratio for the third overtone are represented for each material in figure 7. We can observe that there are angular regions where γ is not real so that no conventional energy trapping can occur. For quartz and berlinite resonators, there are two angular regions where energy trapping exists, they are separated by two small regions. For $GaPO_4$ and LGS energy trapping exists only in one angular region which is very small in the case of $GaPO_4$ but include the AT cut. For the third overtones of the Y rotated cuts of the materials of class 32, very large or very small values of γ and hence very anisotropic modes can be encountered.

In figure 8 the values of the γ coefficients computed for the different overtones of LGS in the vicinity of the compensated cut (situated near the Y one) are represented. Very near the Y cut there are several angular regions where no energy trapping exists for a given overtone. Particularly, energy trapping disappears for the third, the 7th, and the 11th overtone at small positive angles, situated close to the Y cut.

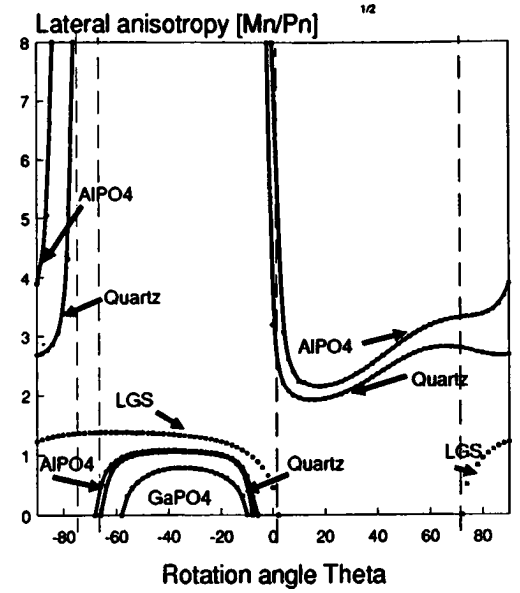


Figure 7: γ coefficients for the 3rd overtone of the piezo-electric mode of Y rotated cuts of class 32 crystals.

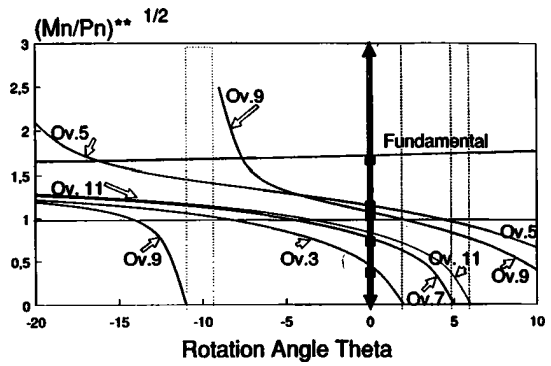


Figure 8: γ coefficients for the different overtones of the piezo-electric mode of Y rotated cuts of langasite.

For these overtones, this implies fast variations and hence noticeable angular or "geometric" sensitivities for the anisotropy of the mode. These, in turn, imply noticeable sensitivities for the properties which are directly related to this quantity such as the values of the elements of the equivalent scheme and the anharmonic spectra. This may lead to the need of using more precise orientations than could be required by the quite low angular sensitivity of the compensated cut.

According to the computed values of γ, the fundamental mode extends in the x direction while the 3rd one is elongated in the z direction. The computations predict a more isotropic behaviour for the 5th and the 7th overtones. On figure 8, we have shown by square dots the experimental observations of the mode shape that were made using X-ray topography (§3.2 and 3.3).

3.1.2 Comparison of filter type resonators: The mode shapes and the properties of filter type plane resonators using the four materials were computed using the same hypothesis: 20MHz fundamental resonance frequency, mass loading =1%, elliptical electrodes respecting in each case the lateral anisotropy (axis ratio= γ) and having the maximal area allowing to have no symmetrical anharmonic. The most important results are given in table 4.

Table 4: Comparison of the properties of filter type plane resonators using the main compensated Y rotated cuts of class 32 materials.

Material	Gamma	2h (um)	2.AX1 (mm)	L (mH)	Co (10-12 F)
Quartz AT	1.261	82.1	3.35	3.542	3.38
AlPO4 AT	1.304	71.2	2.90	3.571	2.97
GaPO4 AT	1.392	61.1	1.65	2.415	1.40
LGS Y	1.710	68.5	1.85	0.884	3.82

In table 4, we can observe that the electrodes dimensions decrease strongly from quartz to gallium phosphate and LGS. The motional inductances and almost the static capacitance decrease some how from quartz to gallium phosphate. This indicate an increased sensitivity to the stay capacitances, which should be carefully minimized or compensated, particularly for VCXO applications (see the end of § 2.2). For LGS, the inductance is much lower and the static capacitance larger, due to the larger value of the dielectric constant (see the expression of the motional inductance in ref. [37]).

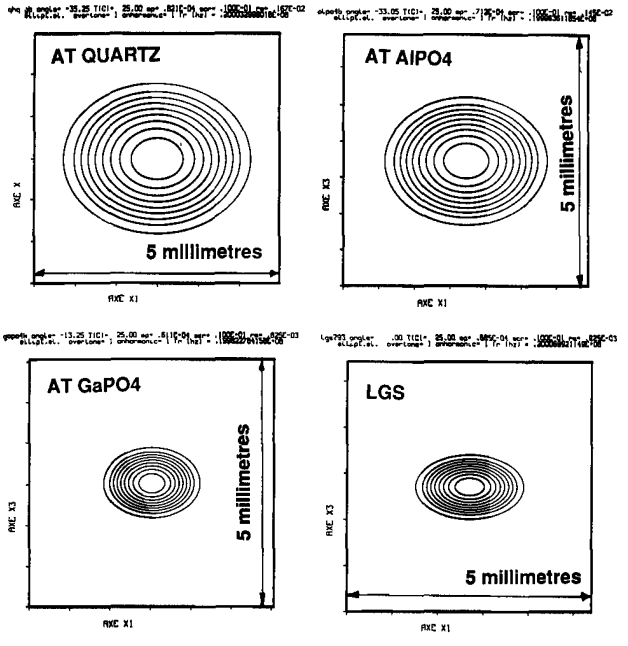


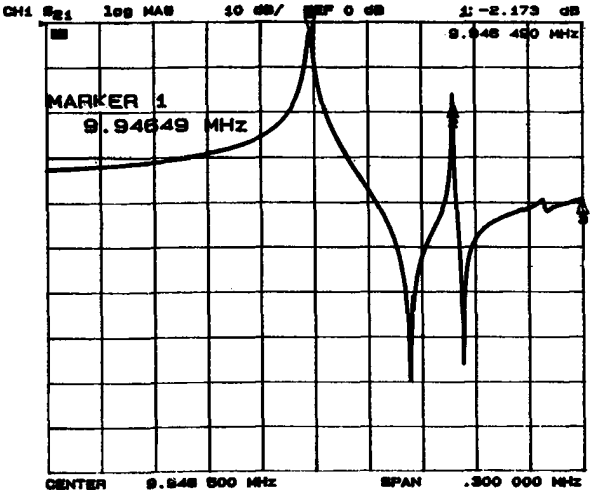
Figure 9: Computed mode shapes for filter type fundamental mode resonators using the main compensated Y rotated cut of class 32 crystals.

The mode shapes are compared in figure 9. They are drawn at the same scale (x1 and x3 coordinates) and the modes are normalized to have the same amplitude at the centre of the resonator $x_1=x_3=0$. The observed diminution of the "size" of the modes when the coupling coefficient increase is very favourable to obtain, at a given frequency, a reduction of the dimensions of the devices.

Another remark must be done about the choice of the hypothesis: due to its symmetry (centro-symmetric) in a transformed coordinate system, the first symmetrical anharmonic is theoretically unexcited with such an electrode geometry. This may allow to use larger electrodes to suppress only the 2nd anharmonic, and so to have lower inductances; but the exact realisation of the symmetry and geometric conditions required to suppress the 1st anharmonic may require a much greater accuracy in the technology.

3.2 Experimental results for plane resonators:

3.2.1 Vanishing of energy trapping: For the 3rd overtone of Y rotated quartz and berlinite resonator, the calculation indicates that energy trapping disappears very near the Y cut while it exists for the fundamental mode. The response curve and the mode shape of a Y cut quartz resonator operating on the fundamental mode are given in figures 10a and 10b. The response curve, and the mode shape observed for this fundamental mode are very characteristic of energy trapping. The response curve corresponding to the third overtone for the same resonator is given in figure 11a . Near the expected resonance frequency of the 3rd overtone, we can observe a non-usual distribution of modes which present all a large attenuation. X-ray topography observations of the mode shape of all the resonances appearing in figure 11a were made. No mode shape similar to those expected for the anharmonics of the 3rd overtone were observed. The topographs have revealed that the modes have very unusual geometries. Four topographs showing the u_1 and u_3 components of two of the



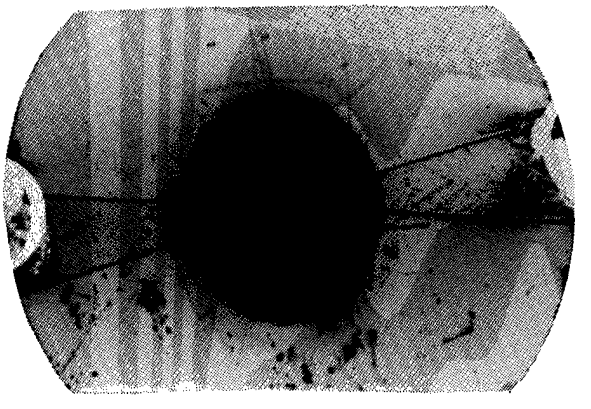


Figure 10: Response curve and mode shape of the fundamental mode of a Y cut quartz resonator.

most intense resonances are displayed in figure 11b,c,d,e. We can observe that the 2 components of these modes have both a very intricate shape and very nearly the same amplitude. We can also see that the components extend up to the edges of the plate particularly in the tab regions. There are also several indications of the existence of some progressive part for these components [32]. A similar electrical response was observed for the 3rd overtone of Y cut berlinite indicating a similar behaviour.

3.2.2 Comparison of computed and experimental *mode shape for quartz, berlinite and gallium phosphate:*

Experimental determinations of the mode shapes of fundamental and overtone resonators using the compensated cuts of these four materials were made using conventional or synchrotron radiation topography [32][46]. These observations were compared to mode shapes computed using a model of the plane resonators. Several of the corresponding results were previously presented in the case of quartz and berlinite. A very good agreement was observed for AT quartz resonators; the mode shapes, the resonance frequencies and the values of the elements of the equivalent scheme are very accurately predicted by the model [37]. For the AT cut of berlinite [17][37][39] the agreement is quite good for all these quantities. A first comparison of experimental and computed values of the elements of the equivalent scheme [17] and the first experimental observations of the mode shape of near AT $GaPO_4$ resonators [33] were recently reported. Further experiments and computations are now being made using the much better crystals and material constants now available.

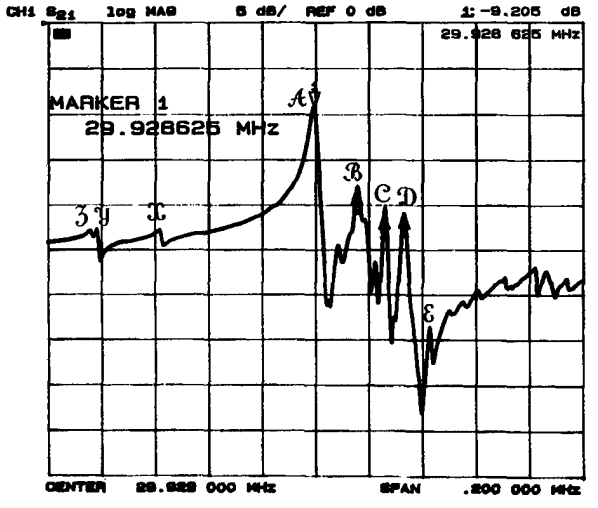
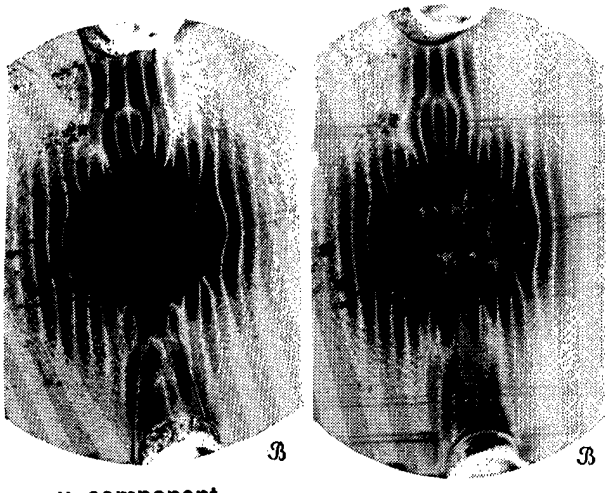
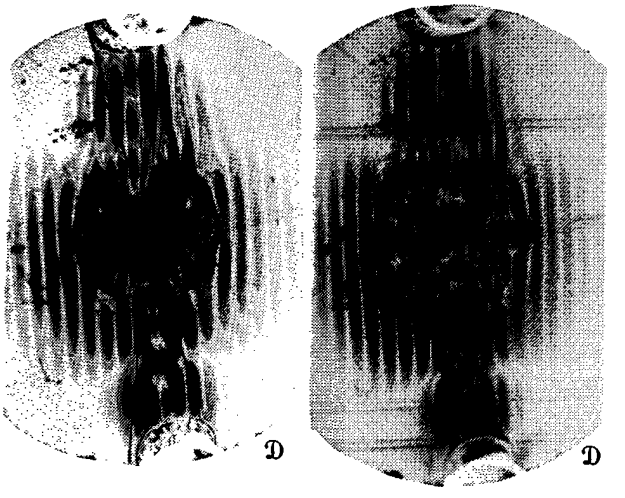


Figure 11a: Response curve near the expected resonance frequency of the 3rd overtone of a Y cut quartz resonator.



u_1 component u_3 component

Figure 11b,c: Mode shape (u_1 and u_3 components) for the attenuated resonance observed near 29.934MHz (mode $\mathfrak{B}$ on figure 11a).

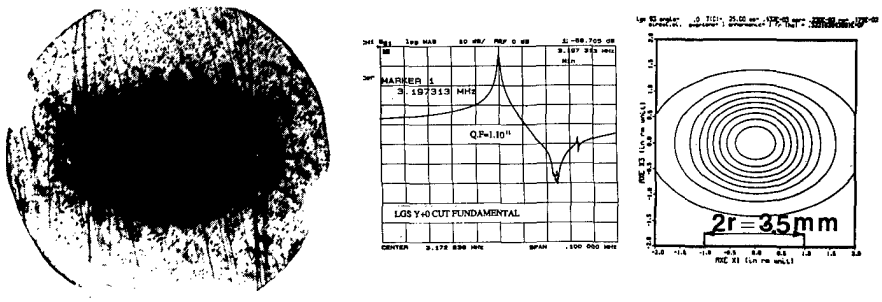


u_1 component u_3 component

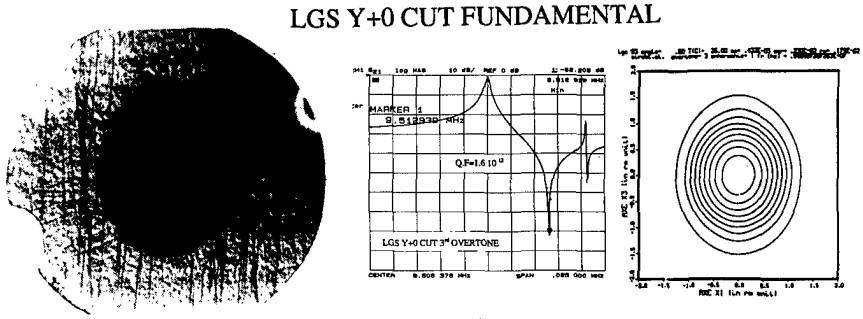
Figure 11c,d: Mode shape (u_1 and u_3 components) of the attenuated resonance observed near 29.947MHz (mode $\mathfrak{D}$ on figure 11a).

3.2.2 Comparison of computed and experimental mode shapes for langasite. On figure 12, the mode shape experimentally observed for the different overtones of a Y cut plane langasite resonator (plate thickness =.432mm, round electrodes 3.5mm diameter, mass loading= .23 %) are compared with the computed ones. The experimental response curves are also displayed for each of these modes. For the fundamental, the third and the fifth overtones modes the agreement between the experimental mode shapes, and the computed ones is good. As predicted by the calculations, the direction of maximum extension of the 3rd overtone (z) is different of those existing for the fundamental and the fifth overtone (x). The anisotropy of the fundamental mode is very important. We can notice, on the response curves of the 3rd and the 5th overtones, that the Q factor is very high (for a plane resonator)

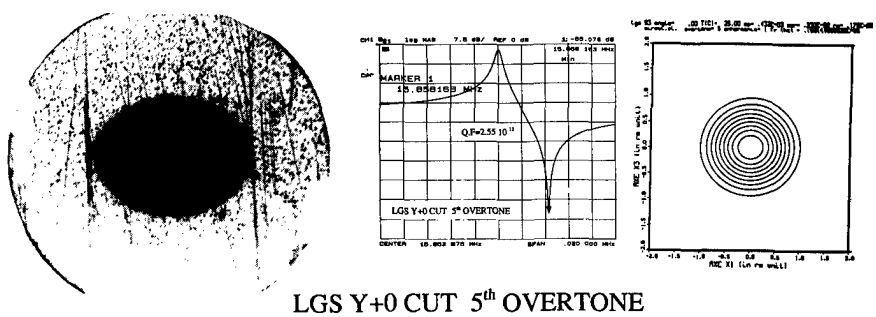
The higher overtones of the same langasite plane resonator (up to the 9th one) display useful responses and present large Q.f products. The agreement between the experimental and the computed mode shapes for the 7th and the 9th overtones is not as good as for the 3rd and 5th overtones. This is probably due to some differences between the material constants determined for pure LGS and those of the doped crystals used here. We must also notice the fact that the calculation of the coefficients of the Tiersten equation involves to evaluate weighted differences of several constants, and hence these coefficients are quite sensitive to small differences in the material constants.



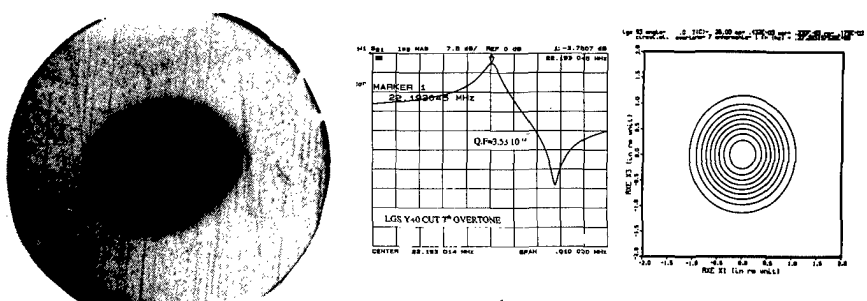
LGS Y+0 CUT FUNDAMENTAL



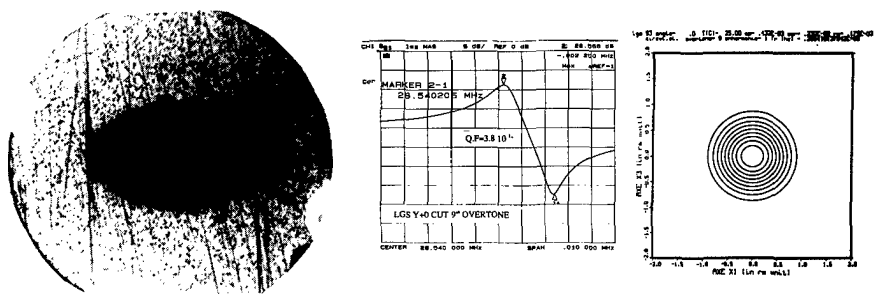
LGS Y+0 CUT 3rd OVERTONE



LGS Y+0 CUT 5th OVERTONE



LGS Y+0 CUT 7th OVERTONE



LGS Y+0 CUT 9th OVERTONE

Figure 12: Comparison of computed and observed vibration modes of a Y cut plane resonator using langasite.

3.3. Experimental results for LGS contoured resonators.

Several plano-convexe Y-cut resonators using 15x15mm square plates cuts in non-doped materials elaborated by Crismatec were made using different designs. The plates were first polished using a cerium oxyde mixt on pitch then using a commercial colloïdal silica dispersion [45] on a clothe. They were etched at 80°C with concentrated phosphoric acid nearly saturated with aluminium phosphate. A second mechano-chemical polishing operation was then made and it was followed by a second etching removing some microns without destroying the polish. The resonators were electroded using Cr/Au metallizations to obtain either a thickness excitation or a lateral field excitation (in the x or in the z directions). In such resonators the energy trapping phenomena results from the thickness variations and the anisotropy is governed by the ratio of Mn to Pn at the power 1/4. So that, modes, less anisotropic, than in plane resonators, are obtained.

3.3.1 Thickness field excitation. On **table 5**, the results of measurements made on a Y cut plano-convexe LGS resonator with thickness excitation are displayed together with the design parameters. High Q factors were measured for the resonance frequencies of several overtones. Still higher Q factors were measured at the antiresonance of several overtones. This is a first indication that much higher values can be obtained at the resonance. It was also observed that electrical impedance level corresponding to these resonances are much lower than what is observed for the corresponding overtone of AT quartz. Further results and the mode shapes observed by X-ray topography for other resonators of a quite similar design can be found in ref. [46]. The mode shapes, are very different of those known for quartz or berlinite plano-convexe resonators (different anisotropy and much smaller spatial extension at equal power). The smaller "dimension" of the modes is predicted by the theories of the plano-convexe resonators [40][42][43][44]. In the theory of H.F.Tiersten [40], it results of the greater values of the coefficients α_n, β_n in the expressions of the Gaussian terms of the Hermitto-Gaussian main displacement (u_1).

Table 5: Properties of the overtones of a Y cut plano-convexe resonator using LGS (thickness excitation).

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

The response curve of the different overtones of the preceding resonators are displayed on figure 13. The measurements of table 5 and the curves were obtained at a source power of 15 dBm. No distortion of the amplitude response was observed even at higher excitation power (up to +20dBm). On the phase response curves, the steep phase slope observed at the anti resonances can be noticed. It much probable that some further optimization of the resonator designs, and of may be of the material, will lead to still higher values of the Q fac-

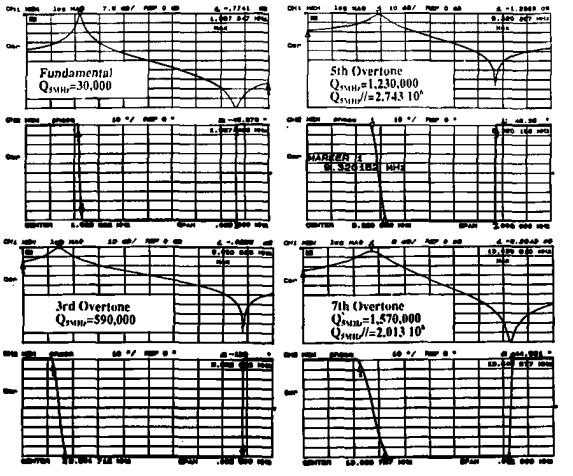


Figure 13: Response curves of the different overtones of the Y-cut LGS plano-convexe resonator of table 5 (thickness excitation).

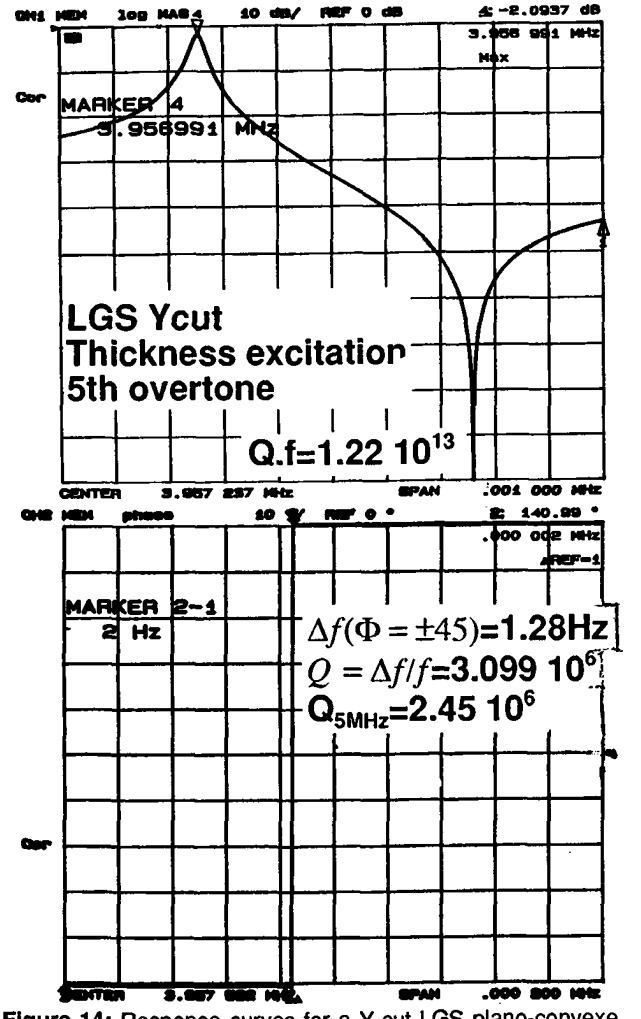


Figure 14: Response curves for a Y-cut LGS plano-convexe resonator displaying a very large Q.f product at the antiresonance of the 5th overtone of the C mode (thickness excitation).

tors for the resonance frequency of the overtone of the slow shear (C) mode. Still higher values of the Q.f product were measured at the antiresonance of other resonators. An example corresponding to observations made for the 5th overtone of a thicker plano-convexe resonator is given in figure 14. The Q.f product measured at the anti-resonance (Q=3.09 10^6 , Q.f=1.5 10^{13} for this resonator) is of the same order of magnitude as the values obtained using the HBAR technique by G.D. Mansfeld [31] whose measurements constitute another very important indication of the possibilities to obtain extremely high Q factors from LGS devices.

3.3.2 Lateral field excitation in the x direction.

The low impedance level predicted and observed for conventional plano-convexe LGS resonators indicate interesting possibilities for lateral field excitation. In table 6, some results obtained for a resonator using a lateral excitation by a field in the x direction are presented. As indicated by several authors and particularly by A.Ballato for quartz and other materials (See the references given in [47]), the x field excite the longitudinal mode and the fast shear mode. We can notice that very high Q factors are obtained for several overtones of the longitudinal (A) mode whereas the overtones of the fast shear (B) seem to present higher acoustic losses with this design. It has been observed that several anharmonics of the even overtones of the C mode that are excited due to their antisymmetry relatively to the collecting electrodes have also interesting properties. The impedance levels obtained with this design are high but they can be some how reduced using the gap parameter. The experiments made have shown that slight variations in the design and in the realization of the resonators can permit to obtain much larger Q factors. This is illustrated in figure 15 by the results obtained with a resonator having a similar design except for a reduced gap (1mm). The Q.f product, then reproducibly measured, is among the highest ones ever measured for a plano-convexe resonator (Q.f =1.65 10^{13} ;Q_{5Mhz}=3.3 10^6). Further experiments with plano-convexe resonators of more varied designs are required to find the maximal achievable values of the Q.f product for the different modes; they migth be surprisingly high.

Table 6: Properties of some modes of a plano-convexe resonator using a lateral field excitation in the x direction.

Y cut 2h=1.746mm Rc=150mm
E field in X direction Gap=1.5mm

Mode	$Fr_{(MHz)}$	$Q_{(Fr)(10^{-6})}$	$Q_{(5MHZ)(10^{-6})}$	$L_{(H)}$	$R_{(Ohm)}$
A5	8.284	0.681	1.296	4.86	371.
A7	11.577	1.154	2.672	6.06	381.
A9	14.869	0.656	1.952	4.69	668.
B5	4.333	0.176	0.153	5.13	791.
B7	6.023	0.256	0.309	8.74	1077.
C4 410	3.192	0.724	0.462	30.95	857.
C6 610	4.769	1.040	0.993	12.95	370

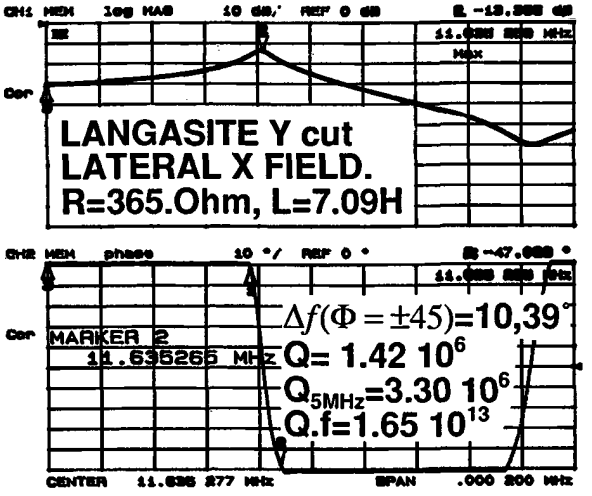


Figure 15: Response curves for a Y-cut LGS plano-convexe resonator showing an extremely large Q.f product at the resonance of the 7th overtone of the A mode (x field excitation).

3.3.3 Lateral field excitation in the z direction.

For LGS, the dielectric constant in the Z direction ε_{33} has a much larger value than those relative to the x direction ε_{11} so that using lateral excitation in the Z direction, leads to reduced impedance level (the inductance is a linear function of $1/\varepsilon$). On figures 16a,b,c, we can observe that then the 4th and the six overtone of the C mode so excited, have impedance level similar to that of 5th overtone quartz using thickness excitation. The Q factor of the 4th overtone is 1 million and those of the 6th overtone is .75 million. The response curve of figure 16c indicates that the distribution of the anharmonics of the 6th overtone is very similar to what is known for the odd overtones excited by a thickness field.

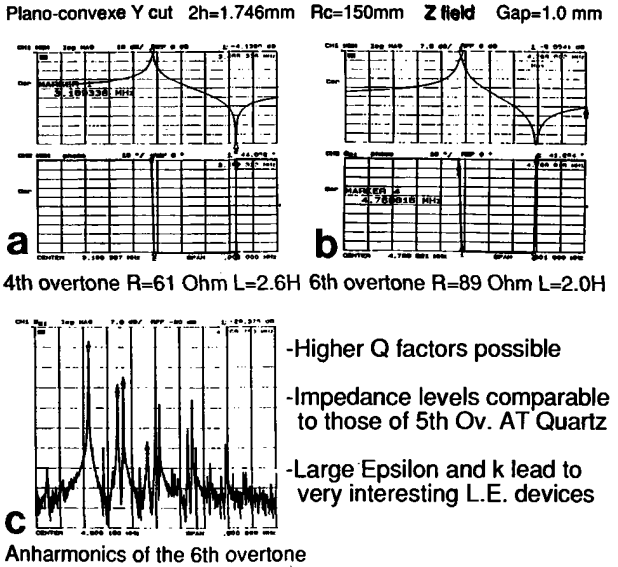


Figure 16: Response curves of the 4th and the 6th overtone of the C mode excited in a Y-cut LGS plano-convexe resonator by a lateral z field.

4 CONCLUSION:

The properties of the most important new piezo-electric materials of class 32 were compared. The calculations and the experiments have shown that quartz and its structural analogs possess an outstanding high thermal stability. They have also shown that the energy trapping properties and hence the design rules of the devices using the new materials can be very different of those known for quartz.

The larger coupling coefficients of $GaPO_4$ and LGS allow to obtain much larger shift for the oscillators and bandwidth for the filters. It permits also to use overtones with much higher ranks than for quartz.

LGS displays extremely reduced acoustic losses and is the prototype of a series of analogs which may also have very interesting properties (including probably for some of them, very large coupling coefficients). LGS is very promising to obtain extremely high Q factor devices for applications requiring very low phase noise.

Gallium phosphate presents an outstanding set of qualities (large coupling coefficients, extremely favourable thermal behaviour, ...). It must be further developed to obtain the large and highly perfect crystals required for the applications to filtering and VCXO's.

Berlinite presents a very interesting compromise of properties, and it is much probably an outstanding surface wave material. Its present state of development is already sufficient to allow in depth investigations of the properties of devices, the obtainment of industrial prototypes and new measurements of the material constants.

On the whole, it appears that the piezo-electric materials of class 32 have very interesting, complementary properties that can cover a very broad range of applications for frequency generation and filtering. The new materials of this crystal class can surpass greatly quartz for all the applications requiring a high coupling coefficient, and also for many other ones demanding a new level of performances on other characteristics (such as: lower losses or reduced phase noise, much reduced dimensions, lower impedance level, high temperature operations, etc...). They are particularly adapted to the new applications to filtering and frequency generation requiring a high level of performances such as in telecommunication equipments.

Acknowledgements: The authors acknowledge the assistance of MM. Caumont, and LeThuault for the elaboration of the studied samples and the helpful contributions of M Joly and Mrs Biblemont.

References:

[1] J.Detaint, J.Schwartzel; Filtering and Frequency Control for the next Generations of Radiocommunication Systems; To appear in Proc. 8th European Frequency and Time Forum Münich March 1994.
[2] A.Ballato, J.G.Gualtieri, Piezoelectric Resonator Materials; Proc.1994 IEEE Frequency Control Symposium (these procedings).
[3] E.D. Kolb, R.A.Laudise; J. Cryst. Growth v.43 p.313 (1978)
[4] H.Poignant, L.LeMareschal, Y.Toudic; Etude de la solubilité du phosphate d'aluminium dans les solutions d'acide phosphorique; Mat.Res.Bull. vol. 14 p.603-612 (1979).
[5] B.H.T. Chai, M.L.Shand, E.Buehler, M.A.Gilleo; Experimental data on piezo-electric properties of berlinite; Proc.1979 IEEE Ultrasonic Symposium p.577 (1979).
[6] J.Detaint, E.Philippot, J.C.Jumas, J.Schwartzel, A.Zarka, B.Capelle, J.C.Doukhan; Crystal growth, physical characterization and BAW devices applications of berlinite; Proc. 39th Annual Frequency Control Symp. p. (1985)
[7] J.Detaint, A.Zarka, B.Capelle, Y.Toudic, J.Schwartzel, E.Philippot, J.C.Jumas, J.C.Doukhan; Berlinite : Characterization of crystals with a low water concentration and design of bulk wave resonators. Proc. 40th Annual Frequency Control Symp. p.101,114 (1986).
[8] B.H.T. Chai , J.P.Hou; Evaluation of berlinite growth at high temperature, Proc. 1st European Frequency and Time Forum p.204 (1987).
[9] E.Philippot, A.Goiffon, J.C. Jumas, C.Avinens, J.Detaint, J.Schwartzel, A.Zarka; New advance in crystal growth of high purity berlinite; Proc. 42nd Annual Frequency Control Symp. p.138-145 (1988).
[10] E.Philippot, A.Goiffon, M.Maurin, J.Detaint, J.Schwartzel, B.Capelle, A.Zarka. Evaluation of high quality berlinite crystal grown in sulphuric acid medium; J.Crystal Growth v.104 p.713-726 (1990).
[11] X.Buisson, A.Zarka, R.Arnaud, B.Capelle, Y.Zheng, J.Schwartzel, J.Detaint, Y.Toudic; Crystal growth of berlinite: methods for growing longer crystals and improving the quality. Proc. 6th European Frequency and Time Forum p.287-291 (1992).
[12] J.Detaint, J.P.Aubry, J.P.Michel, A.Zarka, B.Capelle, R.Lefevre, E.Philippot, J.Schwartzel; Piezoelectric devices using berlinite: Properties and applications. Proc. 1989 URSI symposium on Signals Systems and Electronics p. 433-438 (1989).
[13] G.F.Engel, P.W.Krempl, J.Stadler Proc. Technical aspects of GaPO4; Proc. 3rd European Frequency and Time Forum p.50-56 (1989).
[14] P.Krempl, J.Stadler, W.Wallnöfer, W.Ellmeyer, R.Selic; Present State of GaPO4 Research. Proc. 5th European Frequency and Time Forum p.143-147 (1991).
[15] E.Philippot, A.Ibanez, A.Goiffon, B.Capelle, A.Zarka, J.Schwartzel, J.Detaint. Crystal growth and physical characterisation of $GaPO_4$; Proc. 6th European Frequency and Time Forum p.383-388 (1992).
[16] W.Wallnöffer, J.Stadler, P.Krempl, Temperature dependence of elastic constants of $GaPO_4$ and its influence on BAW and SAW devices. Proc. 7th European Frequency and Time Forum p.653-657 (1993).
[17] J.Detaint, J.Schwartzel, C.Joly, A.Zarka, B.Capelle, Y.Zheng, Y.Toudic, E.Philippot; Energy trapping resonators and integrated filters using the AT cut of quartz like materials. Proc. 1992 IEEE Int. Frequency Control Symposium p.639-647 (1992).
[18] E.Philippot, A.Ibanez, A.Goiffon, M.Cochez, A.Zarka, B.Capelle, J.Schwartzel, J.Detaint A quartz like material: Gallium phosphate $GaPO_4$, crystal growth and characterization. J. Crystal Growth v.130 p.195-208 (1993).
[19] B.Capelle, A.Zarka, J.Schwartzel, J.Detaint, Y.Zheng, A.Ibanez, E.Philippot; X-ray topography study of gallium phosphate crystals and resonators, Proc. 1993 IEEE Int. Frequency Control Symposium p.813-820 (1993).
[20] W.Wallnöfer, P.W. Krempl; A theoretical study of surface acoustic waves in layered quartz homeotype structures Proc. 8th European Frequency and Time Forum 1994.
[21] M.Cochez, J.D.Foulon, A.Ibanez, A.Goiffon, E.Philippot, A.Zarka, J.Schwartzel, J.Detaint. Crystal growth and characterisation of a quartz like material: $GaPO_4$; Proc. of the 1st European Workshop on Piezoelectric Materials Dec. 1993 Montpellier France J. de Physique IV vol 4 n°2 p.C2-183,188.
[22] O.V.Svereva, Yu.M.Mininzon, L.N. Demianets; Hydrothermal growth of OH-group free crystals $AlPO_4$ and $GaPO_4$: The way of twin reducing. To appear in the Proc. of the 1st European Workshop on Piezoelectric Materials Dec. 1993 Montpellier France J. de Physique IV vol 4 n°2 p.C2-19,24.
[23] A.A. Kaminski, B.V. Mill, G.G.Khodzhabagyan; Investigation of Trigonal $(La_{1-x}\ Nd_x)Ga_5SiO_{14}$ crystals: I. Growth and Optical Properties J. Phys. Stat. Sol. V.A80 p.387 (1993).
[24] A.A.Kaminski, I.M.Sylvestrova, S.E.Sarkisov, G.A. Denisenenko; Investigation of Trigonal $(La_{1-x}\ Nd_x)Ga_5SiO_{14}$ crystals: II Spectral Laser and Electromechanical Properties; J. Phys. Stat. Sol. V.A80 p.607 (1993).
[25] S.A.Sakharov, I.M.Larionov, A.V.Medvedev; Application of Langasite Crystals in monolithic filters operating in shear modes; Proc. 1992 IEEE Int. Frequency Control Symposium p.713-723 1992.
[26] A.N. Gotalskaya, D.I. Drezin, V.V.Bezdelkin, V.N.Stassevich Pecularities of Technology, Physical properties and Applications of a new Piezoelectric Material: Langasite $(La_3Ga_5SiO_{14})$; Proc. 1993 IEEE Int. Frequency Control Symposium p. 339-346 (1993).
[27] I.M.Silvestrova, V.V.Bezdelkin, P.A.Senyushenkov, Yu.A.Pisarevsky. Present State of $La_3Ga_5SiO_{14}$ Research. Proc. 1993 IEEE Int. Frequency Control Symposium p. 348-350. (1993).
[28] I.M. Silvestrova, Yu.V.Pisarevsky, V.V. Bezdelkin, P.A. Senyushenkov; New Piezoelectric Materials; Proc. 1993 IEEE Int. Frequency Control Symposium p. 351-352 (1993).
[29] V.V. Bezdelkin, E.E. Antonova Fundamental characteristics of thickness vibrations of langasite plates for BAW resonators. J. de Physique IV vol 4 n°2 p.C2-139,156.

[30] M.F.Dubovik, I.A.Andreyev, Yu.S.Shmaly; Langasite ($La_3Ga_5SiO_{14}$) an optical piezo-electric: Growth and Properties; Proc.1994 IEEE Frequency Control Symposium. (these procedings).
[31] G.D.Mansfeld Measurement of Acoustic Wave Attenuation in $La_3Ga_5SiO_{14}$ using HBAR Technique. Proc.1994 IEEE Frequency Control Symposium. (these procedings).
[32] A.Zarka, B.Capelle, Y.Zheng, J.Detaint, J.Schwartzel; Stroboscopic X-ray topography of quartz resonators. Proc. 42nd Annual Frequency Control Symposium p. 85-92 (1988).
[33] B.Capelle, A.Zarka, J.Schwartzel, J.Detaint, E.Philippot, J.P.Denis; Characterization of Piezoelectric Materials: Old and New Crystals; Proc. of the 1st European Workshop on Piezoelectric Materials; Dec. 1993 Montpellier France. J. de Physique IV vol 4 n°2 p.C2-123,134.
[34] B.J. James A new measurement of the elastic and dielectric constants of quartz; Proc. 42nd Annual Frequency Control Symp. p.146-154 (1988).
[35] D.S. Bailey, J.C.Andle, D.L.Lee, W.Soluch, J.F.Vetelino, B.H.T.Chai Temperature dependence of the material and acoustic wave parameters of berlinite; Proc. 1983 IEEE Ultrasonic Symp. p.335-340.
[36] W.Wallnöfer, P.W.Krempl, A. Asenbaum; Determination of the elastic and photoelastic constants of quartz-type GaPO4 by Brilloin scattering; Phys Rev. Vol.B49 n°15 p.10075-10080. (1994).
[37] J.Detaint, H.Carru, J.Schwartzel, C.Joly, E.Philippot; Optimization of the design of thickness shear devices using the AT cut of berlinite; Proc. 5th.European Frequency and Time forum p.30-40 (1991).
[38] J.Detaint, J.Schwartzel, C.Joly, E.Philippot, Energy Trapping in Plane and Corrugated Resonators; Proc. 41st Annual Frequency Control Symposium p.314-324 (1987).
[39] J.Detaint, H.Carru, J.Schwartzel, C.Joly, B.Capelle, A.Zarka. Determination of the cut-off frequency of unelectroded plates and effects of local thickness modifications in plane resonators. Proc. 43rd Annual frequency Control Symposium p.563-573 (1989).
[40] H.F. Tiersten; Analysis of trapped-energy resonators operating in overtones of coupled thickness shear and thickness twist. J. Acoust. Soc. Amer. v.59 n°4 p.879-883 (1976).
[41] H.F.Tiersten, R.C.Smythe; An analysis of contoured resonators operating in overtones of coupled thickness shear and thickness twist. J. Acoust Soc. Amer. V. 65 n°6 p.1455-1460 (1979).
[42] D.S Stevens, H.F.Tiersten; An analysis of doubly rotated quartz resonators utilizing essentially thickness modes with transverse variations; J. Acoust. Soc. Amer. V. 79 n°6 p.1811-1825 (1986).
[43] H.F.Tiersten, B.J. Lwo, B.Dulmet; An analysis of transversely varying thickness modes in trapped energy resonators with shallow contours. Proc. 1994 IEEE Frequency Control Symposium (these proceedings).
[44] J.Detaint, H.Carru, J.Schwartzel, B.Capelle, A.Zarka; Linear models of the contoured resonators; Proc. 42rd Annual Frequency Control Symposium p.19-28 (1988).
[45] S.Laffey, M.Hendrickson, J.Vig, Polishing and Etching Langasite Crystals; Proc.1994 IEEE Frequency Control Symposium. (these procedings).
[46] B.Capelle, A.Zarka, J.Détaint, J.Schwartzel, A.Ibanez, E.Philippot, J.P.Denis; Proc. 1994 IEEE Frequency Control Symposium (these proceedings).
[47] A.Ballato, E.R Hatch, M.Mizan, B.H.T.Chai, R.S.Tilton, T.J.Lukaszek; Lateral field excitation of berlinite; Proc 38th Annual Frequency Control Symposium p.206-224 (1984)

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

GROWTH OF CRACK-FREE 3-INCH DIAMETER LITHIUM TETRABORATE SINGLE CRYSTALS BY CZOCHRALSKI METHOD

Ryuichi. Komatsu, Satoshi Uda(*), Kazuyasu Hikita and Tadashi Sugihara

Central Research Institute, Mitsubishi Materials Corporation,
1-297 Kitabukuro-cho, Omiya, Saitama 330, Japan
(*) Advanced Products Division, Mitsubishi Material Corporation,
1-5-1 Marunouchi, Chiyoda-ku, Tokyo 100,Japan

Abstract: The mode of occurrence and the origin of cracking of 3-inch diameter $Li_2B_4O_7$ single crystal during growth have been investigated. From these considerations, growth conditions for crack-free single crystal have been examined and crack-free 3-inch diameter lithium tetraborate ($Li_2B_4O_7$) single crystals by Czochralski method have been successfully grown.

Introduction

Lithium tetraborate ($Li_2B_4O_7$) is a piezoelectric material which belongs to the point group 4mm [1]. This single crystal have attracted much attention as a surface wave (SAW) substrate for high frequency because of its high coupling factors (k^2) and low temperature coefficient of frequency (TCF) [2,3] since whatmore[4] first reported. $Li_2B_4O_7$ melts congruently, and consequently it is grown by the Czochralski [5,6] and Bridgman Methods [7]. The most important practical points for growing $Li_2B_4O_7$ single crystals are to prevent cracking and incorporation of voids at the solid-liquid interface during growth. Voids are generated in the melt close to the solid-liquid interface by the accumulation of impurity (almost H_2O) which is contained in starting materials, and are incorporated into the crystal by the fast solidification [8]. The incorporation of voids can be avoided by the use of H_2O-free starting materials. On the other hands, this crystal easily cracks during growth due to the strong habit of specific parting plane to <1,-1,2> direction and high temperature gradient above the melt when $Li_2B_4O_7$ single crystal are pulled to the <110> direction. Resolving this crystal cracking problem is essential to achieving good reproducibility and quality in crystal growth. The growth of crack-free crystals for commercial use has been successfully investigated by various people: Matsumura et al. [5] for the relationship between temperature gradient above the melt and crystal crack, Kamiyama et al. [9] and Sugawara et al. [10] for the relationship between the disarray of growth ridge and crystal crack and Komatsu et al.[11] for the crystal cracking due to the change of melt properties.

We have been studying the increase of growth rate of 3-inch diameter $Li_2B_4O_7$ single crystals by the Czochralski method for the decrease of wafer cost. In our experiments, the increase of growth rate brings about crystal crack, and thereby it needs to clear the origin of crack in order to increase the growth rate.

In this paper, we have examined the mode of occurrence and origin of crystal cracking during growth, and have considered to establish the growth conditions of crack-free 3 inch-diameter $Li_2B_4O_7$ single crystals.

Experimental Procedure

3 inch-diameter $Li_2B_4O_7$ single crystals were grown by the Czochralski (CZ) technique using R.F. heating and an automatic diameter control (ADC) system[11]. The starting material was purified $Li_2B_4O_7$ polycrystalline powder with the molar ratio (B/Li) of 2.00. The growth conditions are summarized in Table 1.

The crystal diameter was 80mm and the length of grown crystal was 100mm. A Pt-Pt:13%Rh thermocouple was placed at the bottom of the Pt crucible, which was used to monitor the melt temperature during the growth. A schematic diagram of the furnace is illustrated in Fig.1. Cracks were classified by the occurrence position of crack and the character of each classified crack was examined.

0-7803-1945-1/94 $3.00 © 1994 IEEE

From those character, the origin of each classified crack was considered.

Table 1. Growth conditions for $Li_2B_4O_7$ single crystals.

Crystal	Growth axis	<110>
	Diameter	80mm ϕ
Control parameters	Growth rate	0.6mm/h
	Rotation rate	0.2 - 5 rpm
Crucible	Pt	130mm ϕ ×130mmh
Atmosphere		Air
Space length (d)		20, 0, -30 mm

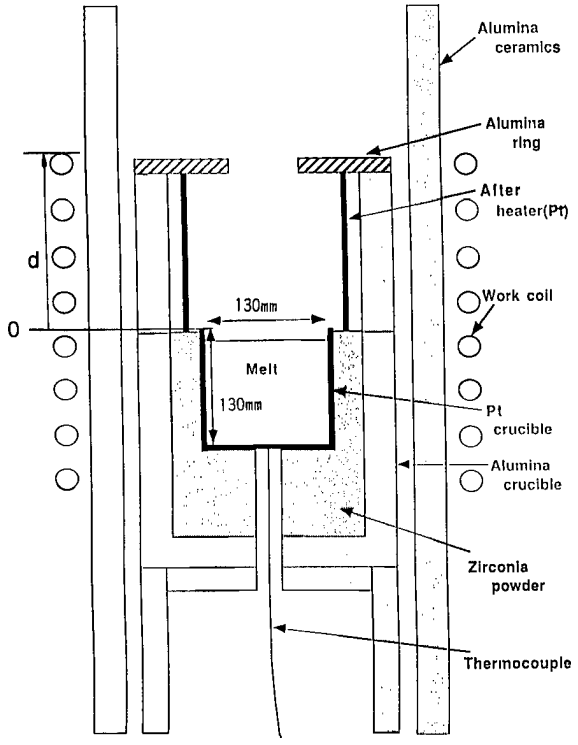


Fig.1. A schematic diagram of furnace system.

Results and Discussions

Crystal growth

Crystal crack occurrence rate during growth was 70 to 80% at the position of d=20mm which is shown in Fig.1. According to the lowering of d, this occurrence rate became small and was almost 0% at the position of d=-38mm. Figure 2 shows crack-free crystals successfully grown under this condition.

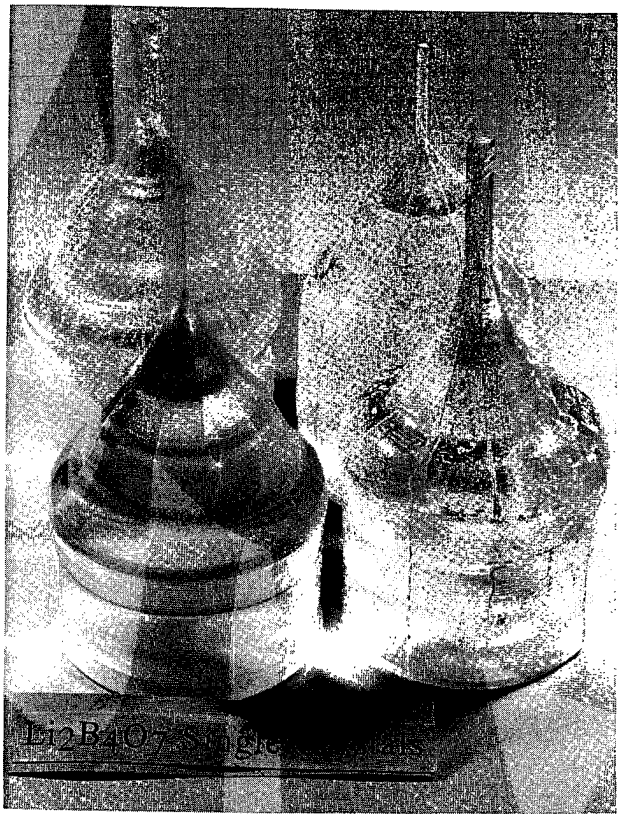


Fig.2. Grown 3 inch-diameter $Li_2B_4O_7$ single crystals.

The effect of crystal rotation rate on crystal cracking was not clear. The orientation of cracks in crystal are varied, although <112> is predominantly observed, as same as that in 2-inch diameter crystal [10]. Although the crystal cracking during cooling is a few, there are main two modes of occurrence of cracking during the growth of 3 inch-diameter single crystal. One occurs over 60mm diameter at a shoulder formation. The other occurs at a later stage of growth after the crystal attains a constant diameter.

Crack at a shoulder formation

This crack occurs when crystal diameter is 60 to

80mm. The example of this crack is shown in Fig.3. The character of this crack is large concave interface to melt as shown in Fig.3. This shape of interface is originated from crystal remelt [12] because the decrease of crystal weight is observed before cracking. Growing crystal is eventually separated from the melt with the progress of crystal remelt, and this separation leads to crystal cracking. It is considered that the crystal remelt is due to the accumulation of latent heat which evolves at the interface during growth. Whenever crystal remelt occurs, the temperature at crucible bottom and its fluctuation increases. Therefore this cracking can be prevented by the temperature control at crucible bottom until crystal diameter attains from 60 to 80mm.

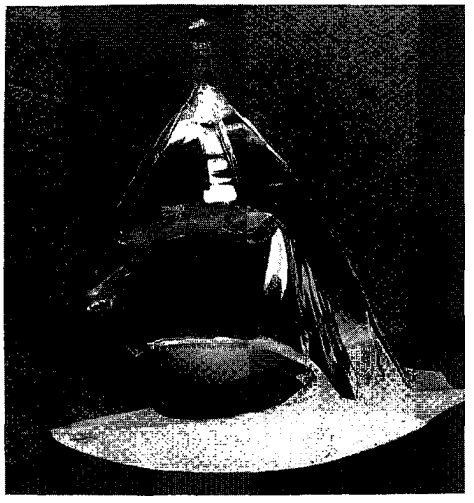


Fig.3. Crack at a shoulder formation.

Crack at a later stage of growth

According to the lowering of d (20mm, 0mm, -38mm), crack occurrence rate became to be small. From the microscopic observation of cracked and crack-free crystal, we found that striation which is banded structure on the crystal surface changes with the decrease of d. Figure 4 shows striations in cracked (d=20mm) and crack-free (-38mm) crystal. Figure 4(a) shows striations, grown under the condition of 70 to 80% crack occurrence rate (d=20mm), while that in crystal at 0% crack occurrence rate (d=-38mm) is shown in Fig.4(b). Striations in crack-free crystal are regular, while that in cracked crystal are irregular. It is reported in Si [13] and semiconductor compound [14] crystals that striation is related to the fluctuation of growth rate. We measured the fluctuation of growth rate of lithium tetraborate which is observed by the change of crystal weight during growth. Figure 5 shows

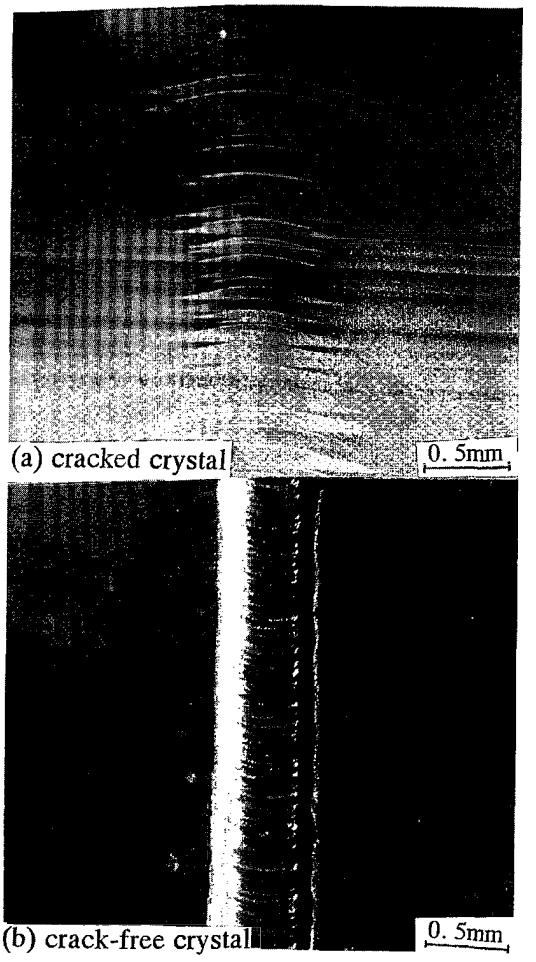


Fig.4. Striations in crack-free and cracked crystal.

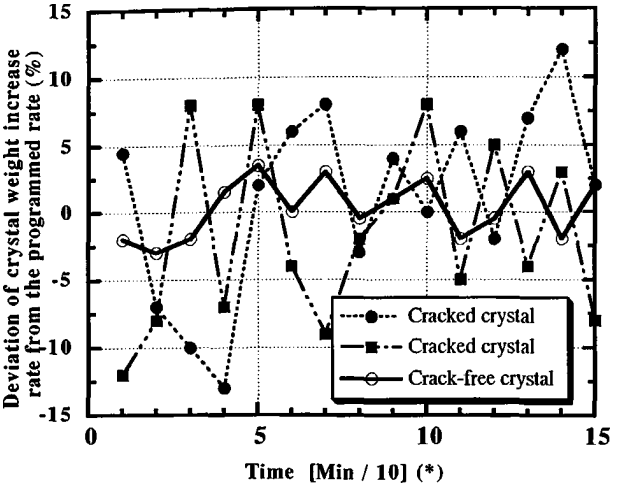


Fig.5. The deviation of the increase rate of crystal weight from the programmed rate.
(*) 0 indicates the time when the length of grown crystal trunk attains 50mm.

the deviation of the increase rate of crystal weight from the programmed rate.

Figure 5 shows that the deviation in a cracked crystal is +-12%, while that in a crack-free crystal is less than +-3%. These results indicate that the large deviation comes from the fluctuation of growth rate, the growth rate fluctuation in cracked crystal is larger than that of crack-free crystal and crack occurs when the fluctuation of growth rate is large.

Chikawa [15] reported in Si that striation is related to the irregular temperature oscillation in melt. Brice [16] shows that the fluctuation of growth rate results from temperature fluctuation in melt below the interface. It was considered that temperature fluctuation in melt below interface is essential to preventing crystal crack, and thereby we examined the temperature oscillation in melt below interface. Figure 6 shows a schematic diagram of measurement system. Pt-Rh thermocouple was inserted into melt below interface, as shown in Fig. 6. Temperature oscillation is normally generated by melt convection. There are two factors controlling melt convection. One is crystal rotation which dominates forced convection. the other is work-coil position which dominates natural convection. work-coil position is denoted as d, as shown in Fig.1.

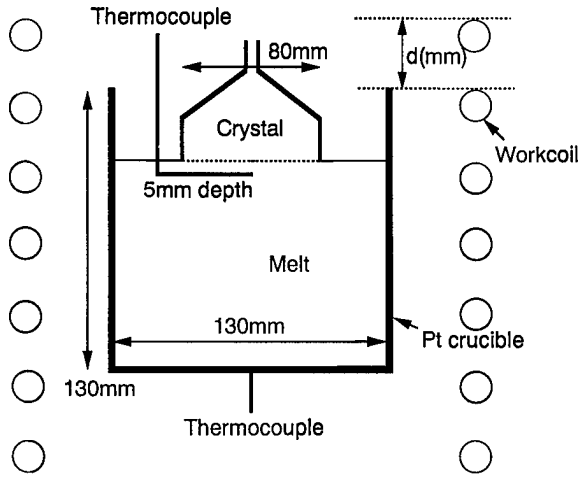


Fig.6. Arrangement of thermocouples, work-coil and Pt crucible. (*) d is defined as $d=H_W-H_C$, where H_W is the height of the top of work-coil from ground level and H_C is the height of the top of Pt crucible from ground level.

The measurement of temperature fluctuation in melt

Typical examples of temperature, recorded in melt below the interface and at the bottom of crucible, are shown in Fig.7.

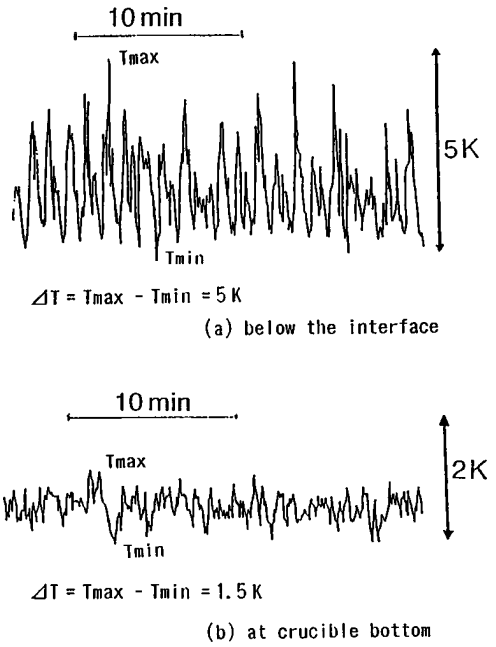


Fig.7. Typical examples of the temperature fluctuation. Crystal rotation rate is 0.3rpm.

This temperature fluctuation seems to be irregular. The difference between maximum and minimum temperature for a period of 10 minutes is regarded as the maximum temperature fluctuation and the relationship between this maximum fluctuation and crystal rotation rate are shown in Fig.8.

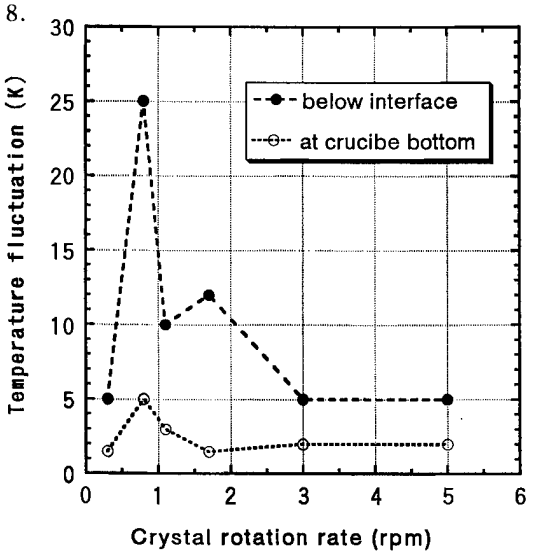


Fig.8. The relationship between maximum temperature fluctuation and crystal rotation rate.

Figure 8 shows the following three important things; the behavior of the maximum temperature fluctuation in crucible bottom is almost as same as that in melt below the interface, the maximum fluctuation at the interface is over 5K and that at crucible bottom is over 1.5K in the range of 0-5 crystal rotation rate (rpm). The origin of largest irregular temperature fluctuation at 0.8rpm is not clear, however it is almost certain that unsteady strong convection of melt is produced when crystal rotation rate is 0.8rpm. The relationship between the crystal rotation and the maximum temperature fluctuation at crucible bottom on three d value (20,0,-38mm) are shown in Fig.9.

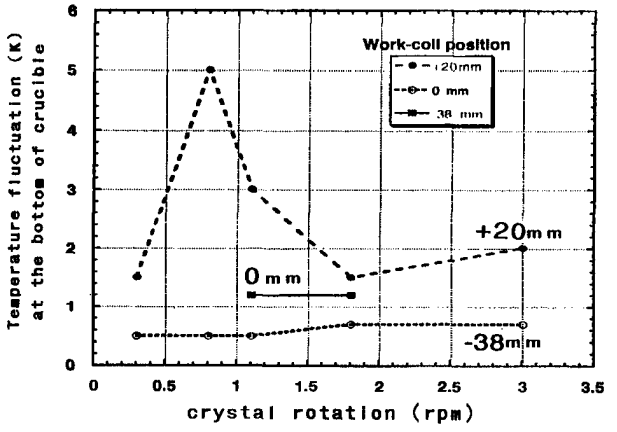


Fig.9. The change of maximum temperature fluctuation at crucible bottom with crystal rotation rate and d values (20,0,-38mm).

Figure 9 shows that the maximum fluctuation decreases according to the lowering of d value, and become to about 0.7K in d=-38mm. According to the lowering of d value, temperature gradient below interface become low, and thereby the melt viscosity in crucible become high. This high viscosity makes melt convection weak and temperature fluctuation which originate from melt convection become to be small. This small fluctuation at crucible bottom indicates that the temperature fluctuation in melt below the interface may be also small.

The effect of temperature fluctuation in melt on crystal growth

The temperature fluctuation may be originated from the convection of melt. Unsteady convection of melt results the irregular temperature fluctuation, and the larger fluctuation causes from stronger convection of melt. It is known that unsteady convection brings about the variation of the growth rate in the growth of metal and semiconductor crystal [17]. We observed that the temperature fluctuation changed from regular to irregular when crystal remelt occurred at the shoulder formation. This shows that unsteady convection may be concerned with the change of crystal growth rate.

We examined the variation of increase rate in crystal weight during growth and revealed that the variation of increase rate in crystal weight was about +-12% in d=20mm, and less +-3% in d=-38mm [18]. Crystal length grown for 10 minutes is about 0.06mm and too short for the crystal diameter to change, and thereby this variation of weight increase mainly results from that of growth rate at the interface. This shows that the growth rate of $Li_2B_4O_7$ single crystal actually is not constant in spite of a constant pulling rate. The large variation of growth rate corresponds to the large temperature fluctuation which results from the strong convection of melt. Crystal cracking rate became to be almost 0% under the condition of d=-38mm which showed the small temperature fluctuation. Hence it is concluded that the origin of crystal cracking is mainly due to the large variation of crystal growth rate at the interface.

There are two possibilities of crystal cracking by the growth rate variation. One is lattice mismatch induced by the variation of impurity concentration due to that of growth rate in the growing crystal. OH^- is detected as main impurity in grown crystal. The other is polycrystallization at the growing interface when the growth rate reaches largest. Since cracking at a later stage of growth after the crystal attains a constant diameter begins at the interface, it may be considered that this cracking results from polycrystallization. On the other hand, crystal cracking during cooling is frequently observed. This crack may be originated from the intense variation of impurity concentration.

Conclusions

There were two modes of occurrence cracking of 3-inch diameter $Li_2B_4O_7$ single crystal during growth. One occurred at a shoulder formation and originated from crystal remelt. The other occurred at a later stage of growth and was mainly related to large growth rate fluctuation. Growth rate fluctuation originated from unsteady convection of melt. Crack-free 3-inch diameter lithium tetraborate ($Li_2B_4O_7$) single crystals by Czochralski method has been successfully grown by the suppressing crystal remelt and large temperature fluctuation in melt.

Reference

[1] J. Krough-Moe, "The Crystal Structure of Lithium Diborate, $Li_2O2B_2O_3$", Acta Cryst.,15, 1962, 190-193.

[2] Y. Ebata, H. Suzuki, S. Matsumura and K. Fukuda, "SAW Propagation Characteristics on $Li_2B_4O_7$", Jpn. J. Appl. Phys.22, Suppl. 22-2 1983, 160-162.

[3] M. Adachi, T. Shiosaki, H. Kobayasi, O. Ohnishi and K. Kawabata, "Temperature Compensated Piezoelectric Lithium Tetraborate Crystal for High Frequency Surface Acoustic Wave and Bulk Wave Device Applications", 1985 IEEE Ultrasonic Symp. Proc., Oct. 1985, 228-232.

[4] R.W. Whatmore, N. M. Shorrocks, C. O'Hara, F. W. Ainger and I. W. Young, "Lithium Tetraboarte: A Temperature-compensated SAW Substrate Material", Elect. Lett., 17, 1981, 11-12.

[5] S. Matsumura, T. Omi, N. Yamaji and Y. Ebata, "A 45X-cut $Li_2B_4O_7$ Single Crystal Substrate for SAW Resonators", 1987 IEEE Ultrasonic Symp. Proc., 1987, 247-252.

[6] Shi-ji Fan, W. Wang, J-J Xiang and J-X Xu, "Bridgman Growth of $Li_2B_4O_7$ Crystals", J. Cryst. Growth , 99, 1990, 811-814.

[7] K. Miwa, Proc. Nihon Kogyokai Koenkai, 1988, 279-280, [in Japanese].

[8] R. Komatsu, T. Suetsugu, S. Uda and M. Ono, "Growth and Characterization of Lithium Tetraborate Single Crystal", Ferroelectrics, 95, 1989, 103-107.

[9] K. Kamiyama, S, Murakami and M. Isogami, "Growth of large sized $Li_2B_4O_7$ Single Crystal using Resistance Heating CZ furnace", Proc. Jinko-kobutsu Toron-kai, Sep., 1993, 23-24, [in Japanese].

[10] T. Sugawara, R. Komatsu and T. Sugihara, "Growth of 3-inch Diameter $Li_2B_4O_7$ Single Crystal using Resistance Heating Furnace", to be published, Jpn. J. Appl. Phys., 1994.

[11] R. Komatsu, S.Uda and K. Hikita, "Mode of Occurrence and Cause of Cracking of $Li_2B_4O_7$ Single Crystals during Growth by Czochralski Method", Jpn. J. Appl. Phys., 32, 1993, 4364-4366.

[12] K. Kitamura, N. Ii, I. Shindo and S. Kimura, "Interface Shape and Horizontal Variations of Al and Ga Contents in Substituted YIG Single Crystals Grown by the Floating Method", J. Crystal Growth, 46, 1979, 277-285.

[13] E. Kuroda, H. Kozuka and Y. Takano, "Effect of Temperature Oscillations at the Growth Interface on Crystal Perfection", J. Crystal Growth, 68, 1984, 613-623.

[14] G. Muller' "Crystal growth from Melt, Crystals 12", Springer-Verlag, 1988, pp.92

[15] J. Chikawa and S. Shirai, "Melt growth of Silicon", Nihon Kesshou Seichou Gakkai-shi, 4, 1977, 2-22, [in Japanese with English abstract].

[16] J. C. Brice, "The Growth of Crystal from Liquids", North-Holland, 1973, 159.

[17] G. Muller' "Crystal growth from Melt, Crystals 12", Springer-Verlag, 1988, 77

[18] R. komatsu, T. Sugihara and S. Uda, "Growth of 3-inch Diameter Lithium Tetraborate Single Crystals", to be published, Jpn. J. Appl. Phys., 1994.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DEFECTS IN THE BRIDGMAN GROWN LITHIUM TETRABORATE OF THREE-INCH DIAMETER

Shi-Ji Fan, Ren-Ying Sun and Jue Wang

Shanghai Institute of Ceramics, Chinese Academy of Sciences
1295 Ding Xi Road,Shanghai 200050,P.R.China

Defects in lithium tetraborate $Li_2B_4O_7$ (LBO) and their influence on performance of devices are concerned by crystal material researchers and device engineers with production of 3-inch diameter LBO crystal by the Bridgman (BR) technique and application and design of surface acoustic wave (SAW) devices fabricated on LBO substrate. The defects in the BR-grown LBO of 3-inch diameter are observed by chemical etching and microscopy. The main defects in the crystal are sub-grain boundaries and dislocations. Growth striation, scattering particle and twin occur in the crystal occasionally. The configuration of the defects, their origins and lessening methods are described also in this paper.

1. Introduction

Lithium tetraborate $Li_2B_4O_7$ (abbr:LBO) crystal as a new piezoelectric and non-ferroelectric substrate material is developed into pilot production by the modified Bridgman technique [1-3] and LBO wafers are fabricated for surface acoustic wave (SAW) devices with middle band width, high stability of central frequency and miniaturization for industrial manufacture of movable communication equipments [4-7]. LBO is a stoichiometric compound in Li_2O-B_2O_3 system with congruently melting point 917°C and belongs to tetragonal crystallographic system. Macro imperfection, core and crack, and X-ray topography of dislocation of density $10^{3\text{-}5}cm^{-2}$ in CZ-grown LBO crystals were reported [5,6,8,9]. Although the BR growth of LBO crystal of three-inch diameter without core, crack, growth striation, scattering particles and twin has been carried out, the researchers of the crystal and devices are concerned about defects in the BR-grown LBO crystals and their influence on device performance.

We have investigated systematically the main defects in the BR-grown LBO crystals of three-inch diameter by chemical etching and X-ray topography and their origin and elimination meanwhile we optimize the growth conditions of large diameter LBO crystals. The results observed by chemical etching are reported in this paper.

2. Experimental

LBO crystals of three-inch diameter are grown by the modified BR technique as the procedure described in reference [1]. The orientations of the crystals are <110> or <001> directions. The samples for optical observation are cut from as-grown boules, finely lapped, sometimes polished in the standard procedure. The defects are revealed by chemical etching in 25% aqueous acetic acid, 15-25°C, 30-120 min and observed by the eyes or with Olympus microscope BH-P-2 and others.

3. Results and discussion

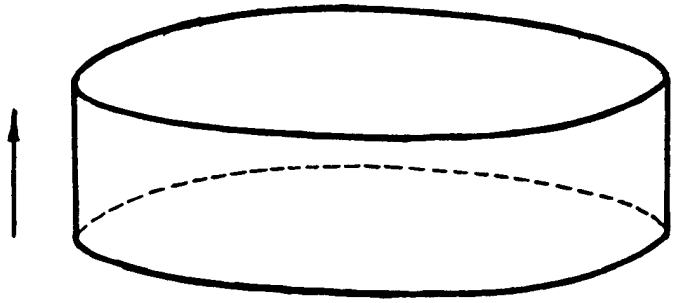
3-1. Sub-grain boundary

After chemical etching, sub-grain boundary on the cross-section (110) face of <110>-grown Φ3" LBO boule can be seen obviously by the eyes through reflection of light from the surface. There is a big difference between the boules, on the (110) surfaces of some boules very dense sub-boundaries such as the net configuration are demonstrated while for other boules, only thin sub-boundaries on the surface. In order to investigate the origin and propagation of sub-boundary during LBO growth, two ends, top and bottom cross-sections, of the boules of 20-50 mm length are observed. As Figure 1 indicates, dense net sub-boundaries on the bottom surface decrease obviously to thin sub-boundaries on the top surface through 20mm growth length. The original sub-boundaries merge with each other or change into range of separated dislocations during the optimized growth. Bulk defects such as core and inclusions create often the original sub-boundaries on the place of diminishing these bulk defects. In Fig. 2, the arrows indicate the inclusions which create the sub-boundary along themselves. Fig. 3

0-7803-1945-1/94 $3.00 © 1994 IEEE

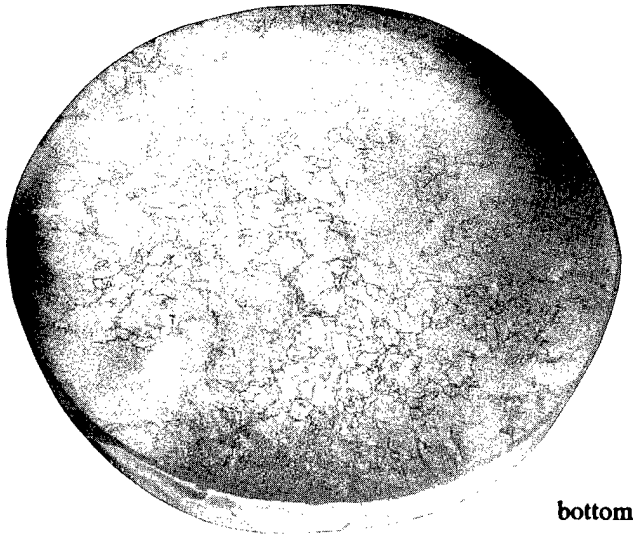


top surface



growth direction <110>

LBO as-grown boule



bottom surface

Fig.1 The dense net sub-boundaries merge each other during LBO BR-growth.

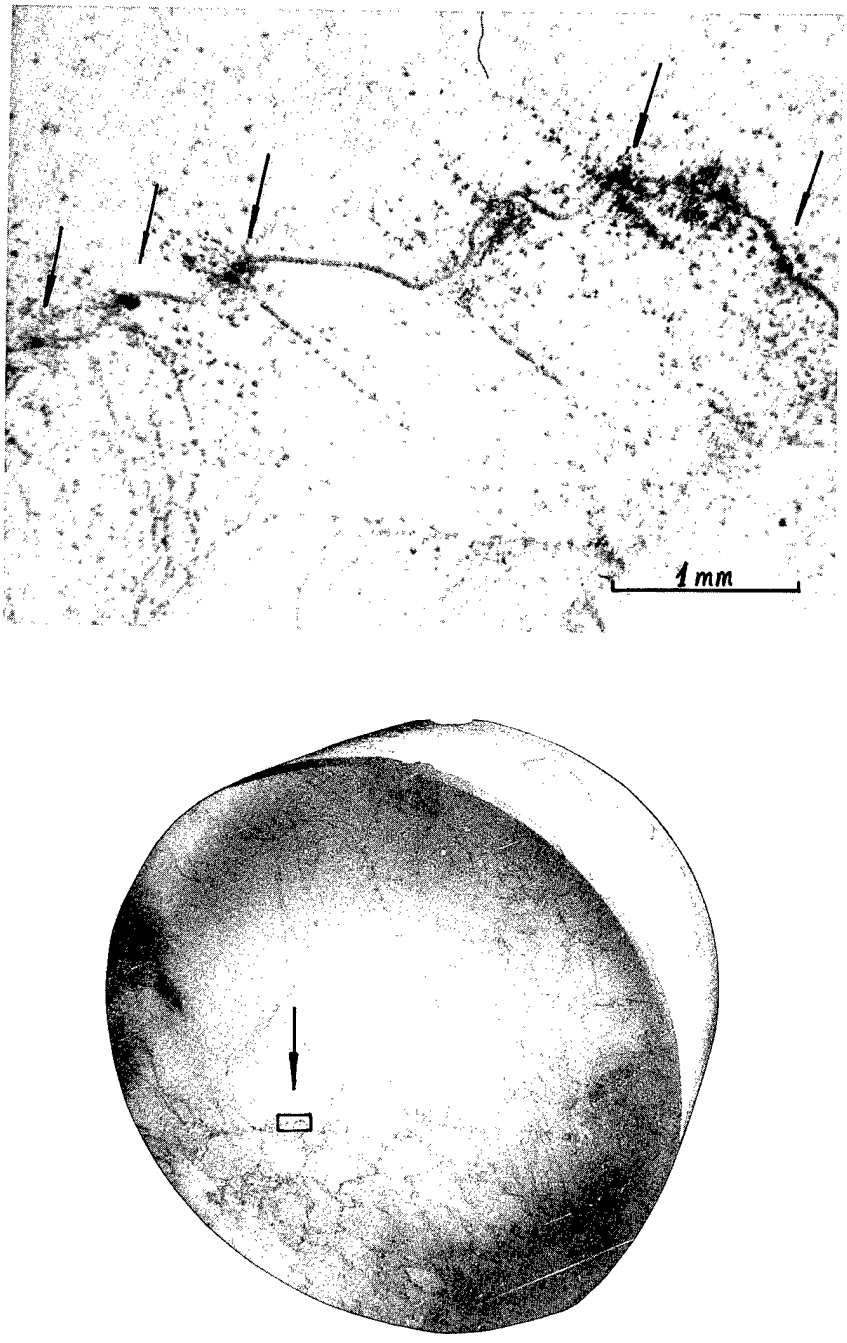


Fig.2 The inclusions along themselves create the sub-boundary in LBO BR-grown crystal.



Fig.3 Few and scattered sub-boundary configuration in the BR LBO crystal

demonstrates a LBO crystal with few and scattered sub-boundaries. As a matter of fact, the sub-boundaries propagate mainly from those in seed crystal so that the best ways of decreasing sub-boundaries in LBO are optimizing the conditions of the crystal growth, specially controlling the shape and stability of solid-liquid (s-l) interface during the growth and choosing the seed with less sub-boundaries.

3-2. Dislocation

Tetrahedral cone of etch pits on (001) face (Fig. 4) and deformed pentahedral cone of etch pits on (110) face (Fig. 5) on the polished surface of LBO crystal can be observed easily by microscope. The densities of etch pits on different areas of the crystal surfaces are quite different (Fig. 6), from $10^{2\text{-}3}$ cm^{-2} to $10^{4\text{-}5}$ cm^{-2}, depending on the distribution of the original dislocations in the seed and the density of bulk defects such as core and inclusions in as-grown LBO crystals. In most of the cases, high density of etch pits appears in the core region or around the inclusions. For example, in Fig. 5, black region at the centre is an inclusion. That means the bulk defects create the original dislocations in as-grown LBO crystal. On the other hand, the dislocations in the seed crystal propagate generally into as-grown LBO crystal. Choosing the seed with low density of dislocation is most effective for decrease of density of dislocation in as-grown boule. Another way to prepare the seed is changing the growth directions of the seed crystal according to the equivalent axis, for example, <110> seed changes into <$\bar{1}$10> or

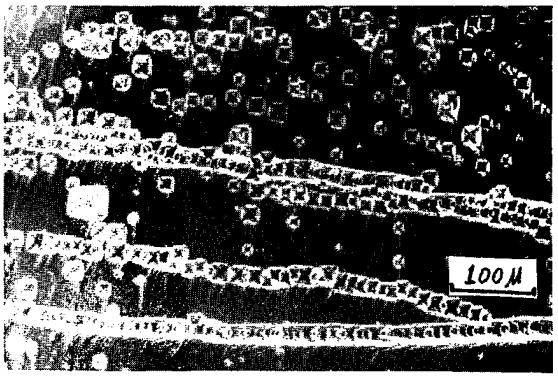


Fig.4 Tetrahedral cones of etch pits and sub-boundaries on (001) face of LBO crystal.

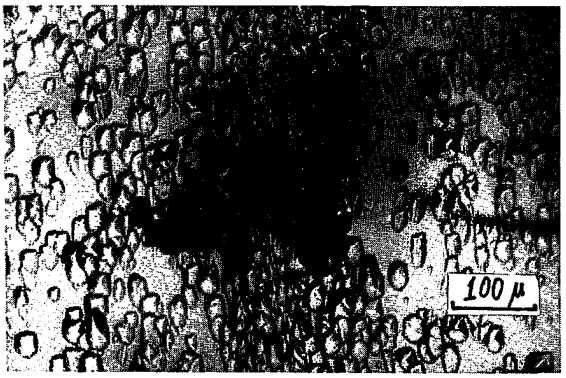


Fig.5 Pentahedral cones of etch pits around an inclusion on (110) face of LBO crystal.

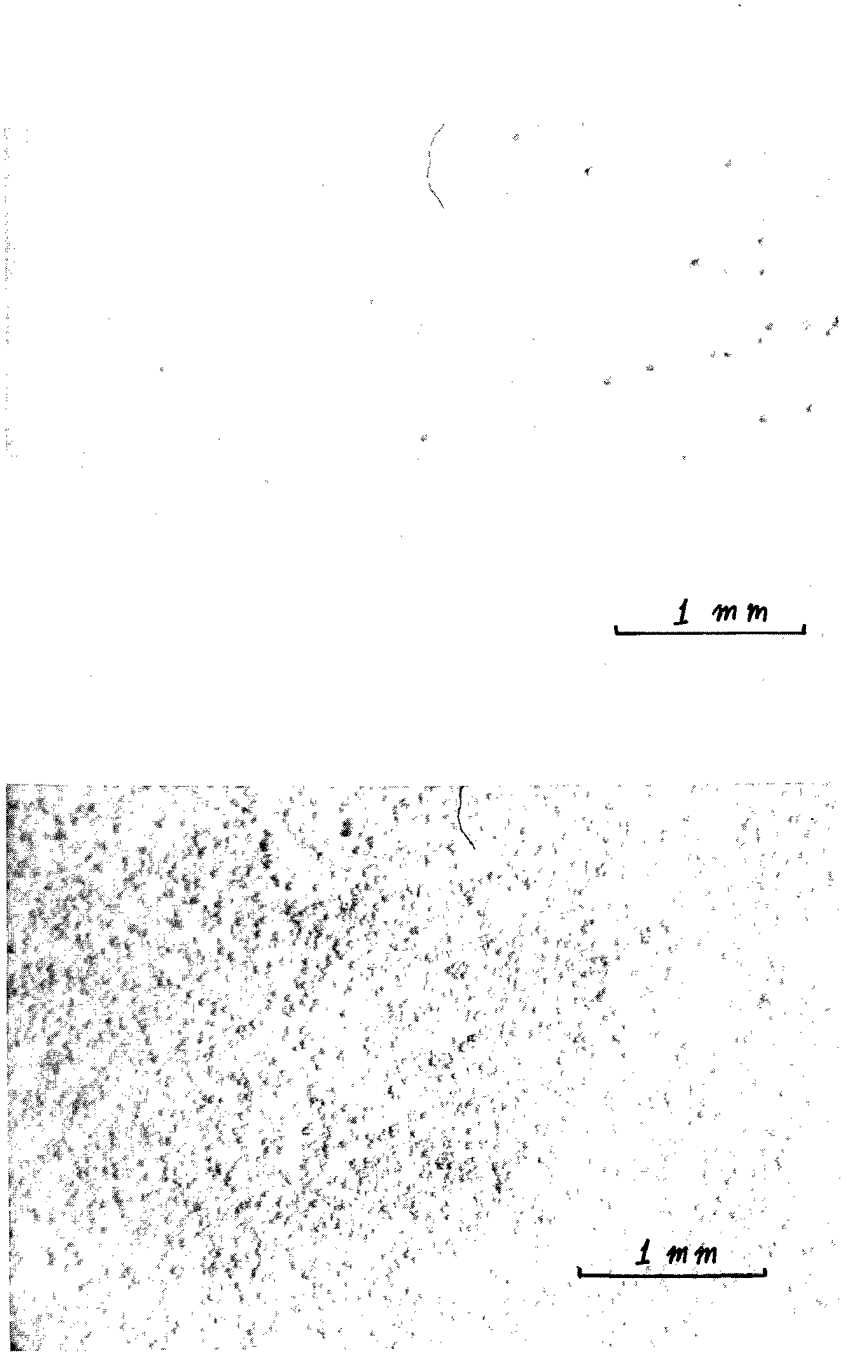


Fig.6 Low or high density of the dislocations on different areas of LBO BR-crystal.

<1$\bar{1}$0> seed. Thus the density of dislocation on the cross-section of the seed in the changed direction decreases.

3-3. Growth striation

In as-grown BR LBO crystals there is (are) while striation(s) of 1-5 mm width in the direction of growth called "growth striation" parallel to the cross-sections of LBO boule. As a matter of fact, growth striation consists of a lot of the constitutional supercooling inclusions. The shape of growth striation is in coincidence with the shape of s-l interface during the growth and the position of growth striation in LBO boule corresponds to the place on which vibration of furnace temperature occurs. Generally, the growth striation forms during the abrupt decrease of furnace temperature more than 0.3-0.5°C per hour. Meanwhile the less deviation from the stoichiometry (Li_2O/Bi_2O_3=0.2146) for melt composition tolerates the more vibration of furnace temperature during the growth without the formation of growth striation. Although the growth striation is a macro-imperfection for LBO boule, LBO wafers of 0.5-2 mm width cut in parallel to the growth striation do not contain the growth striation. Therefore, growth striation do not influence the performance of LBO wafer.

3-4. Scattering particle

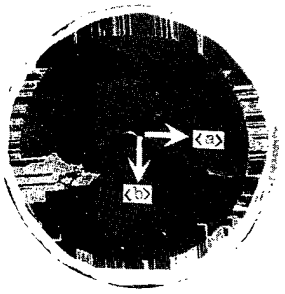
Using He-Ne Laser beam, we observed the whole volume of Φ3" LBO boules in dark circumstance, there are a few scattering particles occasionally. They are filament holes or small inclusions in several tens micrometer and form from the non-cured rift in the seed or growth interface breakdown during occasional vibration of furnace temperature or accumulation of non-stoichiometric component and impurities in local micro-regions.

3-5. Twin crystal

Twinning in LBO crystal appears occasionally also specially for <001> growth of LBO crystal. After chemical etching in acetic acid, two parts of the twin are easily observed. The etching velocities of two parts of LBO twin are quite different from each other. For example, on (001) surface, tetrahedral hills cover one part of the twin while very fine flat on another part of the twin (Fig. 7). The laminar twins create some time in <001> growing LBO at the temperature of near melting point by thermal stress from shrinkage of Pt crucible (Fig. 7). The bulk defects such as inclusions, growing rift from interface breakdown can create original twins also in as-grown LBO crystal by thermal stress so twin in LBO belongs to mechanical twin.

4. Conclusion

The main defects in the BR-grown LBO crystal of 3-inch diameter are sub-grain boundaries and dislocations. They can be decreased by choosing the seed crystal containing less imperfection and optimizing the growth conditions. The influence of the sub-boundaries and dislocations on performance of devices fabricates on LBO substrate is not clear and the comparison of relative data from the substrates with quite different defects configuration is needful to make conclusion.



a. Laminar twins (x0.6).



b. Tetrahedral hills covered on one part and flat surface on another part of LBO twins.

Fig.7 Two parts of twins on (001) face of LBO crystal

Acknowledgments

The authors would like to thank colleagues in their laboratory for LBO crystal growth.

References

[1] 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, pp.811-814, 1990.

[2] 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 pp.46-51, 1991.

[3] S-J Fan, "Properties, Production and Application of New Piezoelectric Crystal Lithium Tetraborate $Li_2B_4O_7$", Proc. 1993 IEEE Frequency Control Symposium,1993, pp.353-358.

[4] 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, pp.123-125, 1987.

[5] S. Matsumura, T. Omi, N. Yamaji and Y. Ebata, "A 45° X-cut $Li_2B_4O_7$ Single Crystal Substrate for SAW Resonators", Proc. IEEE Ultrasonics Symp., 1987, pp.247-250.

[6] H. Abe, H.Saitou, M. Ohmura, T. Yamada and K. Miwa, "Lithium Tetraborate ($Li_2B_4O_7$) SAW Resonators ", Proc. IEEE Ultrasonic Symp., 1987, pp.91-94.

[7] M. Ohmura, H. Abe, K. Miwa and H. Saito, "A 900 MHz SAW Resonator Filter on $Li_2B_4O_7$ Crystal", Proc. IEEE Ultrasonic Symp., 1990, pp.135-138.

[8] D.C.Robertson and I.M.Young, "The Growth and Growth Mechanism of Lithium Tetraborate", J. Material Sci., Vol. 17, pp.1729-1738, 1983.

[9] M. Adachi, T. Shiosaki and A. Kawabata, "Crystal Growth of Lithium Tetraborate ($Li_2B_4O_7$)",Jpn. Appl. Phys., Vol.24, Suppl. 24-3, pp.72-75, 1985.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

AN IMPROVED RESONATOR METHOD FOR THE DETERMINATION OF PIEZOELECTRIC MATERIAL CONSTANTS

John A. Kosinski, Arthur Ballato, and Yicheng Lu*

U.S. Army Research Laboratory
Electronics and Power Sources Directorate
ATTN: AMSRL-EP-MD
Fort Monmouth, NJ 07703-5601

*Rutgers, The State University of New Jersey
Department of Electrical and Computer Engineering
Piscataway, NJ 08855-0909

Abstract

This paper presents an improved resonator method for the determination of piezoelectric material constants. The improved method addresses a fundamental limitation of the measurement methods recommended in the current IEEE Standard on Piezoelectricity: the relations between vibrator response and material constants presented in the Standard are based upon the one-dimensional approximation of an essentially infinite flat plate with a uniform distribution of vibratory motion. The calculation or measurement of the vibrational amplitude distribution is a non-trivial task. The practical result is that the current IEEE 176-1987 resonator method recommendations are of limited usefulness in the determination of "intrinsic" piezoelectric material constants. The limitation can, however, readily be overcome using an improved measurement technique based on measurands unaffected by the vibrational amplitude distribution. In the improved technique, the measurands of choice are the zero-mass-loading, fundamental mode, TE antiresonance or LE resonance frequencies.

Introduction

This paper presents an improved resonator method for the determination of piezoelectric material constants. The improved method was developed during the course of a recently reported $Li_2B_4O_7$ material constants determination [1], as important discrepancies were noted between piezoelectric coupling values obtained using the resonator measurement methods recommended in the current IEEE Standard on Piezoelectricity, IEEE Standard 176-1987 [2], and those obtained using measurements of essentially plane-wave velocities. The magnitude of the discrepancies is illustrated in Table I, wherein several "effective" values of piezoelectric coupling measured using the IEEE 176-1987 resonator method recommendations are compared to the corresponding "intrinsic" values calculated using material constants derived from the measurements of plane-wave velocities. The "effective" piezoelectric coupling values are only 60%-80% of the corresponding "intrinsic" values, leading directly to discrepancies of

TABLE I
COMPARISON OF MEASURED "EFFECTIVE" AND "INTRINSIC" PIEZOELECTRIC COUPLING VALUES OF THE $\bar{c}^{(2)}$ AND $\bar{c}^{(3)}$ MODES OF $Li_2B_4O_7$.

φ	θ	Mode	k(eff)	k
0	28.2	$\bar{c}^{(2)}$	11.3	14.4
0	45	$\bar{c}^{(2)}$	13.3	20.3
45	45	$\bar{c}^{(2)}$	14.4	21.3
45	56.1	$\bar{c}^{(2)}$	16.4	24.4
0	90	$\bar{c}^{(2)}$	28.0	41.3
0	0	$\bar{c}^{(3)}$	12.1	18.4
45	0	$\bar{c}^{(3)}$	12.3	18.4
0	28.2	$\bar{c}^{(3)}$	13.3	18.9
0	45	$\bar{c}^{(3)}$	14.9	23.6
45	45	$\bar{c}^{(3)}$	14.3	23.0
45	56.1	$\bar{c}^{(3)}$	19.3	27.1

0-7803-1945-1/94 $3.00 © 1994 IEEE

10%-25% in the determination of piezoelectric stress constants, as well as secondary discrepancies of varying size in the determination of associated elastic stiffnesses.

The discrepancies between the "intrinsic" and "effective" values of piezoelectric coupling, as well as other important effects, have been identified as arising from the non-uniform distribution of vibratory motion found in practical plate resonators. Such effects are not accounted for in IEEE 176-1987. Further, the measurement and/or calculation of the vibrational amplitude distribution and its effects on the quantities of interest is a non-trivial task. Consequently, the improved measurement method is based upon a set of measurands unaffected by the vibrational amplitude distribution.

Limitation of IEEE Standard 176-1987

Various techniques for determining dielectric, piezoelectric, and elastic constants are covered in Chapter 6 of IEEE 176-1987, with resonator measurements discussed in detail in Section 6.4. The resonator method recommendations are based upon the presumption that the relations between vibrator response and material constants are *accurately* known. This is stated in the Standard as follows:

"For each of the modes of vibration analyzed in Section 4, there is a transcendental expression for the electrical impedance Z(ω) that, in the absence of losses, is exact except for the approximations made in obtaining the equation of motion and boundary conditions."

Such approximations form the primary limitation of the IEEE 176-1987 resonator method recommendations, as the relations between vibrator response and material constants presented in the Standard are based upon the one-dimensional approximation of an essentially infinite flat plate with a uniform distribution of vibratory motion. Unfortunately, some of the basic measurands used in the recommended techniques (C_1 and f_s) are strongly dependent on the non-uniform distribution of vibratory motion found in practical plate resonators. The net result is that such measurements are useful in establishing the "real world" behavior of a particular resonator design, but are of limited usefulness in the determination of piezoelectric material constants

Non-Uniform Distribution of Motion Effects

IEEE Standard 176-1987 recommends three procedures for determining piezoelectric coupling and piezoelectric constants using thin plates, namely 1) measurement of the difference between the fundamental resonance and antiresonance frequencies, 2) measurement of the motional capacitance (using either the slope of the resonator reactance near resonance, measurement of the altered resonance frequency with a series load capacitor (the so-called load frequency), or measurement of f_s, Q, and R_1), and 3) measurement of the fundamental and first or higher overtone resonances. These techniques accurately measure the "effective" piezoelectric coupling of the device under test, and as such are useful in establishing the "real world" behavior of a particular resonator design. However, the basic measurands used in the recommended techniques are strongly dependent on the particular details of the distribution of vibratory motion in the piezoelectric plate. A difficulty arises, therefore, when the measured "effective" quantities are misinterpreted as being the "intrinsic" quantities associated with the ideal one-dimensional plate resonator.

The effects of the non-uniform distribution of vibratory motion on the various elements of the Butterworth-Van Dyke equivalent electrical circuit parameters were first published by Bechmann in 1952 [3]. The "intrinsic" values of R_1, L_1, and C_1 associated with the infinite flat plate are modified by a factor Φ arising from the non-uniform distribution of motion:

$$R_1(\text{eff})=R_1/\Phi, \tag{1}$$

$$L_1(\text{eff})=L_1/\Phi, \tag{2}$$

and

$$C_1(\text{eff})=C_1\cdot\Phi. \tag{3}$$

From equation (3) it follows that

$$k^2(\text{eff})=\Phi k^2. \tag{4}$$

For the air-gap measurements of flat, unelectroded $Li_2B_4O_7$ plates as listed in Table I, Φ ranges from 0.38 to 0.61 with a mean value of 0.45.

The effects of the non-uniform distribution of motion on the critical frequencies may be found by substituting k^2(eff) (and also μ(eff) [4]) into the transcendental equations describing the resonator transimpedance; the thickness-field excitation (TE) resonance frequencies and lateral-field excitation (LE) antiresonance frequencies are among those frequencies strongly affected.

The factor Φ may be calculated from the normalized vibrational amplitude distribution $g(r,\theta)$ as

$$\Phi \equiv \frac{[1/A_e \int_{Ae} g(r,\theta)\, r\, dr\, d\theta]^2}{1/A \int_A g^2(r,\theta)\, r\, dr\, d\theta} \qquad (5)$$

wherein A represents the area of the major plate surfaces and A_e represents the area of the driving electrodes. The amplitude distribution may be calculated using the techniques developed by Tiersten and coworkers [5-8], or determined experimentally (see, for example, Sauerbrey and coworkers [9-11]]), however these are non-trivial tasks.

Measurement Procedure

The limitations of the IEEE 176-1987 recommended methods can readily be overcome using an improved measurement technique based on a properly chosen sample set of measurands unaffected by the vibrational amplitude distribution. In the improved technique, the measurands of choice are the zero-mass-loading (unelectroded), fundamental mode, TE antiresonance or LE resonance frequencies, which are by definition independent of both electrode mass loading and piezoelectric coupling values, and thus by extension are independent of non-uniform distribution of motion effects.

The critical frequencies of unelectroded plates can be measured using TE and LE air-gap test fixtures (see, for example, Guttwein, et al. [12] and Ballato, et al. [13]) in conjunction with a network analyzer/balancing bridge system as shown in Fig. 1. The frequencies are measured in a transmission mode with the shunt capacitance of the resonator balanced out. The TE resonance frequencies are measured with the air-gap as close to zero as possible ($\approx 10\mu m$), while the TE antiresonance frequencies are measured with the air-gap as large as possible (a 4.0mm air-gap is achievable using

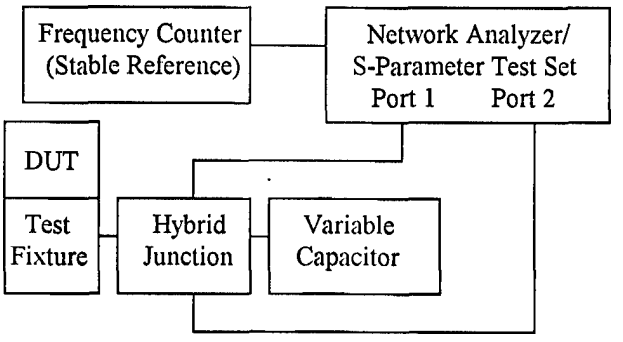


Fig. 1. Critical frequency measurement apparatus.

the test fixture described in [12]). The LE resonance frequencies can be measured using a pair of planar electrodes, separated by a gap on the order of the plate thickness, placed as close to the upper surface of the resonator as possible.

For the most reliable results, multiple harmonics of the various critical frequencies should be measured, and the results checked for self-consistency using Ballato's exact transmission line analogs of the piezoelectric plate resonator [14]. These analogs are exact for both the infinite flat plate with a uniform distribution of vibratory motion and the finite plate with a non-uniform distribution of motion; in the latter case the "intrinsic" piezoelectric coupling and mass loading are simply replaced by the corresponding "effective" values [4]. For the TE case, the resonance frequencies are related as

$$\tan X = \frac{X}{k^2 + \mu X}, \qquad X = \frac{\pi}{2} \cdot \frac{f_{R\mu}^{(M)}}{f_{Ao}^{(I)}} \qquad (6)$$

while the antiresonance frequencies are related as

$$\tan X = \frac{1}{\mu X}, \qquad X = \frac{\pi}{2} \cdot \frac{f_{A\mu}^{(M)}}{f_{Ao}^{(I)}}. \qquad (7)$$

For the LE case, the resonance frequencies are related as

$$\tan X = \frac{1}{\mu X}, \qquad X = \frac{\pi}{2} \cdot \frac{f_{R\mu}^{(M)}}{f_{Ro}^{(I)}} \qquad (8)$$

while the antiresonance frequencies are related as

TE:		M=1	M=3	M=5	M=7	M=9
	$f_{R\mu}^{(M)}$	12.4305	38.3982	64.140	89.853	
	$f_{A\mu}^{(M)}$	12.8401	38.5371	64.2240	89.9144	115.601

1) Use $f_{A\mu}^{(M)}/M$ to find $f_{Ao}^{(1)}$:

12.8401	12.8457	12.8448	12.8449	12.8446

Average value in region of convergence = 12.8448

2) Use $f_{R\mu}^{(M)}$ to find $k_{eff}^{(M)} = \sqrt{X/\tan X}$ (Assuming $\mu_{eff} \approx 0$):

27.8%	28.0%	28.4%	28.5%

3) Use $f_{A\mu}^{(1)}$ to check μ_{eff}=1/X tanX:

0.04%

4) Check critical harmonic $M_c = \dfrac{2k_{eff}}{\pi\sqrt{\mu_{eff}}}$, M_c=10

LE:		M=1	M=3	M=5	M=7	M=9
	$f_{R\mu}^{(M)}$	12.8366	38.5219	64.2005	89.8813	115.558

1) Use $f_{R\mu}^{(M)}/M$ to find $f_{Ro}^{(1)}$:

12.8366	12.8406	12.8401	12.8402	12.8398

Average value in region of convergence = 12.8400

2) Use $f_{R\mu}^{(1)}$ to check μ_{eff}=1/X tan X

0.03%

Fig. 2. Sample data analysis for the $Li_2B_4O_7$ Z-cut longitudinal mode (frequencies in MHz).

$$\tan X = \frac{X}{-\underline{k}^2 + \mu X}, \quad X = \frac{\pi}{2} \cdot \frac{f_{A\mu}^{(M)}}{f_{Ro}^{(1)}} \tag{9}$$

where $\underline{k}$ represents the LE piezoelectric coupling.

A sample data analysis is given in Fig. 2. For the unelectroded plate, the mass loading μ is nominally zero, and the various harmonics of the TE antiresonance or LE resonance frequencies should be integer multiples of the fundamental frequency (1st harmonic, M=1). In practice, small deviations from harmonicity corresponding to a parasitic "effective" mass loading on the order of 0.04% have been observed. Such parasitic effects are readily identified through the recommended measurement of multiple harmonics. In measurements performed to date, the agreement between TE and LE measurements of a given mode and the agreement between repeated measurements using TE or LE have both been on the order of 10^{-4}.

Extraction of Material Constants

Resonator method measurements of either TE $f_{Ao}^{(1)}$ or LE $f_{Ro}^{(1)}$ yield essentially the same information as is obtained from corresponding measurements of plane wave velocities. The extraction of piezoelectric material constants from the measured data is thus accomplished in much the same way as for pulse-echo or high overtone thickness mode data.

In order to obtain piezoelectric material constants from the measured data, the stiffness eigenvalues are first extracted using the well-known relation for the zero-mass-loading, TE antiresonance frequency $f_{Ao}^{(M)}$ of the M^{th} harmonic of the infinite flat plate resonator,

$$f_{Ao}^{(M)} = M \cdot f_{Ao}^{(1)} = \frac{M}{2(2h)} \sqrt{\frac{\bar{c}}{\rho}} . \qquad (10)$$

In (10), M = 1,3,5,... denotes the harmonic number, 2h is the plate thickness, $\bar{c}$ is the piezoelectrically stiffened elastic stiffness, and ρ is the mass density. The zero-mass-loading LE resonance frequency $f_{Ro}^{(M)}$ is also given by (10). Note that the extraction of the stiffness eigenvalues requires not only the measurement of the critical frequencies, but knowledge of the plate thickness and mass density as well. While the critical frequencies can be measured quite accurately, it is difficult to determine the plate thickness and mass density to better than 0.01%. As a consequence, measurement of the critical frequencies to a higher degree of accuracy is of limited usefulness.

The exact procedure to be followed for extracting the desired material constants from the eigenvalue data depends upon the symmetry class of the material being evaluated. This topic has recently been discussed elsewhere [1,15].

Certain elastic stiffnesses may be determined directly from the eigenvalues associated with piezoelectrically unstiffened modes measured using LE. As an example, for symmetry class 4mm, the stiffness eigenvalue $\bar{c}^{(3)}$ of the Z-cut is given by

$$\bar{c}^{(3)} = c_{44}^{E} . \qquad (11)$$

For piezoelectrically stiffened modes, the elastic and piezoelectric parts must be separated. In some cases, the separations are readily determined from the differences between stiffened and unstiffened eigenvalues measured for different orientations. As an example, again for symmetry class 4mm, the stiffness eigenvalue $\bar{c}^{(3)}$ of the Y-cut is given by

$$\bar{c}^{(3)} = c_{44}^{E} + \frac{e_{15}^2}{\varepsilon_{11}^{S}} \qquad (12)$$

Thus, the piezoelectric stress constant e_{15} may be readily determined from the difference between $\bar{c}^{(3)}$ for the Y-cut and $\bar{c}^{(3)}$ for the Z-cut. In other cases, simple separations are not possible and the "intrinsic" constants must be determined as part of a least-squares fit to eigenvalue data from multiple orientations.

Conclusions

The non-uniform distribution of vibratory motion encountered in practical plate resonators cannot be neglected in resonator method measurements of instrinsic constants. An improvement to IEEE 176-1987 which addresses this effect has been developed.

References

[1] J. Kosinski, Y. Lu, and A. Ballato, "Pure-Mode Measurements of $Li_2B_4O_7$ Material Properties," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 41, no. 4, July 1994.

[2] ANSI/IEEE Standard 176-1987, "IEEE Standard on Piezoelectricity," published by the Institute of Electrical and Electronics Engineers, 345 East 47th Street, New York, NY 10017, January 1988, 54 pp.

[3] R. Bechmann, "Über Dickenschwingungen piezoelektrischer Kristallplatten," Archiv. der elekr. Übertragung, vol. 16, pp. 361-368, 1952.

[4] J. Kosinski, A. Ballato, I. Mateescu, and I. Mateescu, "Inclusion of Non-Uniform Distribution of Motion Effects in the Transmission-Line Analogs of the Piezoelectric Plate Resonator: Theory and Experiment," elsewhere in these Proceedings.

[5] H. F. Tiersten, "Analysis of Intermodulation in Thickness-Shear and Trapped Energy Resonators," Journal of the Acoustical Society of America, vol. 57, no. 3, pp. 667-681, March 1975.

[6] H. F. Tiersten, "Analysis of Trapped-Energy Resonators Operating in Overtones of Coupled Thickness Shear and Thickness Twist," Journal of the Acoustical Society of America, vol 59, no. 4, pp. 879-888, April 1976.

[7] H. F. Tiersten and R. C. Smythe, "An Analysis of Contoured Crystal Resonators Operating in Overtones of Coupled Thickness Shear and Thickness Twist," Journal of the Acoustical Society of America, vol. 65, no. 6, pp. 1455-1460, June 1979.

[8] D. S. Stevens and H. F. Tiersten, "An Analysis of Doubly-Rotated Quartz Resonators Utilizing Essentially Thickness Modes with Transverse Variation," Journal of the Acoustical Society of America, vol. 79, p. 1811, 1986.

[9] G. Sauerbrey, "Einfluss der Elektrodenmasse auf die Schwingungsfiguren dünner Schwingquarzplatten," Archiv. der elekr. Übertragung, vol. 18, pp. 617-624, 1964.

[10] G. Sauerbrey, "Amplitudenverteilung und elektrische Ersatzdaten von Schwingquarzplatten (AT-Schnitt)," Archiv. der elekr. Übertragung, vol. 18, pp. 624-628, 1964.

[11] G. Sauerbrey and G, Jung, "Schwingungsmoden plankonvexer Quarzplatten," Zeitschrift agnew. Phys., vol. 24, no. 2, pp. 100-108, 1968.

[12] G. K. Guttwein, T. J. Lukaszek, and A. Ballato, "Practical Consequences of Modal Parameter Control in Crystal Resonators," in Proceedings of the 21st Annual Frequency Control Symposium, April 1967, pp. 115-137.

[13] A. Ballato, E. R. Hatch, M. Mizan, B. H. T. Chai, R. S. Tilton, and T. J. Lukaszek, "Lateral-Field Excitation of Berlinite," in Proceedings of the 38th Annual Frequency Control Symposium, May-June 1984, pp. 206-224.

[14] A. Ballato, "Transmission-Line Analogs for Stacked Piezoelectric Crystal Devices," in Proceedings of the 26th Annual Frequency Control Symposium, June 1972, pp. 86-91.

[15] J. Kosinski, Y. Lu, and A. Ballato, "Pure-Mode Loci in Piezoelectric Plate Resonators: Application to Materials Evaluation in Class 4mm," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 40, no. 3, pp. 258-264, May 1993.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A NEW METHOD FOR THE DETERMINATION OF CONCENTRATIONS OF IMPURITIES IN QUARTZ CRYSTALS

Pierre Zecchini, Khalid Yamni, Bernard Viard and Daniel Dothée

Université de Franche-Comté, Faculté des Sciences et des Techniques
Laboratoire de Cristallographie et Chimie Minérale, 25030 Besançon cedex, France

Summary

Analytical methods, principal component analysis and multivariate regression, have been applied to the spectroscopic infrared data and chemical analyses of quartz crystals in order to find the variations, co-variations and internal relationships between data and samples. These analytical methods generate calibration models which can be used for classification and prediction. So differences and likenesses between samples can be well detected. An attempt has been made to get the chemical analysis of crystals by using only a room temperature infrared spectrum. The absorption in the infrared range was measured between 3300 to 3600 cm^{-1} and the chemical analyses performed by ICP.

Key words: quartz, infrared spectroscopy, principal component analysis, multivariate regression.

Introduction

It is well known that impurities in quartz crystals induce alterations of the physical properties of the material and more particularly, the absorption is changed in the infrared radiation range [1]. All the absorption bands observed over the entire transmission range, between 50000 to 2000 cm^{-1}, are generally used for identification of impurities and rarely for quantitative measurements. A few of them are used to determine the quality of the material [2].

Different workers have developed non-destructive quantitative measurements such ESR [3] and He-Ne laser absorption [4,5] in order to determine the aluminum content and its partition within crystals, or infrared spectroscopy in order to calculate the water content [6]. With these techniques, the amount of the other impurities, such as alkali, iron...,could not be known and, to get them for all the impurities, a chemical analysis of the material is necessary. Until now, the chemical analyses of quartz crystals have been carried out by destructive methods, i.e. inductively coupled plasma (ICP) or atomic absorption.

For each of these techniques, specific samples, in size, orientation or thickness, are necessary and, moreover, the measurements are quite never performed on the same part of a bar. So, for industrial applications, the quality of each bar used and, more particularly the quality of the useful part of the bar, cannot be quickly checked by these methods. In order to overcome this problem, the application of a non-destructive technique has been studied and applied to the quartz material. This method which is in fact a statistical method, can take into account all together the modifications detected over a large range of the infrared spectra of a quartz crystal. It needs only an initial calibration, using one set of data given by an easy to perform technique i.e. room temperature infrared spectroscopy and another set of data which is supposed to be directly correlated to the quality of the bar, i.e. chemical analyses. When the calibration is achieved, a non destructive mean will be proposed to check the quality of bars, by using only the data given by infrared spectra.

Experimental procedures

Infrared transmission measurements were made between 3300 and 3600 cm^{-1} using a single beam Nicolet Magna 750 FTIR spectrometer. In this study, 52 polished crystals were measured at room temperature, normal to the Z-direction. Some of them were also measured normal to the Y-direction. The usual beam size was 5 mm wide. The samples used for this investigation were generally Y-cuts, between 0.5 and 1.9 cm thick.

The chemical analyses were made using a Perkin-Elmer Inductively Coupled Plasma 6500. The impurities were measured out after dissolution of 4 to 8 grams of quartz in a hydrofluoric solution. To use more than one emission radiation for each element, when that is possible, leads to a better reliability of the measurements. The low contents of impurities needed a particularly procedure for the dissolution of the samples:

- all the solutions used for cleaning (acetone, hydrochloric, nitric and hydrofluoric acids) were high-pure solutions. The water used to rinse was high-pure too.

0-7803-1945-1/94 $3.00 © 1994 IEEE

- the samples were broken by a thermal shock in high-pure water after heating 10 minutes at 1000 °C. The analysed quartz material was weighed after a 15 hour heating at 120 °C in a Teflon beaker. Then, the hydrofluoric dissolution was made in that beaker. It needed at least 2 or 3 days at 140°C.

- after evaporation, a 2% ultrapur nitric acid solution was used to prepare the solution to be analysed. 2 ppm of scandium were added to the solution to be used as an internal standard.

The analytical methods

The study of the data from infrared spectroscopy and from chemical analyses was performed by using a commercial software, "The Unscrambler" [7]. It runs in good conditions with a computer 486 DX 16 MB. This software can be used to solve different problems of identification and analyses [8]. The first step is the normalisation of the infrared data in order to become free from the thickness and defects on the surface of the samples. Then, absorbance or derivation can be used.

Four fundamental parts must be differentiated in the possibilities of this software:

- in one set of data, i.e. absorbance at different wavelengths for various samples, the data decomposition is carried out by PCA (the principal component analysis) in order to find which are the most important wavelengths for the total variation in all the data set. This information leads to make the model called the PCA-model.

- to establish regression relationships between sets of data, i.e. infrared absorbance and the chemical analyses, two well known data sets must be first used. These training sets lead to create a new model which will be the calibration model. This is carried out by PCR (the Principal Component Regression) or by PLS (the Partial Least Squares regressions). PLS-1 handles one variable at time, i.e. on the one hand, Al content or Na content, and on the other hand, the absorbance data. It must be distinguished from PLS-2 which, simultaneously, handles several variables, i.e., on the one hand, Al- and alkali-contents and on the other hand infrared data.

- the third part is the classification of unknown samples which can be achieved by using the PCA or PCR models.

- the last part is the prediction of the chemical analysis. By using the calibration models (PCR or PLS-1), the amount of impurities can be estimate in unknown samples from the infrared spectra.

PCA, PCR and PLS are projection methods. The many dimensional co-ordinate system made using the data sets are reduced in two- or three-dimensional representations where the samples can be localised. The "scores" are the co-ordinates of the samples in the new axes, the principal components. The estimation of the chemical contents is given numerically or graphically with a standard deviation.

Experimental results

Table I gives the known or measured characteristics of the samples used. A comparison between their commercial quality was made by the calculation of the value of

$$\alpha = \frac{1}{\text{thickness}_{cm}} \log_{10}\left(\frac{T_{3800}}{T_{3500}}\right),$$

even if the transmission T_{3800} at 3800 cm^{-1} and T_{3500} at 3500 cm^{-1} were not exactly measured with the standard conditions which, until now, need a double beam spectrometer.

Table 1: The characteristics of the samples

sample	α	origin	sample	α	origin
1	0.045	syn	27	0.032	B
2	0.038	syn	28	0.029	B
3	0.030	syn	29	0.028	A vs
4	0.036	syn	30	0.030	C
5	0.059	syn	31	0.029	A
6	0.040	syn	32	0.046	B s
7	0.032	syn	33	0.039	C s
8	0.037	syn	34	0.033	B(+X)
9	0.039	syn	35	0.187	B (-X)
10	0.041	syn	36	0.028	natural
11	0.038	syn	37	0.018	D
12	0.041	syn	38	0.022	D
13	0.040	syn	39	0.022	D
14	0.029	syn	40	0.023	D
15	0.029	A	41	0.025	D
16	0.024	A	42	0.026	D
17	0.022	A	43	0.029	D
18	0.022	A	44	0.027	D1
19	0.028	A	45	0.028	D1(Z)
20	0.024	A vs	46	0.032	D2
21	0.030	A vs	47	0.032	D2 (Z)
22	0.034	A vs	48	0.019	D3
23	0.023	A vs	49	0.020	D3(Z)
24	0.027	A vs	50	0.022	D3
25	0.029	B (Z1)	51	0.022	D3
26	0.031	B (Z2)	52	0.017	D3

A, B, C are different quartz manufacturers
Z1, Z2, +X, -X are the growth zones of the same Y-cut
D is a Premium Q run, (Z): Z-dir. observation
D1: Na-doped, D2: K-doped, D3 Li-doped
s, vs: sold as swept or vacuum swept material
syn: synthetic quartz sold before 1987.

Part 1: PCA data decomposition

1 - results

Several different steps are shown on figures 1 to 8. The reduction of the data - the 311 absorption values measured for each of the 52 samples, between 3600 and 3300 cm^{-1}, data space 0,964 cm^{-1}, - gives the "scores" in the new axes system. Figure 1 is a three-dimensional projection. As the axes PC #1, PC #2 and PC #3 which explain respectively 94, 5 and 1 % of the full information, this projection explains all the relative space distribution of the samples. As the axes 1 and 2 explain 99 %, the two-dimensional projection (PC #1,PC #2), figure 2, shows that samples 35 and 36 are "out of range" and are located far from all the other. These "outliers" are a natural crystal (sample 36) and a -X zone of a synthetic crystal (sample 35). All the other synthetic crystals are located in the same part of the graph.

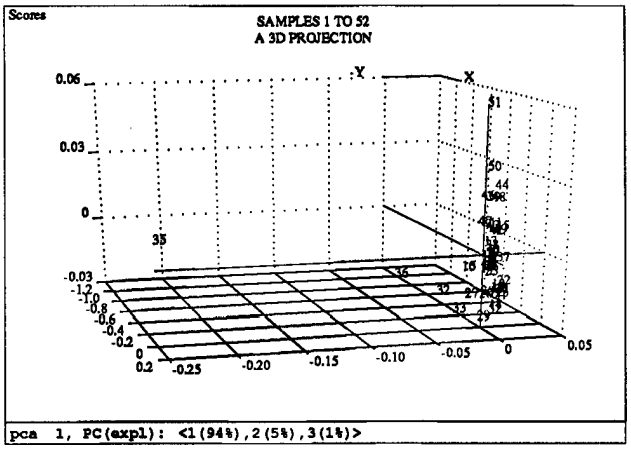


Figure 1: Each sample gets its new co-ordinates in the set of axes PCA #1, PCA #2 and PCA #3.

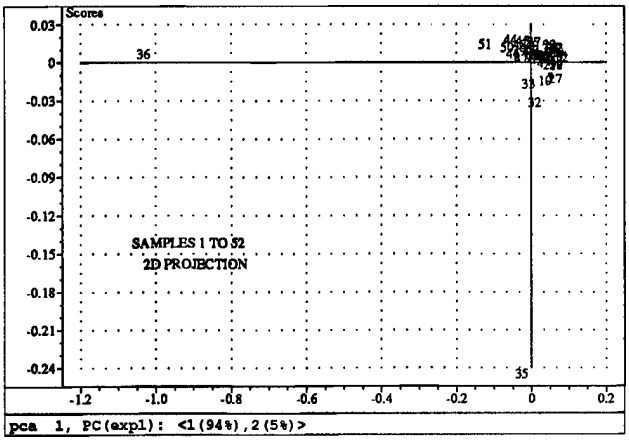


Figure 2: On this two-dimensional projection, the samples 35 (-X zone) and 36 (natural crystal) are well differentiated from the other synthetic quartz crystals.

On figures 3 and 4, the outliers are removed. The new calibration gives a new distribution of the samples. The (1,2) projection which explains 98 % and the (1,3) projection which explains 95 %, show several different information which can be correlated to the samples. Quite all of the samples from the Premium-Q run (samples 37 to 52) are located in the same part of this projection, except sample 52. They are differentiated from the other commercial runs. For this particular run, all the treatments have an effect on the position of the representative point:

- the as-grown crystals, samples 37 to 43, remain very close, except the sample 39
- the Na-doped samples 44 and 45, the K-doped samples 46 and 47 and the Li-doped samples 48 to 51 can be located in different parts of the projections
- the highest and the lowest α -values (respectively for sample 32 and sample 52) are located in two opposite parts on the (1,2) projection plane.

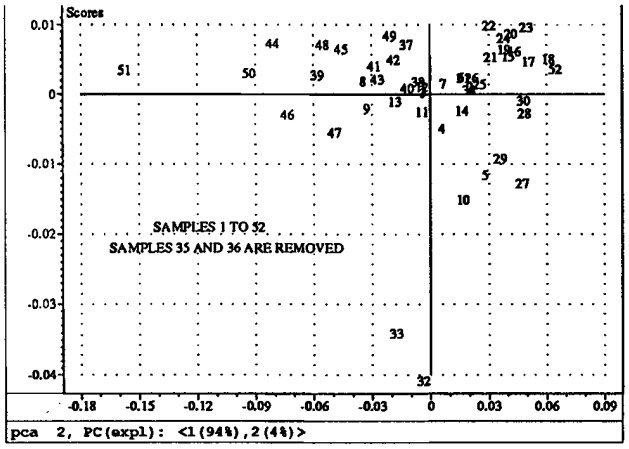


Figure 3: As the samples 35 and 36 are removed, a better projection can be given for the remaining samples.

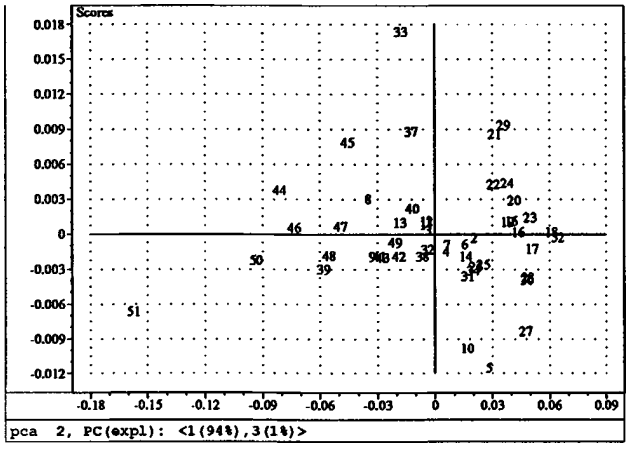


Figure 4: This other 2D-projection gives complementary information for the remaining samples.

At least, both these projections are needed to observe variation and co-variation for all the 52 samples.

On figure 5, only the samples 1 to 34 are studied. The new calibration indicates that the older produced samples (1 to 14) are quite all in the upper left part of the two-dimensional (1,2) representation. These samples were cut from marketed bar grown before 1987. All the other, samples 15 to 34, were sold in 1992 or in 1993.

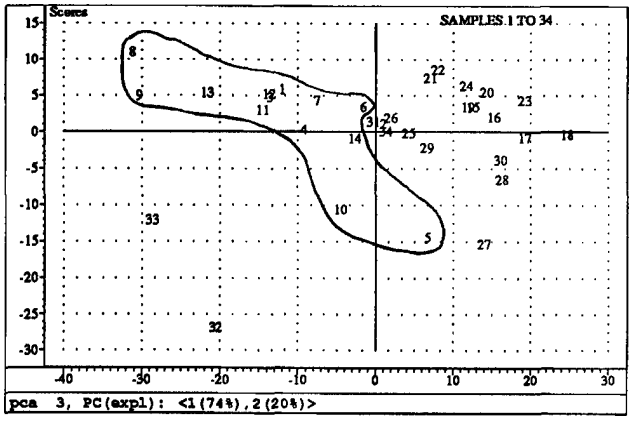


Figure 5: Dispatching and gathering according to the known information or particular positions.

Figures 6 to 8 are one-dimensional representations. These projections, respectively on the axis 1 (explaining 74 %), 2 (explaining 20%) or 3 (explaining only 4 %), show this difference - the co-ordinate values are negative or positive mainly depending on their "period of production" and their hydrogen content, figure 6 - or the particularities of some samples (5, 10, 27, 32, 33), figure 7, or a difference between samples which looked similar on figures 6 and 7 like the samples 5 and 10 or 32 and 33, figure 8.

2- discussion

The PCA decomposition and the knowledge about the different samples cannot be easily correlated by using only one of the specific characteristics such as production, treatment, or commercial quality value. However, it is able to well indicate on the one hand, some waited likenesses and differences, and on the other hand, some unexpected points. The PCA analysis takes into account all together the known and unknown characteristics of the samples which have an effect on the infrared absorption range.

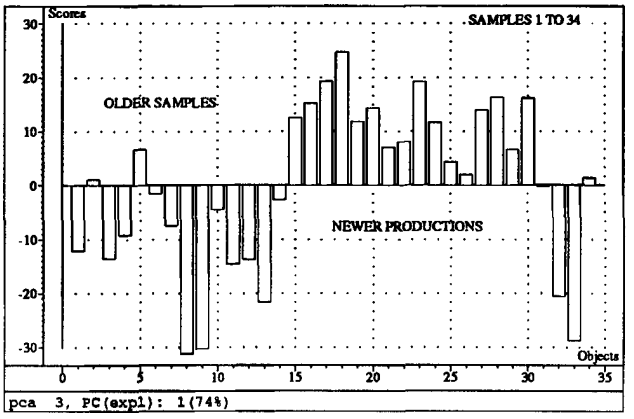


Figure 6: This one-dimensional projection on PC #1 shows the maximal difference between the samples.

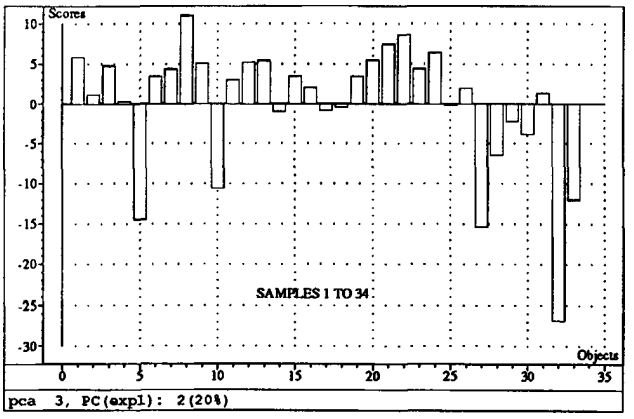


Figure 7: Explanation given on PC #2, (20 %).

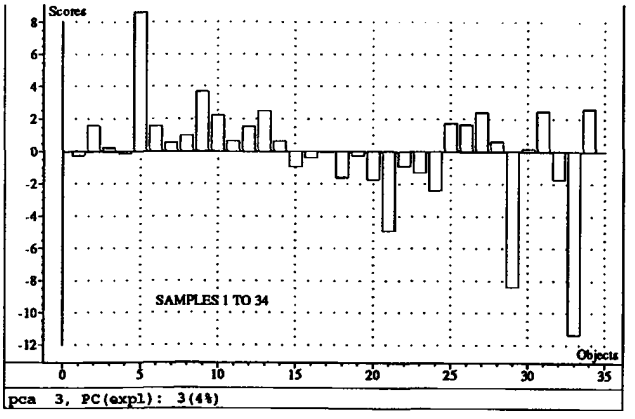


Figure 8: The 4% of explanation on PC #3 gives nevertheless complementary information.

Part 2: multivariate calibration

1 - the calibration model

The calibration model is made by using the infrared data and the chemical analyses of the samples 15 to 34. The chemical analyses of these samples are given in table 2.

1 -1: Principal Component Regression

The PCR analysis is resumed by the figures 9 to 12.

table 2: ICP chemical analyses (samples 15 to 34)

sample #	K	Li	Na	Al
15	0.09	0.39	0.66	1.62
16	0.00	0.02	0.21	0.70
17	0.11	0.02	0.23	0.36
18	0.20	0.02	0.53	0.45
19	0.38	0.01	0.33	0.74
20	0.95	0.01	2.71	0.95
21	1.32	0.02	2.68	1.84
22	0.44	0.01	1.06	0.79
23	0.62	0.00	0.51	0.67
24	0.38	0.01	0.15	0.53
25	0.00	0.08	0.00	1.57
26	0.00	0.08	0.00	1.93
27	0.41	0.16	0.79	0.96
28	0.40	0.19	0.77	2.70
29	0.12	0.03	0.51	2.72
30	0.30	0.38	0.69	2.91
31	0.47	0.30	0.44	1.80
32	0.25	0.59	0.23	1.35
33	0.39	2.01	0.51	1.93
34	0.00	0.61	3.36	2.92

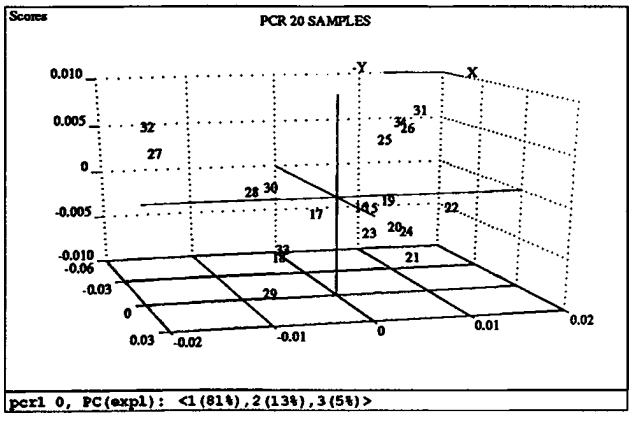


Figure 9 : A three -dimensional representation using the principal components #1, #2 and #3

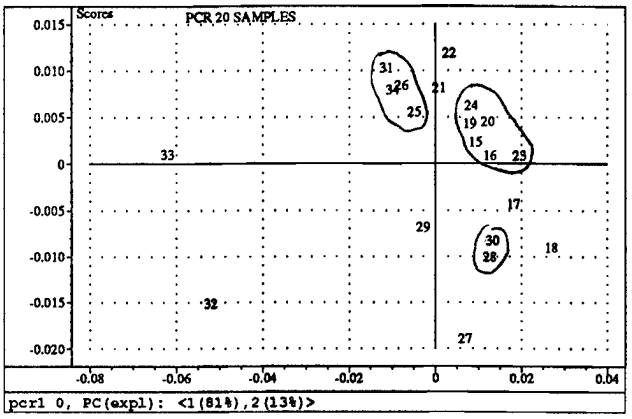


Figure 10 : Gathering of similar samples

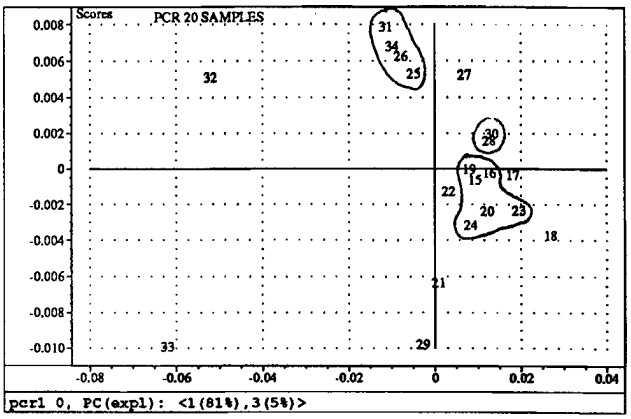


Figure 11 : Similar samples remain together.

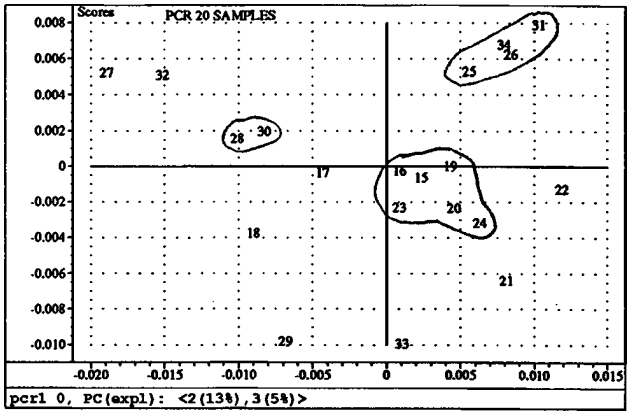


Figure 12: This projection is confirming the similar samples

As indicated on these previous figures, samples can be gathered together. In fact, similar samples remain together whatever the modes of projection are. Some correlations can be made between analyses and roundup:

- samples 28 and 30 are similar,
- samples 25, 26, 31, 34 have a high Al- content,
- samples 15 to 23 are from the same manufacturer. On figure 10, sample 21 which has a high Al-content is close to samples 25, 26, 31 and 34. It is far from these samples on figures 10 and 11. That can be due to the difference in Na-content. The cleanest samples 17 and 18 are on each projection at the opposite side of all these samples.
- sample 33 remains always alone. Its Li-content is the highest measured by ICP in these 20 samples. Moreover, its aluminum concentration and its α-value are high too.

1 - 2: The PLS calibration

A PLS-1 calibration model is created for K, Li, Na and Al using the infrared data and the chemical analyses given in table 2.

2 - Applications of the created model

Five unknown samples (noted A, B, C, D and E) were measured using only their infrared spectra and the calibration model. The α–values of these samples are reported in table 3.

table 3: α–values of the 5 unknown samples

A	B	C	D	E
0.0260	0.0365	0.0266	0.0328	0.0303

2 - 1: Classification

From the infrared spectra, the classification of the new samples can be made by using the PCA or PCR model.

Figures 13 (a, b, c) show the position of each new sample in comparison with the samples of the training set. Sample A is equivalent to the samples 28 and 30, samples C and E move together with the samples 15, 16 and 19 in each projection. Samples B and D look like similar to sample 27 on figure 13(b) but a large difference is obtained on figure 13(a) or on figure 13(c).

In comparison with the values given in table 3, these observations indicates that a classification using only the commercial α-values is not valid.

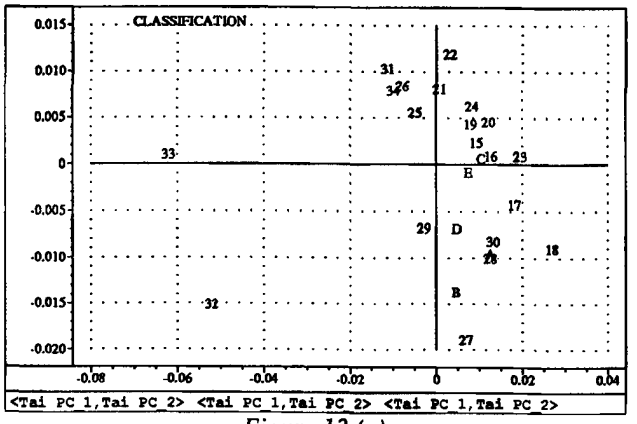


Figure 13 (a)

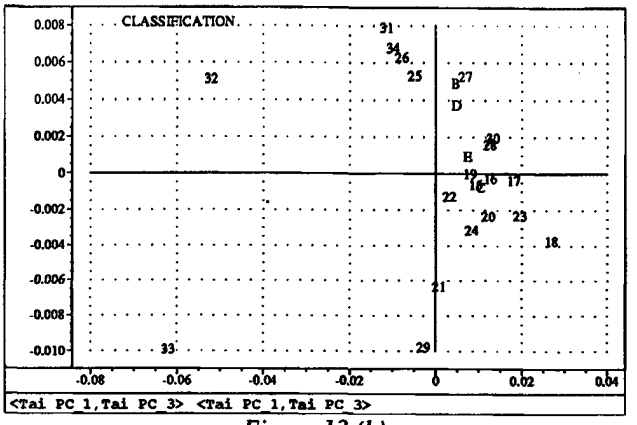


Figure 13 (b)

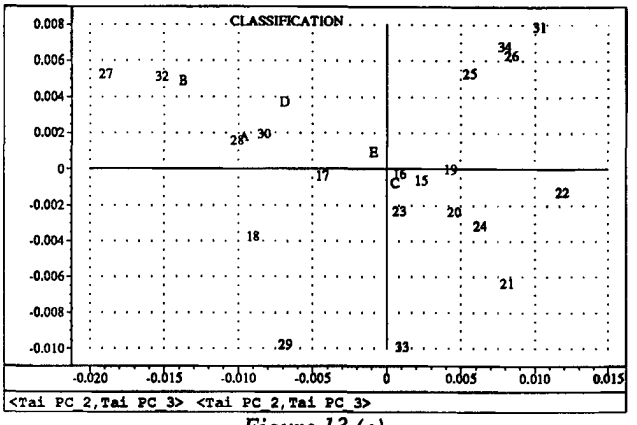
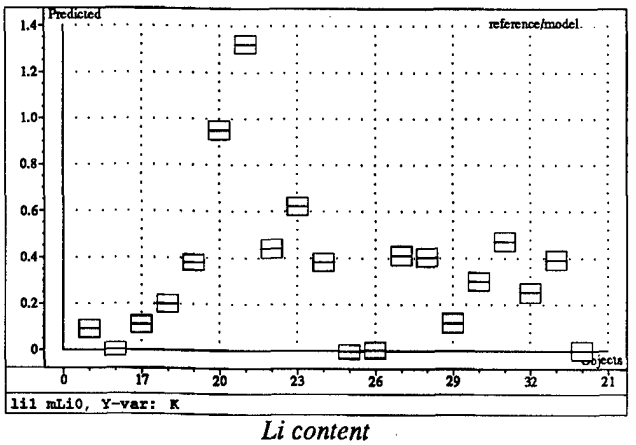


Figure 13 (c)

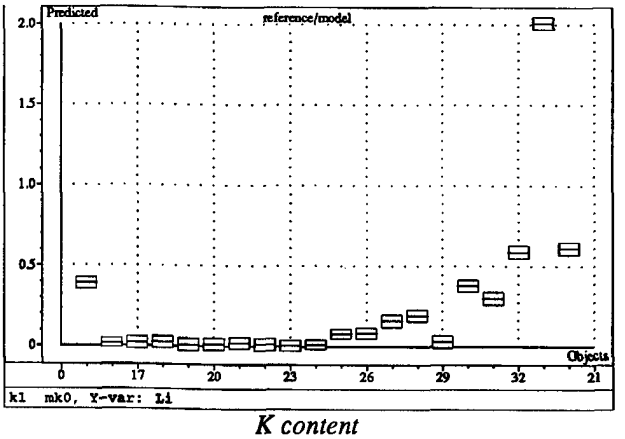
Figure 13: Dispatching of the new samples within the model.

2-2: Prediction of the impurity contents

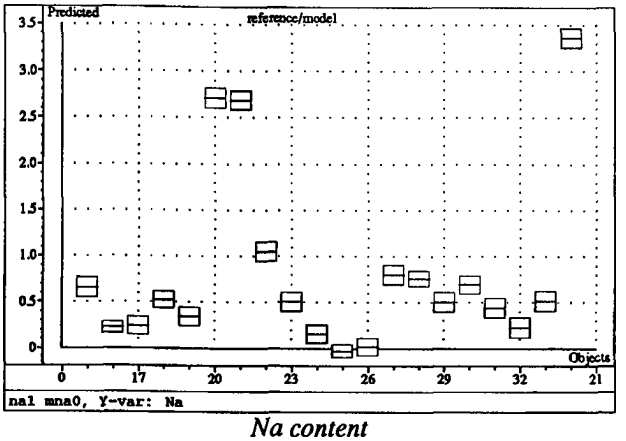
Each PLS-1 calibration model can be used with the infrared data of the training set. The graphical results are given on the figures 14 and 15 for the samples 15 to 35.



Li content



K content



Na content

Figure 14 : The re-calculated alkali contents for the training set

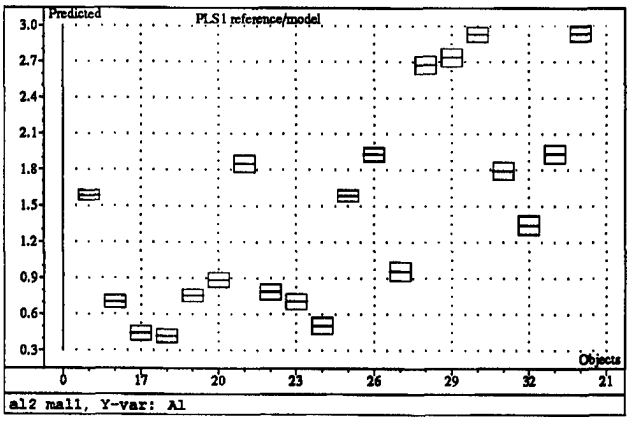


Figure 15: The re-calculated Al content for the training set

The infrared spectrum of the sample 21 from the training set was added to the infrared spectra of the unknown samples. The PLS-1 calibration carried out using or the absorbance or the first derivative, permits now a prediction of the impurity concentration of the 6 samples using only the infrared data.

In table 4a, the estimate impurity contents and those measured by ICP are reported. Using or the absorbance (Abs) or the first derivative (D1), the concentrations predicted for aluminum remain close. For 2 samples, a good accuracy is obtained between PLS and ICP. For sodium, the estimation is good using the absorbance, but worse using the first derivative. For potassium and lithium, the values are too low to make a significant comparison. The standard deviation of each method is compared in table 4b.

table 4 a) chemical analyses

		A	B	C	D	E	21
	ICP	3.13	2.46	2.28	1.05	2.77	1.84
Al	Abs	1.73	2.41	0.89	1.27	1.42	1.83
	D1	1.05	2.81	0.48	1.13	1.26	1.85
	ICP	0.16	0.27	0.23	0.39	0.35	0.02
Li	Abs	0.46	0.31	0.32	0.14	0.15	0.02
	D1	0.15	0.51	0.11	0.02	0.05	0.01
	ICP	0.14	1.22	0.84	0.73	0.75	2.68
Na	Abs	0.11	1.24	0.98	0.47	0.97	2.68
	D1	0.60	3.45	0.52	1.73	1.32	2.68
	ICP	0.25	0.39	0.01	0.14	0.02	1.32
K	Abs	0	0	0	0	0	1.32
	D1	0	0	0.19	0	0.22	1.32

table 4 b) standard deviation

		A	B	C	D	E	21
	ICP	0.68	0.28	0.55	0.68	0.27	
Al	Abs	0.13	0.20	0.04	0.12	0.05	0.05
	D1	0.08	0.15	0.04	0.09	0.04	0.07
	ICP	0.01	0.01	0.03	0.01	0.01	
Li	Abs	0.09	0.13	0.03	0.09	0.03	0.03
	D1	0.02	0.04	0.01	0.02	0.01	0.01
	ICP	0.01	0.10	0.02	0.05	0.01	
Na	Abs	0.23	0.38	0.08	0.24	0.09	0.10
	D1	0.05	0.07	0.03	0.06	0.03	0.05
	ICP	0.03	0.04	0.01	0.01	0.02	
K	Abs	0.07	0.13	0.03	0.08	0.03	0.04
	D1	0.02	0.03	0.01	0.02	0.01	0.02

3 - discussion

Using the PCR model, the classification for the new samples shows differences and likenesses between each of them. Never the α-values nor the chemical analysis results are sufficient for the full explanation of their relative positions. The model permits to find with a good accuracy the impurity concentrations in the sample 21 from the training set even if this sample is introduced as an unknown sample. However, for the other unknown samples, a larger difference between the predicted impurity contents and those determined by the ICP chemical analysis is noted.

The model used was built with 20 reference samples. In this training set of quartz crystals, the variation of each impurity content is not wide enough to generate all the possibilities of their effects on the infrared spectra. The calibration model must be developed not only by increasing the number of the samples but by adding "better" samples. That will increase the validity of the results.

The hypothesis that each impurity has not only a specific effect on the infrared spectra, but that it induces consequences according to the other impurities and their relative concentration, can be made. In that case, the classification using a PCR or the PCA model seems to be the best way to qualify crystals than any of each qualification test used separately .

Conclusion

In this study, a method for a non destructive qualification of quartz crystal, using the principal component analysis and multivariate calibrations has been applied.

This method leads by using a training set of well known samples to models.

The models can be used as a clean way for environment for the prediction of the impurities concentrations without a chemical analysis or for the classification of new samples.

To reach the impurity contents with a good accuracy, the calibration model must be developed by increasing the number of "selected" samples having a larger variation of impurity contents.

The qualification of quartz crystals can be carried out by using more than few wavelengths of the infrared spectrum on which the effect and the concentration of all the impurities could act and interact. It is possible with this statistical method to "read" the infrared spectra of new samples and to classify them in comparison with reference crystals.

References

1 - A. Katz, "Hydrogen in alpha-quartz", Philips Res. Reports , vol. 17, pp. 133-195 and 201-279, 1962

2 - J.C. Brice and A.M. Cole, "The characterization of synthetic quartz by using infrared absorption3, Proc. of the 32nd Annual Frequency Control Symposium, pp.1-10, 1978

3 - M. E. Markes and L.E. Halliburton, "Defects in synthetic quartz: Radiation-induced mobility of interstitial ions", J. Appl. Phys., 50 (12), 1979

4 - H.G. Lipson, "Infrared and laser spectroscopic characterization of aluminum defects in cultured quartz", Proc. of the 40th Annual Frequency Control Symposium, pp. 63-69, 1986

5 - P. Zecchini and H.G. Lipson, "Defect distribution mapping in quartz", Proc. of the 43rd Annual Frequency Control Symposium, pp. 490-496, 1989

6 - M.S. Paterson, "The determination of hydroxyl by infrared absorption in quartz, silicates glasses and similar materials", Bull Min, 105, pp 20-29, 1982

7 - "The Unscrambler" from CAMO (Computer-Aided Modelling A/S), Trondheim, Norway

8 - H. Martens & T. Næs, "Multivariate Calibration", Wiley & Sons Ltd, ISBN 0-471-93047-4, 1993

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

GROWTH OF HIGH QUALITY QUARTZ CRYSTAL AND ITS APPLICATION TO TEMPERATURE SENSORS

S. Ishigami and M. Sato, Tokyo Denpa Company Ltd.
6-11, Chuo 5 Chome Ohta-Ku, Tokyo, Japan

F. Uchiyama, K. Agatsuma and K. Tsukamoto
Electrotechnical LaboratoryUmezono 1, Tsukuba, Japan

ABSTRACT

Synthetic quartz crystal has been grown in a platinum lined autoclave measuring ϕ 20 cm $\times$ 300 cm. The seed, measured purity of the liquid solution and the nutrient are carefully selected to minimize impurities and dislocations in the "as grown" quartz crystal. The three-layer autoclave has been newly designed to grow a highly pure and inclusion free synthetic quartz crystal. The liner and the fixtures such as basket, frame and baffle plate for the autoclave are all made from platinum. Etch channels, etch pits and inclusions are not detectable in this synthetic quartz crystal using a detection limit of 10 μm.

A wide-range cryogenic temperature sensor is fabricated from this highly pure synthetic quartz crystal. An optimum Y cut plate is mounted in a hermetically sealed enclosure 2mm in diameter and 6mm in length to realize a cryogenic temperature sensor covering a wide temperature range from 4.2K to 400K. The sensor operates in fundamental mode at about 28MHz through either an extended cable connection or wireless echo sensor system. Although degradation in crystal impedance and an abnormal frequency change of the temperature sensor was initially found at about 60K, these disadvantages have been eliminated by applying a "sweeping method".

INTRODUCTION

In 1990, Tokyo Denpa Co., Ltd. started a contract with Science and Technology Agency of Japan. The subject of the contract was "Wireless Cryogenic Thermometer and Application to Cryogenic Temperature Control System", and it consisted of three parts, the growth of high quality quartz crystal, the development of the temperature sensor and the temperature control system for cryogenic temperature. The investigation was continued for two and half years for these three themes. In this paper, the results of crystal growth and the performance of the temperature sensor are described.

Growth of High Quality Quartz Crystal

Many researchers have been studying on purifying synthetic quartz crystals to an ideal level by using superior seed, solution liquid and lascas, and by selecting better growth conditions, but nowadays more superior performance for synthetic quartz crystal is required for the sophisticated application such as temperature sensing system, and high frequency fundamental quartz crystal products.

The researchers were concerned about contaminations and impurities that come from the inner wall of the autoclave or from the nutrient, lascas. To minimize these unnecessary objects, various researches have been done for a long time. An application of a liner made from the precious metal like gold or silver is one of the efforts for this purpose.

In 1988, Rovert Irvine et al.[1] of Sawyer Research Products, Inc. presented a study result on growing a synthetic quartz crystal using a silver lined autoclave. In this study, they used a sizable silver liner and acquired useful data for successive researchers.

We started a study for an ideal autoclave structure and soon understood that a liner was effective to minimize the defects in a synthetic quartz crystal. Then we decided to employ platinum as a liner material, because gold is too soft for a large vessel and silver tends to scatter in an autoclave.

The investigated items are as follows.

- Design for autoclave
- Design for platinum liner
- Growth of crystal without liner
- Growth of crystal with platinum liner
- Sweeping method

0-7803-1945-1/94 $3.00 © 1994 IEEE

Cryogenic Temperature Sensor

It is well known that a certain quartz crystal unit is suitable for a temperature sensor because it has a frequency output that can be easily converted to a digital signal.LC-cut [2], AC-cut and Y-cut have been used for temperature sensing, but the operating temperature range is comparatively narrow, 77K–523K. The early works on a quartz cryogenic temperature sensor were done by Flynn et al.[3] and Smagin et al.[4]. Flynn reported that the frequency of the quartz resonator decays almost linearly with reducing temperature but had a kink near 60K.

We cut an optimum Y cut plate, which has a maximum temperature coefficient, from the high quality synthetic quartz crystal and used it as a cryogenic temperature sensor because it could show the satisfactory sensitivity at lower temperature. Since a smaller size is desirable for the temperature sensor, we investigated a rectangular quartz crystal plate, blank, and mounted it in a small cylindrical enclosure measuring $\phi 2 \times 6$mm.

The investigated items are as follows.

Improvement in sensitivity
Reducing enclosure size
Highly reliable enclosure
Approximation for accurate measurement

AUTOCLAVE DESIGN

In order to achieve an ideal autoclave, we discussed the following points as a basic concept.

Sizable autoclave to acquire effective data for mass production (height: 3 m, inner diameter: 20 cm)

Use of a platinum liner, vessel, to minimize inclusions and impurities

Pressure-balanced platinum liner, vessel, to avoid contaminations

Platinum made hardware such as basket, baffle plate, frame and suspension

Absolute and stable temperature difference between dissolution and growth zone

At the first stage of this study, we planned to make a true platinum liner on an inner wall of the autoclave by plating method or clad metal method or by plating with explosion. But our research results did not indicate that these methods were realistic, because the size of the autoclave was too large for these methods. Therefore, we decided to use a platinum vessel that had thin practical wall to reduce the cost.

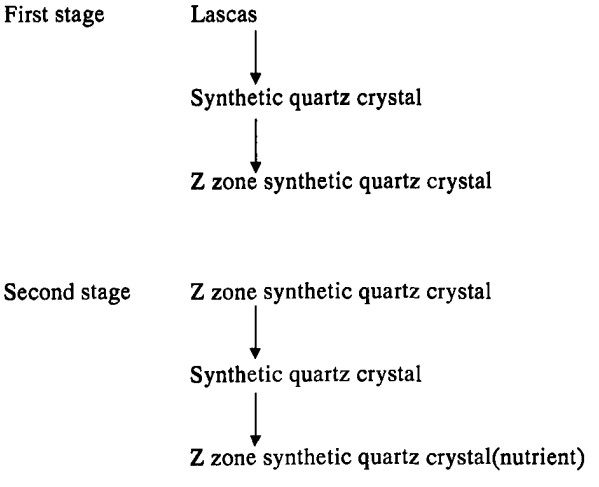
Figure 1 illustrates the final structure of the autoclave. The autoclave consists of an outer vessel made from an iron alloy, iron-chromium- vanadium, and an inner platinum vessel and a frame, made from the alloy, to protect the platinum vessel. As can be seen in the figure, dissolution and growth zone are insulated from the outer vessel with the platinum liner. Sub baffle plates are attached to the iron frame to prevent convection in the gap between the platinum liner and the outer vessel. At the beginning of this program, some innovative structure was applied at the top of the platinum vessel to keep the purity in the platinum vessel with keeping the pressure balance, but later on, the structure was removed because a more efficient method was devised.

To keep accurate size and lower cost for the platinum vessel, it was fabricated with four thin platinum plates 0.3 mm thick. The plates were formed and butt-welded to form the vessel and was finished on a lathe.

Figure 2 shows photographs of the autoclave and the platinum vessel with the iron frame.

GROWTH OF QUARTZ CRYSTAL

After forming a protective layer on the inner wall of the outer vessel, a series of experimental runs were conducted to find optimum run conditions. The quartz crystals were grown three times in the outer vessel without the platinum vessel to determine the optimum run conditions. To obtain a superior performance, the nutrient was carefully purified as follows.



Processing the lascas in this method, the impurities were reduced drastically. Table 1 shows the impurity levels.

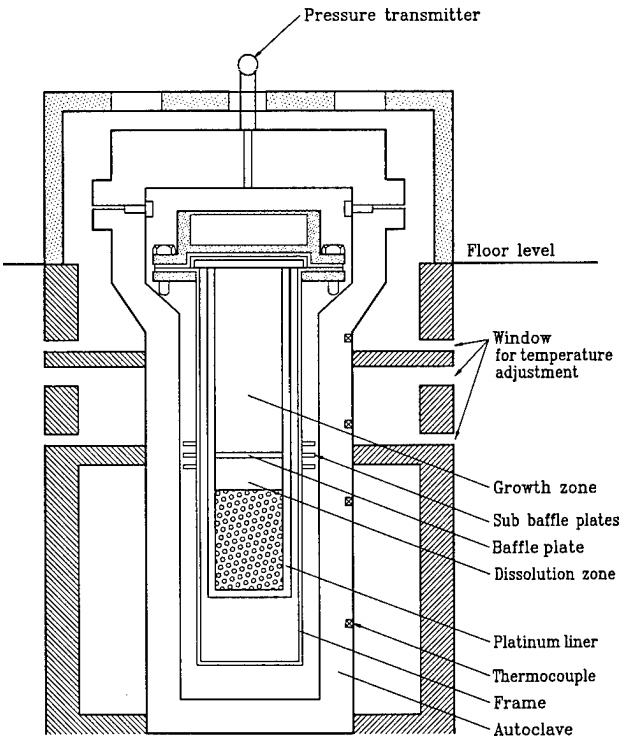


Figure 1. Final Structure of the Autoclave

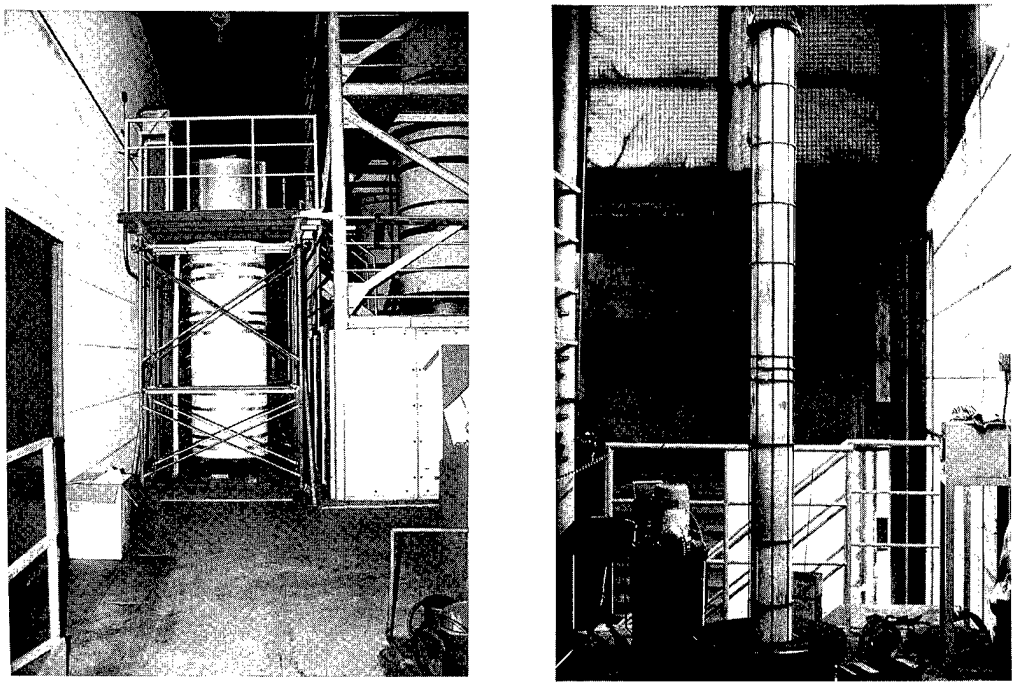


Figure 2. Photographs of the Autoclave and the Platinum Vessel with Iron Frame

Table 1. Impurity Level of raw material unit: ppm

	Lascas	Z zone of synthetic quartz crystal	Z zone crystal for nutrient
Aluminum	52.2	9.4	1.2
Iron	41	2.9	1.0

Table 2 presents the run conditions for the growth without platinum vessel.

Table 2. Run Conditions (without platinum vessel)

Items (Unit)	Run without Platinum Vessel
Raw Material	synthetic quartz
Weight of Raw Material (kg)	38.9
Quantity of Seed (pcs)	41
Surface Area of Seeds ($x10^3cm^2$)	4.403
Solution	Sodium Hydroxide
Solution Concentration (%)	4
Temperature at Growth Zone (degree centigrade)	333.7
Temperature at Dissolution Zone (degree centigrade)	359.5
Temperature Difference (degree centigrade)	25.7
Pressure (kg/cm^2)	1543
Run Days (day)	51

Opening the autoclave after 51-day-run, comparatively large 41 pieces of synthetic quartz crystals were obtained. The performance of the crystals were checked in detail and is shown in Table 3.

Table 3. Performance of crystal grown without platinum vessel

Item	Performance
Growth Rate (mm/day) Z axis	0.419
Growth Rate (mm/day) X axis	0.256
Q ($x 10^6$)	2.67
Etch Channel Density (pcs/cm^3)	3.12
Inclusion Density (pcs/cm^3)	0.23
Impurity Level (ppm)	
Aluminum	2.4
Sodium	0.3
Iron	1.0

As can be seen in the table, the crystals show superior performance and can be classified as an optical grade.

GROWTH WITH PLATINUM VESSEL

Acquiring the fundamental data for run conditions, the platinum vessel was installed in the autoclave and we started to grow high quality quartz crystal using the platinum vessel.

Prior to an actual crystal growth, we had to consider a method to keep the pressure balance to maintain the cleanliness of the inside of the platinum vessel. At the beginning, we attached a mechanism that consisted of distilled water and siphon structure. However, it was so delicate to handle that we soon removed it and studied a new method. The new method we finally devised was to adjust degree of fills for the platinum vessel and for the space between the outer vessel and the platinum vessel. Several experiments were conducted to confirm the efficiency and eventually, the optimum ratio of fill was determined.

Following the experimental runs, two actual growth runs were done in the platinum vessel. The growth conditions and the results are shown in Table 4 and 5, respectively.

Table 4. Growth Conditions

Items (Unit)	First Run	Second Run
Raw Material	sysnthetic quartz	synthetic quartz
Weight of Raw Material (kg)	15.2	16.0
Surface Area of Raw Material ($x10^3cm^2$)	1.95	NA
Weight of Seeds (kg)	0.64	0.88
Quantity of Seed (pcs)	31	33
Surface Area of Seeds ($x10^3cm^2$)	1.92	2.13
Solution	Sodium Hydroxide	Sodium Hydroxide
Solution Concentration (%)	4	4
Temperature at Growth Zone (degree centigrade)	314.0	324.3
Temperature at Dissolution Zone (degree centigrade)	369.8	379.7
Temperature Difference (degree centigrade)	55.8	55.4
Pressure (kg/cm^2)	1476.8	1590
Supplied Electric Power (kW)	11563.0	14435.5
Run Days (day)	50	59
Growth Rate (mm/day)	0.361	0.428

Table 5. Results of Growth

	Mass Production	Autoclave without platinum liner	Autoclave with Platinum Liner (1)	Autoclave with Platinum Liner (2)
Etch Channel Density (pcs/cm^2)	30	3.12	0.1	0 (ND)
Etch Pit Density (pcs/cm^2)	70	6.2	1.8	0 (ND)
Inclusion Density (pcs/cm^3)	4.84	0.23	0.16	0 (ND)
Q ($\times 10^6$)	2.48	2.67	2.2	3.0
Impurity Level (ppm)				
Iron	3.1	1.0	0.3	0.02
Aluminum	4.9	2.4	1.5	0.69
Sodium	3.4	0.3	0.5	0.48

ND: not detectable (limitation of detection is 10μm)

At the second run, extremely high quality synthetic quartz crystals were obtained.

Etch channels were checked with X ray topography. Figure 3 shows the topography of the etch channel free crystal grown in the platinum vessel and Figure 4 is a photograph to show the surface of the crystal.

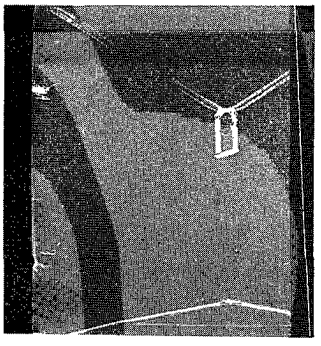


Figure 3. Topography of the Etch Channel Free Crystal

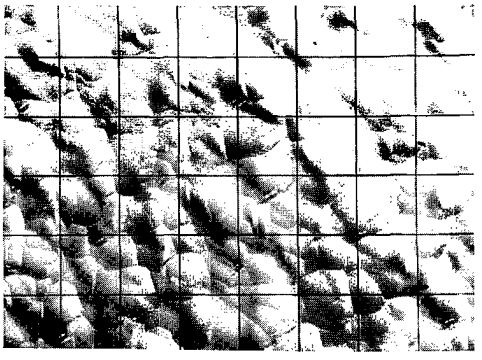
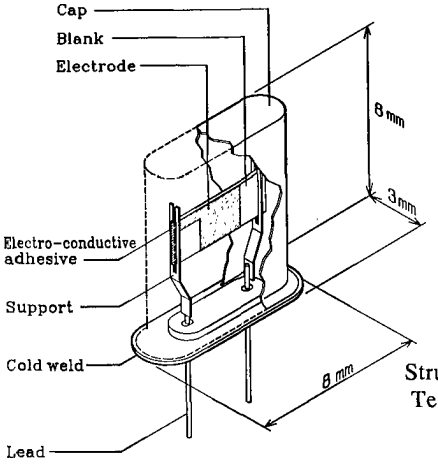


Figure 4. Photograph of the Crystal Surface($\times$150)

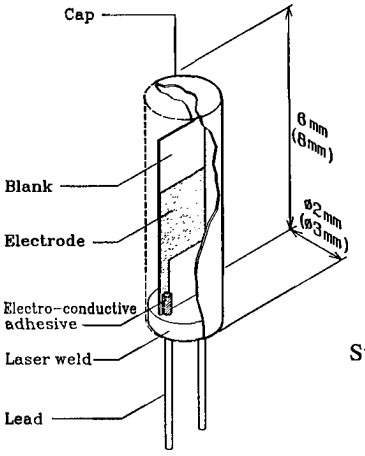
CRYOGENIC TEMPERATURE SENSOR USING HIGH QUALITY QUARTZ CRYSTAL

SENSOR CONSTRUCTION

In order to obtain high resolution and reliability, we have developed and tested a newly designed quartz crystal temperature sensor operated at about 28MHz in fundamental mode. The construction of the temperature sensor is shown in Figure 5. The quartz crystal blank was sliced in optimum Y cut, which is called Ys cut, from the high quality synthetic quartz crystal. After depositing the silver electrode on the blank, it is bonded to the leads on one end with electro-conductive adhesive. The enclosure is sealed with a radial compression method in helium atmosphere and is sealed again with laser beam welding to keep the high reliability. The cantilever support system for the blank also gives high reliability by reducing thermal stress in it.



Structure of Existing Temperature Sensor



Structure of Improved Temperature Sensor

Figure 5. Structure of the Sensor

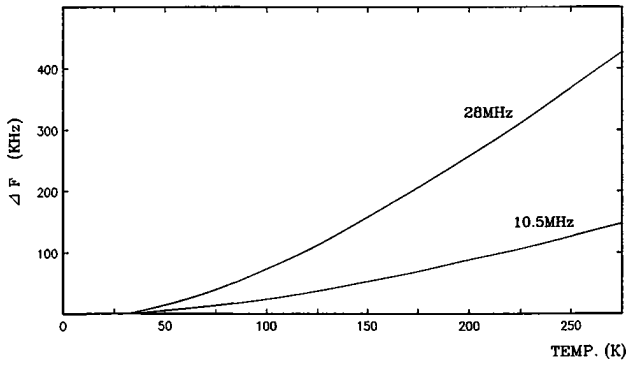
TEMPERATURE DEPENDENCE OF THE RESONANT FREQUENCY

The newly designed small quartz crystal temperature sensors have been made with several different aspect ratios of the quartz blanks. The sensors have been calibrated by using a platinum resistor, a thermocouple, Au-0.07 atom% Fe/ Chromel, and a germanium resistor. The quartz crystal units, sensors, are operated in thickness shear mode through extended cable connection or a radio wave from a transmitter. The resonant frequency of the optimum Ys cut quartz crystal varies non-linearly but continuously from room temperature to liquid helium temperature. The dependence of the frequency and CI, crystal impedance, are shown in Figure 6, and the sensitivity of the sensors are listed in Table 6.

Table 6. Sensitivity of the Sensors

Temperature(K)	Sensitivity (Hz/K)	
	Sensor Frequency 10.5 MHz	Sensor Frequency 28 MHz
10	1	10
20	45	100
25	90	200
50	350	700
100	550	1300
150	650	1800
200	800	2300
250	850	2550

As it is obvious from the data above, 28 MHz sensor has the higher sensitivity than old 10.5 MHz sensor.



Frequency-Temperature Characteristics

Figure 6. Frequency-Temperature Characteristics of the sensors

REPRODUCIBILITY

Since the reproducibility of the sensor is the most important property for cryogenic temperature sensor, several quartz crystal temperature sensors were selected and subjected to the thermal shock test. The temperatures for this test were 283K and liquid helium temperature, 4.2K. The sensors were soaked in the methanol bath and liquid helium bath alternately. A standard quartz crystal temperature sensor was used in the methanol bath to keep the temperature at 283K ± 1mK. The test results are shown in Figure 7. The sensors show excellent reproducibility of frequency within 2.6 ppm and 0.3 ppm, 0.001K, typically.

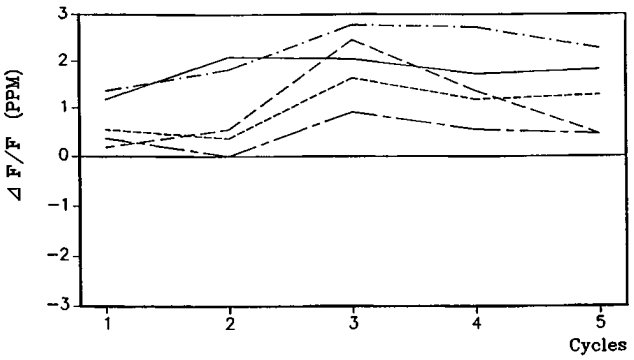
IMPROVEMENT IN PERFORMANCE BY USING SWEPT QUARTZ CRYSTAL

The synthetic quartz crystal has a loss peak at about 60K and it causes the deterioration, dip, in crystal impedance at this temperature. The loss peak is probably due to contaminant, sodium ions, in the quartz crystal as reported by Fraser [5]. Since the dip in crystal impedance causes abnormal frequency shift, the dip should be removed for accurate measurement. In order to eliminate the dip, sweeping method was applied to the synthetic quartz crystal. The sweeping conditions are shown in Table 7.

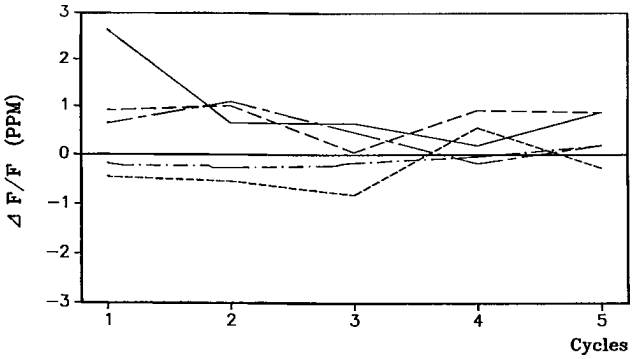
Table 7. Sweeping Conditions

Item	Condition
Test sample size	10 x 8 x 200 mm (lumbered type)
Electrode	Gold
Atmosphere	Air (1 atm.)
Voltage	DC 1 kV
Temperature Change	Heated to 500 deg. at a rate 20deg/hour
	Keep 500 deg. for 24 hours
	Cooled down at a rate 20deg/hour

A new quartz crystal temperature sensor was made from the swept crystal and the crystal impedance at about 60K was checked. Figure 8 shows the test results for the sensors made from swept and unswept quartz crystal. As can be seen in the figure, the dip in crystal impedance is eliminated for the sensor made from the swept quartz crystal. Figure 9 indicates the measurement error due to the dip in crystal impedance of the temperature sensor. The error has been improved to a great extent by using the swept quartz crystal.

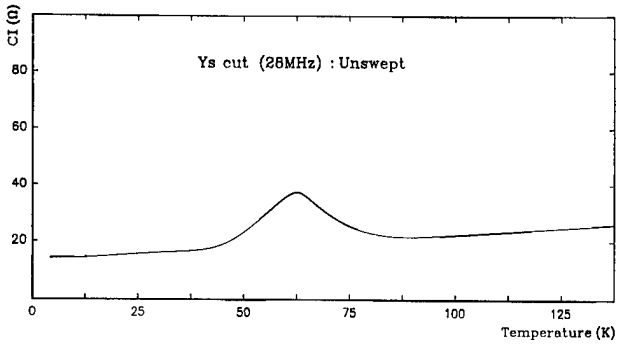


Frequency Shift of Existing Temperature Sensor After Thermal Shock

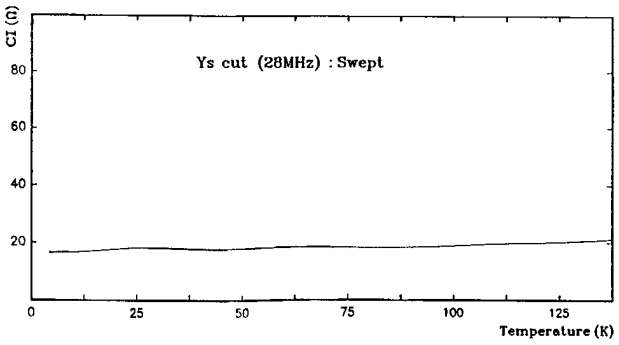


Frequency Shift of Improved Temperature Sensor After Thermal Shock

Figure 7. Results of the Thermal Shock Test

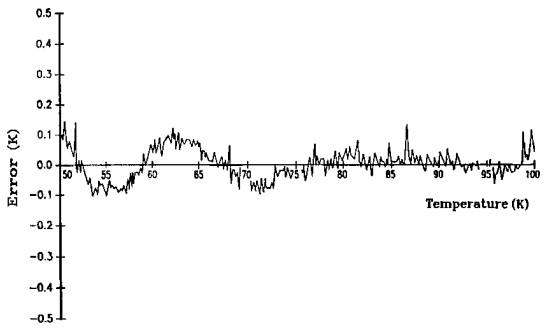


C1-Temperature Characteristics of Temperature Sensor made from unswept quartz crystal

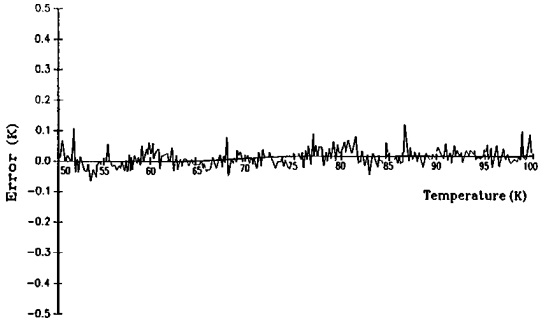


C1-Temperature Characteristics of Temperature Sensor made from swept quartz crystal

Figure 8. Temperature Dependenceof the Crystal Impedance



Error of Unswept Sensor



Error of Swept Sensor

Figure 9. Error of Temperature Measurement

WIRELESS ECHO SENSOR SYSTEM

Since the measurement at cryogenic temperatures are so sensitive to the thermal flow through the wire of the sensor, a wireless sensing system called wireless echo sensor system has been investigated. The system is based on the stored resonance energy and the emission of electro-magnetic wave of the sensor, quartz crystal unit, at the resonance frequency. The basic concept is illustrated in Figure 10. The sensor assembly consists of a quartz sensor and an antenna coil, sensor antenna, as shown in Figure 11. The temperature is converted to the resonance frequency of the quartz crystal temperature sensor. The echo wave from the quartz resonator is emitted through sensor antenna and is received by a receiving antenna, TX-RX antenna, which also serves as a transmitting antenna. The temperature is determined by measuring the frequency of the echo waves.

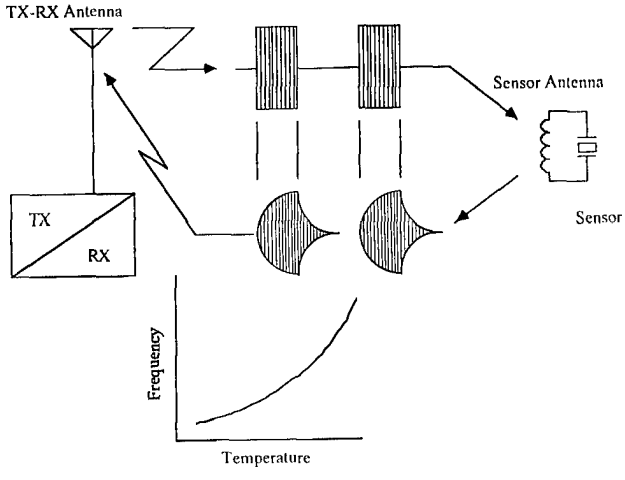


Figure 10. Principle Diagram for Wireless Echo Sensor System

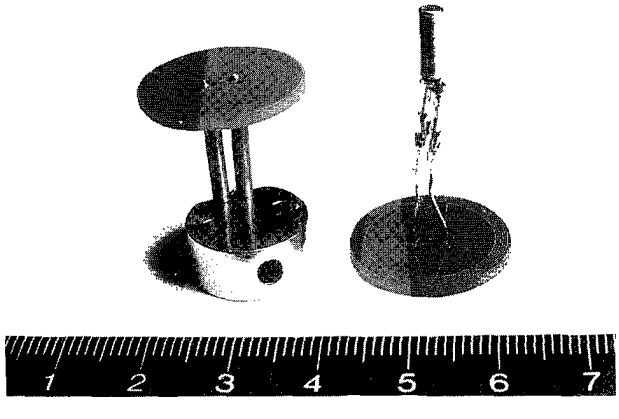


Figure 11. Photograph of the Wireless Temperature Sensor

CONCLUSIONS

By using a platinum vessel as a liner in an autoclave, a high quality synthetic quartz crystal has been produced. The density of etch channel, etch pit and inclusion is not detectable at the criteria of 10 μm. A technique for pressure balance for the platinum vessel is devised to prevent the contamination from the outside of the platinum vessel.

A cryogenic temperature sensor has been developed using the high quality synthetic quartz crystal. The quartz crystal temperature sensor is operated at 28MHz in fundamental mode and the sensor measures the temperature over 25K to 400K with sufficient resolution of 200Hz/K.

A wireless temperature sensing system called wireless echo sensor system has been developed to avoid the measurement error caused by the thermal flow through the wire of the sensor. Since battery is not required for the sensing element, it is suitable for the cryogenic temperature measurement.

ACKNOWLEDGEMENTS

The authors would like to thank T. Tajima for his advice and assistance during this investigation. The authors would also like to thank M.Watson of Spectrum Electronics for his assistance.

REFERENCES

[1] Robert Irving, Jonathan Foise, Edward Leeson, and Gary Johnson
"Recent Experiments in a Silver Lined Autoclave"
Proceedings of the 42nd Annual Frequency Control Symposium, 127, 1988

[2] D. L. Hammond, C. A. Adams, P. Schmidt
"A Linear Quartz Crystal Temperature Sensing Element"
Proceedings of the 19th Annual ISA Conference, 1964

[3] T. M. Flynn, H. Hinnah, and D. E. Newell
"An Improved Cryogenic Thermometer"
Advances in Cryogenic Engineering, vol. 8, 334, 1962

[4] A. G. Smagin and B. G. Mil'shtein, *"Quartz Frequency Thermometer"*
Instruments and Experimental Techniques, vol. 12, 932, 1970

[5] D. B. Fraser, *"An Elastic Effect of Alkali Ions in Cristalline Quartz"*
Journal of Applied Physics, vol. 35, No. 10, 2913, 1964

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

OUTGASSING OF QUARTZ

D.E. Pierce[1,2], R.A. Murray[1], R. Lareau[1], S. Laffey[3], and J.R. Vig[1]

[1]U.S. Army Research Laboratory, Fort Monmouth, N.J. 07703-5601
[2]National Research Council-Resident Research Associate
[3]Vitronics Inc., Eatontown N.J. 07724

ABSTRACT

The outgassing of quartz crystal plates was measured with a high resolution mass spectrometer using thermal desorption techniques. As the quartz was heated from 40°C to 1200°C in a vacuum, the mass spectrometer identified several types of gases. The primary species that appeared were NH_3, OH, and H_2O. The bonding sites of the OH and H_2O are discussed, including the temperature dependence of the desorption peaks which is used to calculate the activation energies of the surface states. Bakeout times and temperatures to insure clean surfaces are calculated. The effect of bakeout conditions on aging is also discussed.

The presence of NH_3 was unexpected, but it is believed that the NH_3 was due to the sample preparation which included etching the samples in ammonium bifluoride.

I. INTRODUCTION

Many operations need to be performed correctly to make a high precision quartz resonator. The quartz needs to be grown with low defect and impurity density and then cut to the correct angles with a high degree of accuracy. Next, the quartz blanks need to be lapped, contoured, polished, electroded, mounted, and sealed into a vacuum enclosure. If any of these steps are performed incorrectly, the performance of the crystal is degraded. However, even when everything is done correctly, the aging is rarely lower than parts in 10^{12}/day.[1] Many factors are thought to contribute to aging: stress relief in the mounting structure, mass transfer to or from the resonator surfaces due to adsorption and desorption, changes in the electrodes, changes in the quartz material, and leaks into the enclosure.

It has been suggested, that the outgassing of quartz is a possible aging mechanism, however, the magnitude of this mechanism has been unknown. Although a large amount of information is available on the outgassing characteristics of vitreous SiO_2[2-4], no reports on the outgassing of quartz crystals could be found.[1]

One of the main obstacles to reducing the aging is that after the crystal enclosure is sealed, there is no way to tell what is happening inside. In the past, one of us (JV) had tried to open metal enclosures and analyze the interior gas using a mass spectrometer. It was found that punching a hole into an unsealed enclosure released as much gas as punching a hole into a sealed resonator. Measuring what was originally in the enclosure was impossible with such a technique. Since we cannot look into the resonators after they are sealed, we decided to look at what came off the quartz using a thermal desorption mass spectrometer (TDMS).[5-7]

Heating speeds up the removal of impurities from surfaces and increases the diffusion rate through the bulk.[3] Most manufacturers include a high temperature bakeout in their processes, but this bakeout is usually limited by the mounting structure components or electrodes to approximately 300°C. The bakeout of the quartz itself is limited to 573°C, due to a phase change (although changes can take place at lower temperatures, especially if the quartz is stressed). Several papers suggested that hydrogen, the single largest impurity in cultured quartz, can be mobile even without the electric field used in sweeping.[8-11] It has been reported that the hydrogen concentration changed at room temperature over a period of years.[12] If this were true, it could certainly affect the performance of resonators. So, we were interested in determining the outgassing of quartz crystals as a function of temperature. Our initial findings on the outgassing from quartz are reported in this paper.

II. EXPERIMENTAL

The outgassing experiment centers on the ability to heat processed quartz blanks in an ultrahigh vacuum while monitoring the vaporizing species. Two methods of heating were used; tantalum cell and laser.

U.S. Government work not protected by U.S. copyright

A. Tantalum Cell Heating:

Pieces of processed quartz blanks were introduced into a tantalum foil cell to which a tiny thermocouple was attached (see Figure 1). The cell was spotwelded to a vacuum feedthrough which could be introduced into the vacuum chamber for analysis. Cell temperature was ramped in a linear fashion up to 1500°C with feedback control of a 100 amp power supply using a fast PID (proportional-integral-difference algorithm) temperature controller. Typical ramp rates were about 8K/sec. Fast ramp rates were advantageous in determining small quantities of outgassing species.

Quartz pieces were placed in the cell in air and then placed under vacuum. Once the appropriate vacuum was achieved, the cell was heated and vaporizing species and their outgassing rates were determined by line of sight detection with a mass spectrometer.

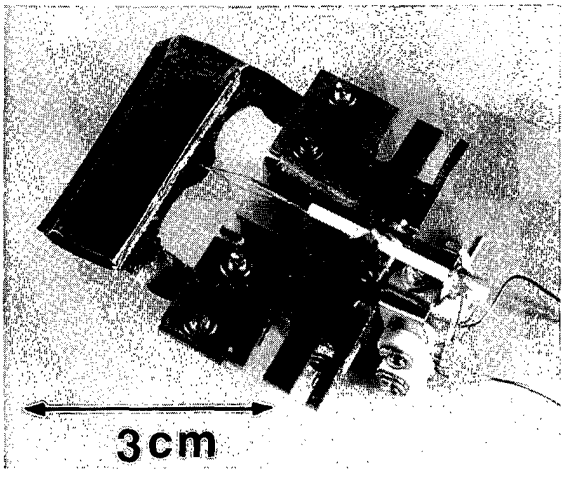


Figure 1: Tantalum vaporization cell

Initial experiments with the cell alone, without any sample, revealed that water desorbed from the surface of the cell. Tantalum metal is covered with an oxide that adsorbs water. The region of the tantalum cell water desorption peak was useful as an internal sensitivity calibration to determine the relative quantities of water and other species desorbing from the quartz sample.

B. Laser Heating:

The other type of heating consisted of directing CO_2 laser radiation onto the entire back of a quartz blank (approximately 1 cm diameter-see Figure 2).[6] A blank suspended by two small gold coated molybdenum clips could be precisely positioned such that the laser beam was centered on the backside of the crystal while the front side was directed toward the entrance of the mass spectrometer ion source. The laser beam delivery system included a series of lenses that defocussed the beam to approximately cover the entire surface of the quartz blank.

A pyrometer which was focussed on the front side of the crystal utilized a chopped pyroelectric detector and ZnSe optics. The low temperature limit of this pyrometer was approximately 120°C. The pyrometer used a narrow band IR filter with transmission between 5.5 and 7 microns. This filter was chosen because quartz has a high emissivity in this wavelength range (approx 0.90).[13] The pyrometer was calibrated using a thermocouple joined to the surface of a quartz blank and the estimated accuracy of the quartz surface temperature measurement was about +/- 10 degrees.

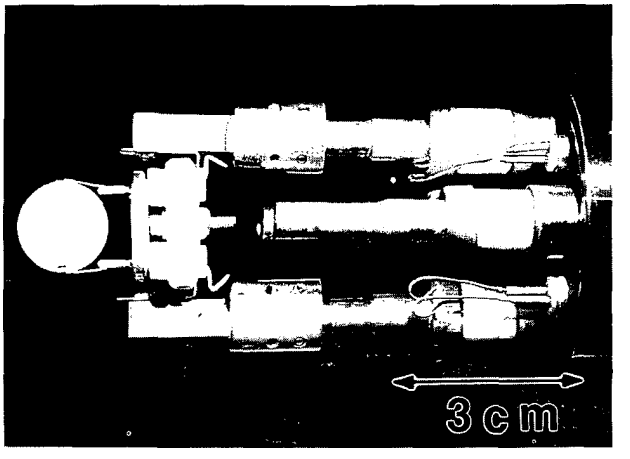


Figure 2: Quartz blank in laser desorption holder

C. Analysis System:

The multi-chamber vacuum system consisted of a preparation chamber, analytical chamber, and mass spectrometer chamber. The analytical chamber used two turbopumps in series and included a liquid nitrogen trap with an integral titanium sublimator. Typically, a sample could be analyzed within one to two hours in the mid-10^{-9} torr range after being introduced from air.

The analysis chamber was uniquely designed for thermal desorption experiments and material outgassing studies. Line of sight to the mass spectrometer ion source could be maintained from the quartz blank surface or the opening of the tantalum cell. Several viewports were present which gave a clear view of the sample in the analysis position. One viewport was used for introduction of CO_2 laser radiation and another used for sample surface

temperature measurement using the specially designed pyrometer.

The mass spectrometer has ultra-high mass resolution provided by the triple focussing (electrostatic-magnetic-electrostatic) geometry. The advantage of high mass resolution is that molecules very near in mass can be completely separated. Of particular relevance to this study was the ability to separate NH_3(17.027amu) and OH(17.003amu).

D. Sample Preparation:

The quartz samples used in this work were grown by three different manufacturers. Thermo Dynamics, Sawyer Research Products and Motorola. The IR absorption values at 3500cm^{-1} and the H content are shown in Table 1.

	Thermo-Dynamics	Sawyer	Motorola
$alpha_{3500}$	0.061	0.19	0.08
Total H content (relative to atomic Si)*	300 ppm	900 ppm	400 ppm

* Special high alpha bars provided for outgassing study

Table 1: Quartz samples

The samples were prepared by: lapping using a five and then a three micron aluminum oxide abrasive, cleaned in boiling trichloroethane (TCA) and then ultrasonically cleaned in TCA, cleaned using RCA standard cleaning solutions 1&2 then UV ozone cleaned, etched in saturated ammonium bifluoride for 20 minutes at 85°C then thoroughly rinsed in hot 18 megohm-cm deionized water and blown dry with UHP N_2.

III. RESULTS

A. Outgassing using tantalum cell:

In Figure 3, the first set of desorption spectra shows outgassing from quartz in the tantalum cell. NH_3^+,OH^+,H_2O^+, and CO_2^+ were seen as well as minor components that appear to be hydrocarbons. In all samples, water was detected and along with the H_2O^+ signal, OH^+ was also detected. In every case the OH^+ mirrored the H_2O^+, maintaining a constant OH^+/H_2O^+ signal ratio typical of hydroxyl ion produced as a water byproduct in the ion source of the mass spectrometer. No evidence for molecular OH desorption was obtained. Also, no elemental hydrogen (H or H_2) or any fluoride species were detected. Fluoride might be expected as a result of the ammonium bifluoride etch. We looked for fluorine with Auger electron spectroscopy in our early experiments and none was detected.[14]

Beginning about 110-120°C, a desorption state of water is clearly seen both for the Cell and the Cell+Sample spectra (see Figure 3). The upper water spectrum represents the cell which was brought out into air and then returned to the vacuum system and analyzed. The 180°C peak constitutes water adsorbed on the cell surface most likely in the molecular form. This peak shows a considerable tail at higher temperature (220-600°C) and probably represents surface hydroxyls (H-bonded?-see discussion). At even higher temperatures (>750°C) a continuous outgassing of water begins that rises with temperature. Further experiment showed that this water is due largely to outgassing of the tantalum cell support assembly and vacuum surfaces. With high temperature heating of the cell, the assembly temperature rises due to thermal conduction and radiation heating.

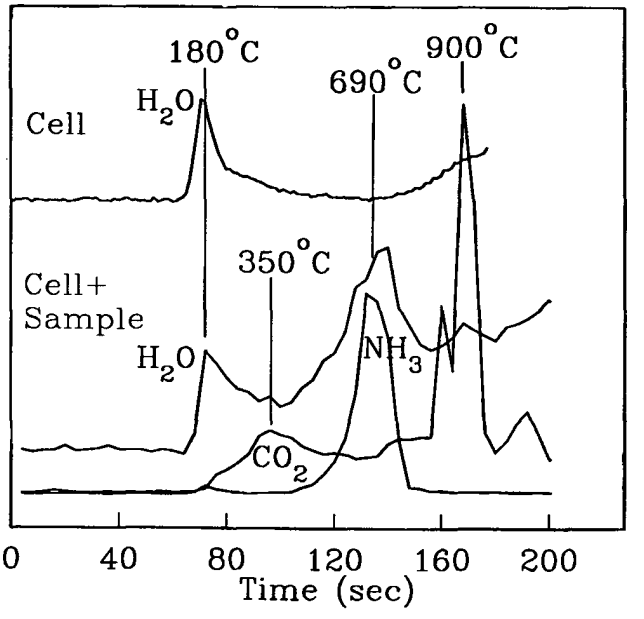


Figure 3: Outgassing of quartz in tantalum cell

In the set of desorption spectra labelled "Cell+Sample", the water desorption spectrum has changed considerably as a result of the presence of the quartz sample. As expected, desorption from the cell surface is evident from the peak at 180°C. In addition there is a distinct peak at higher temperature (labelled

"690°C"). At higher temperatures, a continually rising water signal begins which is similar to that seen with the empty cell.

In addition to water, NH_3 and CO_2 are seen desorbing at various temperatures. The CO_2 is the first species detected after the cell surface water and has a peak at 350°C. In addition, CO_2 is detected continually with subsequent peaks in the 900°C region. The origin of the CO_2 is not clear, but may be related to the growing of the quartz in presence of carbonate which might leave some carbonate species that can be desorbed.

At cell temperatures above about 400°C, outgassing of water and ammonia begin from the quartz and both peak at a cell temperature of about 690°C. The peak heights of the three species shown in Figure 3 for the Cell+Sample are the actual relative mass spectrometer signals. Figure 3 does not include any correction for the instrument's sensitivity to the various species, however, the inclusion of normalization would have only a minor effect on the relative signal intensities.

The tantalum cell temperature was measured by a thermocouple spotwelded to the surface. Linear temperature ramps were obtained. One problem however, was that as the cell was ramped, the actual temperature of the quartz pieces inside (1 to 7 pieces) was unknown. A calculation of the temperature lag, based on a heat balance model of the cell, predicted a considerable lag. Since the quartz pieces varied in number, size and shape, the variations in actual sample temperature among runs would be considerable.

The tantalum cell analysis was particularly useful in determining the identity and quantity of various outgassed species as well as a relative temperature of their production. However, as a result of the lag in temperature, a direct correlation of the cell outgassing results to resonator stability is obscured. Of particular interest are the "690°C" peaks because, based on the thermal model of the cell, the temperature at which the quartz begins to outgas was actually about 120°C!

B. Outgassing using laser heating:

In order to obtain more precise outgassing temperature information, quartz blanks were analyzed by laser thermal desorption with pyrometric detection of the surface temperature. The laser was adjusted to give a ramp rate on the order of 10K/sec over the range 150-400°C. Figure 4, shows the desorption spectra of water and ammonia from a quartz blank. Here it can be seen that both species begin outgassing at the same time. The onset temperature is estimated to be about 110-120°C. The ammonia peaks somewhere in the 200°C region, however, the water did not show a distinct peak (until 475°C). In some of the samples analyzed with the laser under the same conditions, a pronounced peaking around 200-300°C was seen for both ammonia and water.

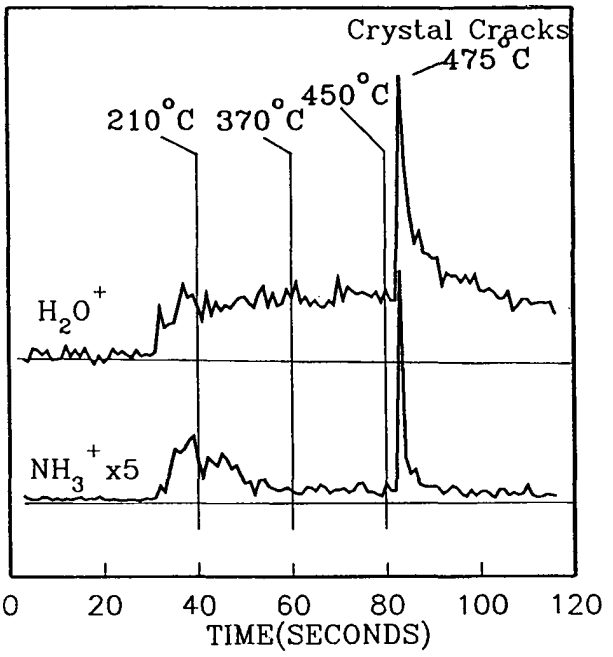


Figure 4: Outgassing of quartz using laser heating

All quartz desorption results showed a close relationship between water and ammonia at the onset, as if they are mutually bonded or bonded in some similar way to the surface. In some of the samples, the ammonia signals were comparable in size to that of water (Figure 3, 690°C region) and in other cases were 20% to 30% of the water signal (see Figure 4, 690°C region).

In Figure 4, the water signal remains well above the baseline in the region 120°C-450°C. Several mechanisms, operating simultaneously or independently, can explain the persistence of the H_2O signal in this region. Multiple states of bound water can be considered which involve varying degrees of hydrogen bonding and various types of water-quartz surface bonding. Also, inefficient pumping of water in the vacuum system can be considered. This mechanism could be used to explain a high temperature tailing off of water desorption states, where the pressure rise from a water desorption state would decrease with some time constant. Since water peaks were seen in some cases (for example the water on the tantalum cell surface), the system's pumping speed for water appears to be fast enough to prevent significant quantities of residual water from lingering in the vacuum system. Finally, the contribution of near surface and bulk

water and water related species can be considered. Water may be moving continually to the surface and desorbing as the temperature increases.

At higher temperatures, a sharp desorption peak of both water and ammonia occurs at about 475°C, at which point the crystal fractures. As is well known, crystalline quartz undergoes a phase transition at 573°C and it is possible that the cracking under laser heating is related to this transition. Since the laser heating is directed onto one side of the crystal during analysis, a thermal gradient through the crystal exists that can induce stresses leading to fracture. In addition, the laser beam profile is not flat across the crystal surface and the variation in heating across the face also induces stresses.

In Figure 5, the laser temperature ramp associated with the desorption spectra in Figure 4 is shown. At the point of crystal fracture, the temperature immediately drops. This drop results from the loss of crystal fragments. The remaining pieces held by the clips are still being heated by the laser and the temperature is still going up with continued radiation. The pyrometer only senses the radiation output from the target area and when a fraction of that area is lost, an apparent drop in temperature results.

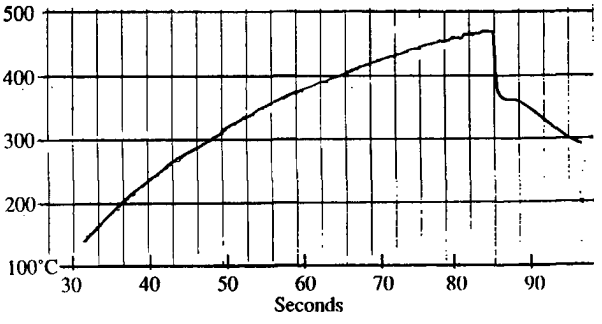


Figure 5: Quartz temperature ramp under laser heating

The instantaneous desorption as the crystal cracks may be related to the increase in surface area along the fracture lines. When this internal surface is exposed with the crystal at high temperature, the water near the surface may be escaping rapidly. Ammonium bifluoride solution can progress deep into the crystal during the etching process forming etch channels and an artifact of the ammonium bifluoride may have been trapped by capillary action and released suddenly upon cracking.

Rapid outgassing associated with crystal fracture occurred in samples under laser heating, however, with the tantalum cell, heating did not lead to fracturing. This can be explained by the fact that smaller pieces were used and heating inside the cell produces a more uniform temperature field, with lower thermal stresses.

Figure 6, shows the ammonia desorption from three samples performed in succession under identical conditions. It can be seen that the pattern of desorption is the same for each, but the amount of ammonia is different. The three samples were processed together but the surface of the Q-16 sample had noticeable etch damage which was assumed to be a clustering of etch pits. The increased desorption signal correlates with this observation. Thermal desorption analysis may provide a measure of etch pit and channel density. Preliminary results (not shown) also revealed a strong relationship between water desorption and surface roughness, where samples etched to give greater surface roughness desorbed considerably larger amounts of water.

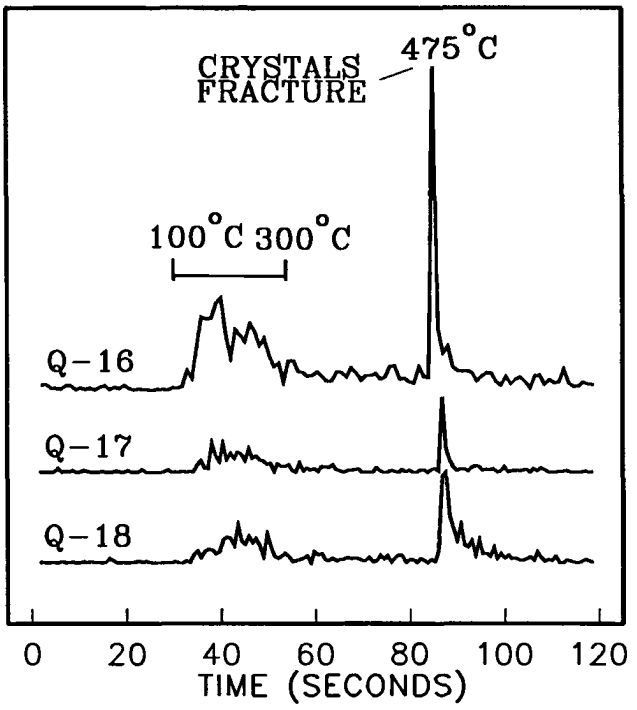


Figure 6: Laser desorption of ammonia from quartz

IV. DISCUSSION

Considerable work has been done studying adsorption on silica which can be used as a model for adsorption on quartz surfaces. In particular, the adsorption of water onto oxide surfaces has been studied in great detail.[15-17] As a result of extensive infra-red studies, various states of water-silica bonding have been identified.[18-25] There are three primary categories: molecularly adsorbed, hydroxyl formation involving hydrogen bonding, and isolated hydroxyl formation. These

various types of bonding of water have different temperature stabilities which are directly relevant to the discussion of the outgassing properties of quartz.

Molecular adsorption: Water termed molecularly adsorbed, resides on top of a hydroxyl covered surface and can exist as multilayers at lower temperatures. As SiO_2 is placed under vacuum, some water immediately desorbs at room temperature. As a result of differing degrees of hydrogen bonding, molecular water persists to above room temperature. It has been suggested [20],that molecularly bound water can be desorbed from the surface after heating to a temperature of about 150°C.

Hydroxyls/H-bonded: Water can form a stronger chemical bond by reacting directly with the surface to form hydroxyls. For example, infra-red studies have indicated that water will adsorb and readily react with certain siloxane bridge bonds Si-O-Si to form two Si-OH units.[20-25] In the reverse reaction, two Si-OH bonds can be condensed to form a siloxane bridge and water. The temperature at which the condensation reaction takes place depends on the nature of the Si-OH bonds involved. As determined by titration of surface hydroxyls, the silica surface can be hydroxylated to the extent of $4.6/100A^2$.[19] Similarly, ^{22}Na labelled NaOH quartz was found to have $4.25\ OH/100A^2$.[17] Based on a calculation of the packing density of hydroxyls, the maximum number of hydroxyls for quartz(001) and quartz(101) surfaces was predicted to be $9.6/100A^2$ and $6.0/100A^2$ respectively.[17]

With a high surface coverage of OH, hydrogen bonding occurs between nearby hydroxyls. The hydrogen bonding is believed to have the effect of weakening the Si-OH bond.[20] As silica is heated, first the molecularly adsorbed water desorbs followed by water originating as hydrogen bonded hydroxyls. Infrared absorption indicates that the hydrogen bonded hydroxyls are removed over a broad temperature range to about 500°C where they are completely removed.[18-20] The broad range can be explained as varying degrees of hydrogen bonding destabilization of the Si-OH bond. One important aspect to note is that the process is rapidly reversible upon exposure to water vapor.

Hydroxyls/non H-bonded: Hydroxyls which are not hydrogen bonded are divided into two types, Type A: isolated single hydroxyls and Type B: isolated pairs. The isolated pairs have been explained as geminal, $Si(OH)_2$ and vicinal, HO-Si-O-Si-OH (or nearby).[19,25] Both have OH groups geometrically too far apart to H-bond. The Type B groups are less thermally stable and will condense liberating water at a lower temperature than Type A. In the temperature range above 500°C to about 800°C the Type B group are condensed preferentially. Type A, which are considered truly isolated, persist to at least 1000°C.[19]

Outgassing studies of glass particles of various composition showed outgassing states spanning the temperature range of about 90°C to 380°C.[26] Another study considered the effect of particle size on the outgassing of water from magnesium-alumino-silicate glass with a water content of approximately 400ppm.[27] In that study water desorption states were observed in the range 120°C to 500°C. These states were not dependent on particle size and were judged to be associated with the surface. At higher temperature, another distinct water desorption peak was observed but the peak position was dramatically affected by particle size. This type of water was considered to be originating from the bulk of the glass material.

V. CONCLUSIONS

Quartz crystals can be expected to have water related surface species similar to those detected on silica surfaces. In addition, quartz has internal hydrogen that exists in various forms, such as molecular water, Al-OH at aluminum substitution sites, and Si-OH at defects and dislocations.[28] Typical hydrogen concentrations for synthetic quartz range from 250-2500 H-atoms per million Si atoms.[12] It has been suggested that molecular water is the dominant hydrogen species.[11]

With respect to resonator performance, outgassing from the surface or bulk contributes to aging. In this outgassing study, we hoped to characterize the nature of hydrogen in quartz by following the thermal desorption of hydrogen bearing species such as H_2, OH, and H_2O and in the process discovered the presence of NH_3, CO_2, and other minor species.

Desorption of water from quartz can be seen in Figure 4, where water was detected continually after the sample reached about 110°C. The onset temperature identifies the most weakly bound state. If we assume the peak of this state to be on the order of 200°C and we further assume first order desorption kinetics and a standard preexponential factor of 10^{13}/sec, an activation energy of 28 kcal(119kJ) can be calculated using Redhead's desorption analysis method[5]. Higher temperature desorption results from states with higher activation energies. Desorption centered at 300°C and 400°C can result from states with activation energies of about 35 kcal(146kJ) and 41 kcal(172kJ) respectively.

The practical implication of the outgassing analysis can be seen when baking time and temperature is considered. The half-life of a desorption state can be estimated from kinetic parameters and the results of representative calculations are given in Table 2.

If the bakeout of resonators was 100°C, it would be impossible to remove water having activation energies of desorption of 35 and 41 kcal. The more weakly bound water however, with a lower activation energy (28 kcal) is easily removed (note: 10 half lives=5 hours which removes 99.9%). At 200°C the 35 kcal state is removed in a reasonable time and at 300°C the 41 kcal is removed.

	Activation Energy (kcal)/ Desorption Peak Temperature (°C)		
Bake Temp.	28 kcal/ 200°C	35 kcal/ 300°C	41 kcal/ 400°C
100°C	30 min.	7 years	10^4 yrs
200°C	1 sec.	3 hours	72 days
300°C	3 msec.	2 sec.	1 hour

Table 2: Desorption half-life of water

The efficiency of the baking step in resonator fabrication is only as good as the vacuum that the resonator is contained within. As mentioned earlier with respect to studies done on silica, hydroxyls will immediately reform on surfaces which have been baked to remove water (100°C-400°C). If these hydroxyls can be assumed to be bound with activation energies in the 28-41 kcal range a significant contribution of outgassing to resonator aging can be expected as a result of adsorbed or readsorbed water.

As an example, consider a single monolayer of hydroxyl groups present on an ideally flat quartz surface. Assume also a moderate bonding to the surface, with an activation energy for desorption of 35 kcal. If the resonator was operated at a constant 100°C, e.g. in an OCXO, then 1/2 of the monolayer could be expected to have desorbed in the first seven years. A 5MHz fundamental mode crystal is about 1 million molecular layers thick and the loss of a 1/2 monolayer of water will constitute a change in frequency on the order of 1/10 ppm (adjusting for atomic mass), which will result in a frequency change of 4 parts in 10^{11}/day.

300°C Bakeout	Surface coverage (monolayers)	Desorption rate at 100°C (monolayer /day)	Aging rate per day
0 hr.	1	8×10^{-7}	8×10^{-13}
1 hr.	0.5	4×10^{-7}	4×10^{-13}
10 hrs.	0.001	8×10^{-10}	8×10^{-15}

Table 3: Aging rate due to a monolayer of 41 kcal bound water

The aging rate at 100°C can be calculated for quartz baked at 300°C assuming one monolayer of water with a binding energy of 41 kcal. The more weakly bound water will be rapidly removed at this temperature and need not be considered. In Table 3, the bake time, surface coverage after baking, desorption rate at 100°C, and aging rate are given. It can been seen that a bakeout temperature of 300°C will remove sufficient water, so that an aging rate of better than 1 part in 10^{12} can be achieved. Note, that in this simplistic example no provision for the effect of water under the surface has been considered. A more detailed study of surface and bulk outgassing is currently underway.

VI. ACKNOWLEDGMENTS

We would like to thank Karen Blisnuk for her help in sample preparation and Don Eckart for his photographic assistance. We would also like to thank Motorola, Sawyer Research Products, and Thermo Dynamics for providing samples and IR data, and Ray Filler for helpful comments and editing.

VII. REFERENCES

1) J. R. Vig and T.R. Meeker, Proceedings of the 45th Annual Symposium on Frequency Control, pp. 77-101, 1991

2) R. Glang et al., "High Vacuum Technology" in Handbook of Thin Film Technology, L.I. Maissel and R. Glang, ed., McGraw Hill, pp. 2-39 to 2-59, 1970.

3) W.G. Perkins, Journal of Vacuum Science and Technology, v.10, pp. 543-556, 1973.2)

4) P.A. Redhead, J.P. Hobson, and E.V.Kornelson, Physical Basis of Ultra-High Vacuum, Chapter 10, Chapman and Hall, 1968.

5) P.A. Redhead, Vacuum, v.12, pp. 203-211, 1962

6) D.E. Pierce, R.P. Burns, and K.A. Gabriel, Thin Solid Films, v.219: pp.312-317, 1992

7) D.E. Pierce et al.,Tribology Transactions, v.34, pp. 205-212, 199

8) D. Griggs, Geophys. J. R. Astr. Soc., v. 14, pp. 19-31, 1967.

9) J. D. Blacic, Tectonophysics, v. 27, pp. 271-294, 1975.

10) A.C. McLaren, J.D. Fitz Gerald, and J. Gerretsen, Phys. Chem. Minerals, v. 16, pp. 465-482, 1989.

11) R. D. Aines, S.H. Kirby, and G.R. Rossman, Phys. Chem. Min., v. 11, pp. 204-212, 1987.

12) J.C. Brice, Rev. Modern Phys. v. 57, no. 1, pp. 105-146, 1985.

13) R.T. Rewick and R.G. Messerschmidt, Spectroscopy, pp. 36-39, August 1992.

14) J.R. Vig et al., Proceedings of the 28th Annual Symposium on Frequency Control, pp. 96-108, 1974.

15) D.A. Tompkins, "Surface Chemistry of Oxides", Discussion of the Faraday Society, No.52, pp. 205-275, 1971.

16) H.P. Boehm and H. Knozinger, in Catalysis-Science and Technology, v.4, J.R. Andersen and M. Boudart, Eds., Springer, Berlin, 1984.

17) H.P. Boehm, Angewandte Chemie, International Edition, v.5, pp.533-622, 1966.

18) J.B Peri, Journal of Physical Chemistry, v.70, pp. 2937-2945, 1966.

19) C.G. Armistead et al., Journal of Physical Chemistry, v.73, pp. 3947-3953, 1969.

20) M.L. Hair, Journal of Non-Crystalline Solids, v.19, pp. 299-309, 1975

21) B.A. Morrow, I.A. Cody, and L.S.M.Lee, Journal of Physical Chemistry, v.80, pp. 2761-2767, 1976.

22) T.A. Michalske and B.C. Bunker, Journal of Applied Physics, v.56, pp. 2686-93, 1984.

23) B.C. Bunker, D.M. Haaland, T.A. Michalske and W.L. Smith, Surface Science, v.222, pp.95-118, 1989.

24) B.C. Bunker et al., Surface Science, v.201, pp.406-426, 1989.

25) B.A. Morrow and A.J. McFarlan, Journal of Physical Chemistry, v.96, pp.1335-40, 1992.

26) H. Wittkopf, Vacuum, v.37, pp. 819-23, 1987.

27) M.A. Frisch and E.A. Geiss, High Temperature Science, v.26, pp. 175-186, 1990.

28) R.A. Murray, Proceedings of the 45th Annual Symposium on Frequency Control, pp. 9-16, 1991.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

HIGH-TEMPERATURE ACOUSTIC LOSS OF AT-CUT QUARTZ CRYSTALS

Joy E. Ferris and J. J. Martin
Department of Physics
Oklahoma State University
Stillwater, OK 74078-0444

Abstract

Sensor applications are increasing the use of quartz crystals at high-temperatures. At these temperatures the interstitial alkalis thermally escape from the traps and drift in the shallow potential wells of the Z-axis channel. Their random motion leads to a thermally activated acoustic loss which may degrade oscillator performance. Point defects may also lead to acoustic loss peaks; Fraser has reported a peak near 300° C in swept quartz. We have investigated the acoustic loss of unswept (Li-compensated). Na-compensated, and swept (H-compensated) AT-cut resonator blanks from room temperature to over 400° C. In swept quartz, we find a loss peak centered at 600 K which increases linearly with the aluminum content. This peak most likely results from relaxation of the Al-OH center. The acoustic loss of unswept and Na-compensated resonators remained relatively constant up to about 500 K; at higher temperatures it increased exponentially with $T(K)^{-1}$. The Al-OH loss peak in the swept blanks with higher Al content caused them to have a larger loss in the 400 to 600 K range than unswept blanks

Introduction

Sensor applications such as pressure transducers and thermometers are causing an increased use of quartz crystals at high temperatures. [1-4] Quartz contains interstitial alkali ions which are trapped near the substitutional aluminum or other defects. At high temperatures the alkali ions thermally escape from the traps and drift in the Z-axis channel. Their motion in the shallow potential wells of the channel leads to an acoustic loss which increases exponentially at higher temperatures.[5,6] An increasing acoustic loss corresponds to an increasing series equivalent resistance which may degrade the oscillator's performance. Sweeping or electrodiffusion replaces the alkalis with hydrogen and might be expected to remove the high temperature acoustic loss.[7,8] However, Fraser has reported an acoustic loss peak near 300° C in natural and cultured quartz that had been electrolyzed (swept).[5] More recently, Koehler[9] and Lipson, *et al.*[10] have also reported the presence of this high temperature acoustic loss peaks in swept quartz.

Anelastic loss arises when a defect or impurity has two (or more) equivalent orientations separated by a potential barrier.[5] The maximum loss occurs when the thermally activated hopping rate matches the frequency of the crystal. The loss, ΔQ^{-1}, is given by

$$\Delta Q^{-1} = \frac{nG\omega\tau}{1+\omega^2\tau^2} \qquad (1)$$

where n is the number of centers per unit volume, G measures the coupling to the crystal vibration, ω is the angular frequency of the vibration, and τ is the relaxation time for reorientation of the defect. Since the relaxation is thermally activated

$$\tau = \tau_o \exp(E_b / kT) \qquad (2)$$

where E_b is the barrier height, and τ_o contains the number of equivalent orientations and the attack rate. The loss is usually determined by measuring the crystal's Q or series resistance as a function of temperature. As the temperature is swept ΔQ^{-1} reaches a maximum at a temperature, T_{max}, where $\omega\tau = 1$. Loss peaks are often described in terms of T_{max}; the 53 K peak due to the Al-Na center is, perhaps, the classic case in quartz crystals. At high temperatures, the alkali ions thermally escape from the Al-site and drift in the shallow potential wells of the Z-axis channel. The number of free ions, n, is given by

$$n = (c/2)^{1/2} N_o \exp(-E_a / kT) \qquad (3)$$

where c is the mole fraction of Al traps, N_o is the number of SiO_2 molecules per unit volume, and E_a is the association energy between the ion and the trap. The square root dependence comes about because two defects

0-7803-1945-1/94 $3.00 © 1994 IEEE

are created: the mobile ion and the fixed ionized trap.[11] The resulting acoustic loss would be described by Eq. 1 with n replaced by Eq. 3 and, therefore, would contain two activation energies., E_a and E_b.

Experimental Procedure

Cultured quartz from a number of sources was used for these experiments. All of the crystal blanks were taken from the Z-growth region of the stone with the exception of an R-growth stone, HA-A, of high aluminum content. Table I summarizes the characteristics of the samples. EPR measurements were used to determine the aluminum concentration,[12] and low temperature infrared scans were also made on the samples. Plano-convex 5 Mhz 5th overtone resonator blanks were commercially fabricated from the bars.

TABLE I. Samples

Sample	Source	Al(ppm)	Mech. Q (Li)
X67	Rome Lab	0.2	2.6×10^{-6}
PQ-E	Sawyer	10-12	2.5×10^{-6}
HA-A	Russian	60-70	2.0×10^{-6}

After fabrication, the blanks were sodium, lithium, or hydrogen swept using our standard process.[13] The swept blanks have been well characterized through their use in a number of our earlier investigations.[14-17] The Na-swept HA-A blank had been irradiated to about 7 Mrad in 1987-88 and it was decided to investigate its high temperature acoustic loss during the anneal process. All of the other blanks were well annealed at temperatures greater than 450° C before they were used for this study.

Acoustic loss versus temperature measurements over the room temperature to 300° C range were made using our automated transmission method system.[18] Above 300° C, the acoustic loss was measured using our log-decrement system.[14] The resonator blanks were mounted in a gap-holder which was placed in an evacuated furnace for all of the measurements. The gap holder consisted of ceramic end caps with an AT-cut quartz ring spacer. The 15 mm diameter blanks were supported only on the outer 1 mm of their circumference. A thermocouple was attached to the gap-holder.

Results

Figures 1 and 2 show the acoustic loss versus temperature results for the 0.2 ppm Al sample and the 60 ppm Al sample respectively. These results represent the best and worst cases. The loss for the Li- and Na-compensated blanks increases rapidly with temperature above about 500 K for both Al concentrations.

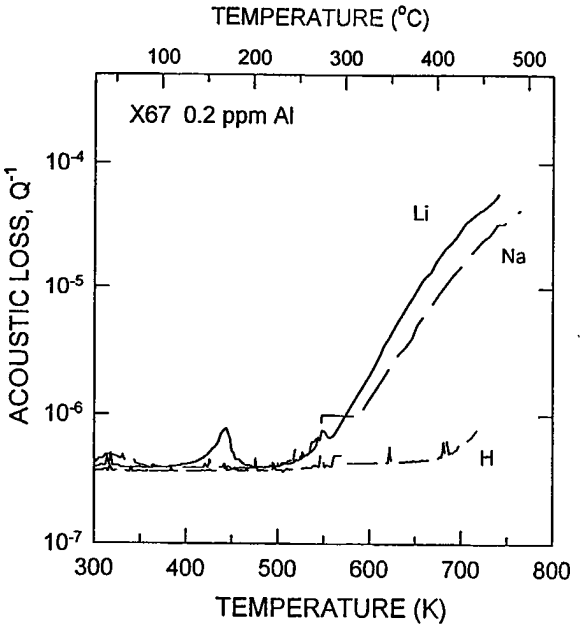


Fig. 1. The acoustic loss vs. temperature curves for the Li-, Na-, and H-compensated 0.2 ppm Al (X67) samples are shown.

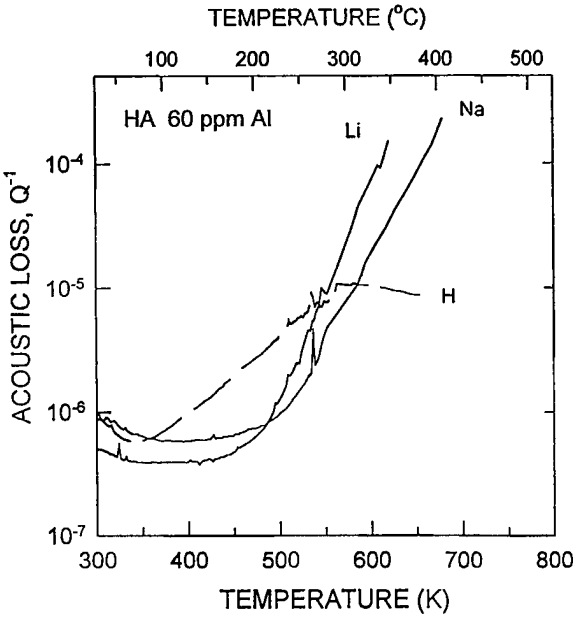


Fig. 2. The acoustic loss vs. temperature curves for the Li-, Na-, and H-compensated 60 ppm Al (HA-A) samples are shown.

As expected the results for the 60 ppm samples are larger than for the 0.2 ppm samples. As reported by Fraser[6] the high temperature loss is larger for Li-compensated samples than for the Na case. The swept or H-compensated 60 ppm sample shows a loss peak near 600 K while the loss in the swept 0.2 ppm sample is small and constant up to 700 K. The peak near 600 K is the same peak reported previously in swept quartz.[5,9,10] The results for the 12 ppm Al sample lie between those for the 0.2 and 60 ppm samples. The 12 ppm sample shows the 600 K loss peak.

Discussion

H-compensated

Figure 3 compares the acoustic loss versus temperature results for the three H-compensated (swept) samples.

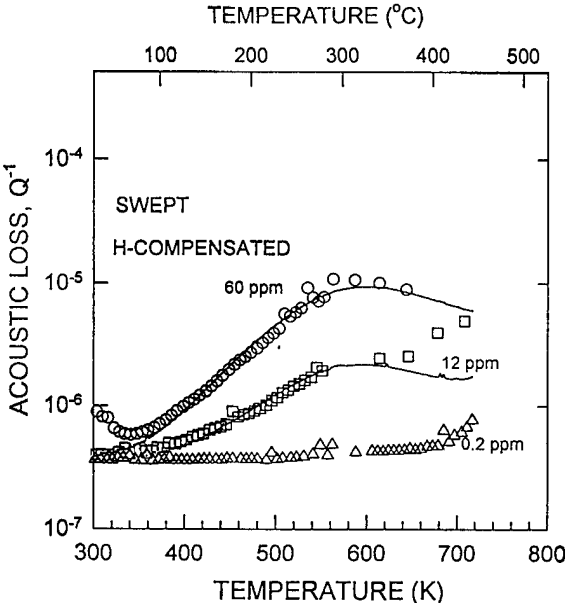


Figure 3. The acoustic loss curves are shown for the three H-compensated samples.

Below 650 K the results for the 0.2 ppm sample are essentially independent of temperature. They represent the loss due to the interaction with the thermal phonons. The higher Al concentration samples show a loss peak near 600 K. This peak is the same as the one reported earlier.[5,9,10] Sweeping removes the alkalis trapped by the aluminum and replaces them with hydrogen. Thus, the Al-M center is changed into the Al-OH center which is responsible for the 3367 and 3306 cm^{-1} infrared bands. No loss peaks below room temperature have been found that could be attributed to the Al-OH center. If we assume that the results for the 0.2 ppm sample represent the "intrinsic" acoustic loss then the difference between the data for the other two samples and that for the 0.2 ppm sample gives the height of the 600 K peak.

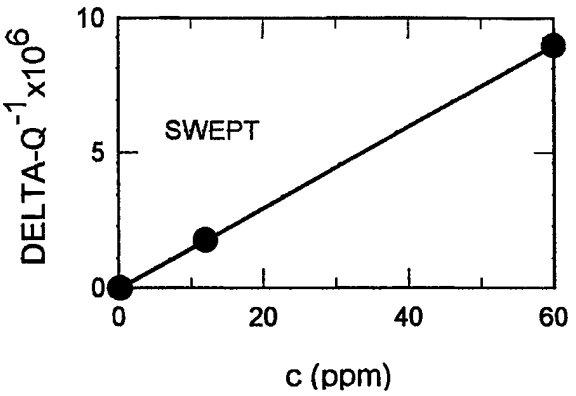


Fig. 4. The difference at 600 K between the curves for the 60 and 12 ppm Al concentration samples and the 0.2 ppm versus Al concentration is shown.

Figure 4 shows that this difference increases linearly with the aluminum concentration. Therefore, the 600 K loss peak in swept quartz must result from reorientation of the Al-OH center. Such relaxations are likely. The infrared bands broaden with increasing temperature due to thermal motion. Using low temperature electron irradiation, Sibley, *et al.*[19] were able to alter the relative population of hydrogen between long and short bond oxygens; the normal population recovered upon warming to room temperature. Thus, hydrogen is able to jump between the oxygens adjacent to the aluminum; the entire center relaxes during such a jump. The curves through the data points for the 12 and 60 ppm samples were calculated by adjusting the parameters in Eq. 1 and Eq. 2 to fit the data. Table 2 gives the parameters found from the calculation.

TABLE II. Al-OH acoustic loss peak parameters.

G (ppm Al)$^{-1}$	E (eV)	τ_o (sec)
$3x10^{-7}$	0.36	$3x10^{-11}$

There are two "versions" of the Al-OH center. One responsible for the 3367 cm^{-1} band and the other the 3306 cm^{-1} band. We assume that one corresponds to the hydrogen on the "short" oxygen bond and the other the hydrogen on the "long" oxygen bond. It is probable that

there are actually two overlapping loss peaks, both centered near 600 K. Lastly, the upturn at higher temperatures shown in Fig. 3 may be due to residual alkalis or to hydrogen moving in the channel.

Alkali compensted

The acoustic loss for the 0.2 ppm Li- and Na-compensated samples increases rapidly above 500 K. The small slowing of the increase at higher temperatures suggests that the T_{max} for the Li and Na loss peaks is well above 700 K. Thus, $\omega\tau >> 1$ and the $\omega^2\tau^2$ in the denominator of Eq. 1 dominates giving

$$\Delta Q^{-1} \approx nG\omega\tau_o \exp(-E_b / kT). \quad (4)$$

Eq. 3 gives the number of free ions. If we combine the multiplicative constants in Eqs. 3 and 4 we find that the loss, Q^{-1}, can be written as

$$Q^{-1} \approx Q_{back}^{-1} + C\exp(-E / kT) \quad (5)$$

where Q_{back}^{-1} is the background loss. The constant C is proportional to the square root of the aluminum concentration, c. The activation energy, E, contains both the association energy of the ion with the aluminum and the barrier height:

$$E = \frac{E_a}{2} + E_b. \quad (6)$$

This activation energy is the same as that found in ionic conductivity. Then for $T << T_{max}$ the loss increase due to the ions moving in the channel is thermally activated. As shown in Fig. 5 the data for the Li-compensated samples give a straight-line on a $\log(Q^{-1})$ versus 1000/T. The data for the 0.2 ppm aluminum sample deviate at the higher temperatures because the $T << T_{max}$ approximation is breaking down. The lines through the points represent a fit of Eq. 5 to the data. The fitting parameters are given in Table III.

TABLE III. Lithium high temperature loss parameters.

Al (ppm)	Q_{back}^{-1}	C (arb)	E (eV)
0.2	4.1×10^{-7}	2.8×10^{3}	1.1
12	4.1×10^{-7}	3.3×10^{4}	1.1
60	4.1×10^{-7}	1.2×10^{5}	1.1

The activation energy of 1.1 eV is higher than that reported by Fraser[6] but similar to the values found in sweeping Li into quartz.[15]

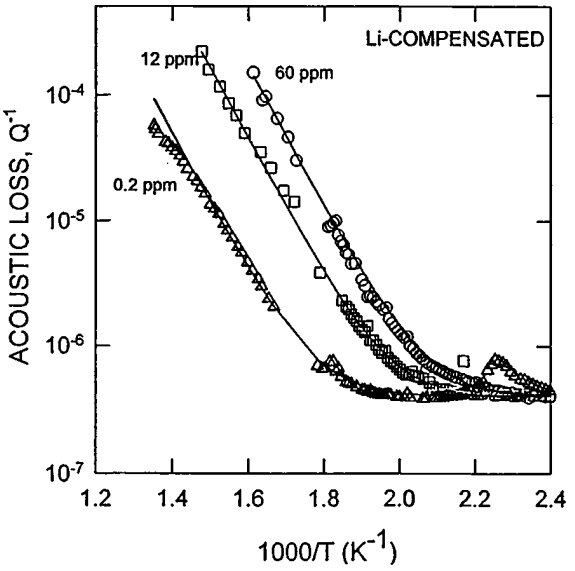


Figure 5. This log (Q^{-1}) vs 1000/T plot shows that the high temperature acoustic loss for the Li-compensated samples is thermally activated.

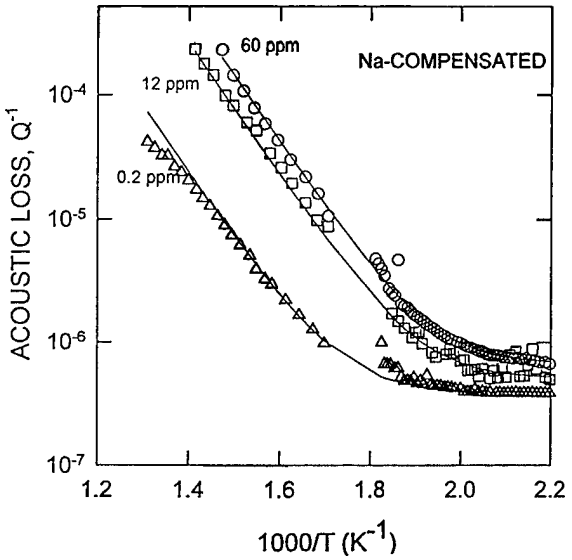


Figure 6. This $\log(Q^{-1})$ vs 1000/T plot shows that the high temperature acoustic loss of the Na-compensated samples is thermally activated.

Figure 6 shows that the high temperature acoustic loss for the three Na-compensated samples is also thermally activated. The lines through the points

represent a fit of Eq. 5 to the data. The fitting parameters are given in Table IV.

TABLE IV. Sodium loss fitting parameters.

Al (ppm)	Q_{back}^{-1}	C (arb)	E (eV)
0.2	3.9×10^{-7}	600	1.05
12	5.1×10^{-7}	6.5×10^{3}	1.05
60	6.7×10^{-7}	1.2×10^{4}	1.05

The activation energy is only slightly smaller than the 1.1 eV value found for lithium and is similar to our results from sweeping. In both cases, the activation energy contains both the association energy and the barrier height.

If we assume that the aluminum site is the source of the alkali ions then the constant C in Eq. 5 is proportional to the square root of the aluminum concentration. As shown by the log-log plot of C versus the aluminum concentration in Fig. 7, C does have the expected square-root dependence.

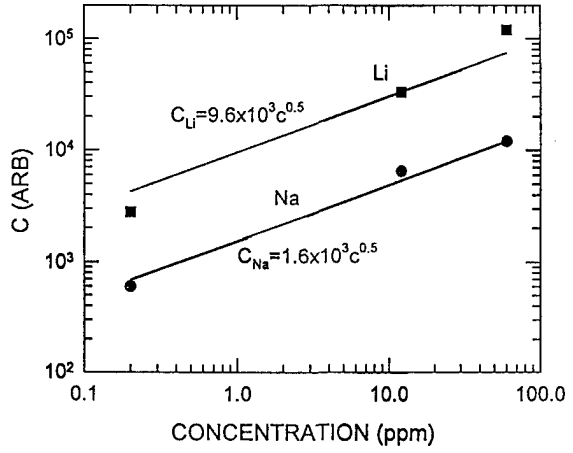


Figure 7. The constant C in Eq. 5 increases as the square-root of the aluminum concentration.

Anneal study of an irradiated Na-compensated sample

As part of an earlier study, the Na-compensated 60 ppm (HA-A) sample had been irradiated to about 7 Mrad[17] and it was decided to track its high temperature acoustic loss during the anneal process. Irradiation converts the Al-M^{+} centers into a mixture of Al-OH centers and Al-hole centers. Since HA-A quartz has a high aluminum content relative to its other defect content irradiation at room temperature produces a large concentration of Al-hole centers. The Al-hole center has an optical absorption band that peaks near 500 nm in the visible; this band is responsible for the color of smoky quartz. The Al-hole visible absoprtion band and the Al-OH infrared band allowed us to track both centers during the anneal. The anneal process was carried out as follows. First, the infrared Al-OH absorption was measured at 80 K and the visible range Al-hole absorption was measured at room temperature then the blank was place in the high-temperature acoustic loss system and data collected as it was heated to the desired anneal temperature. The infrared and visible optical scans were then repeated and the sample measured to the next anneal temperature which was increased by 25 K. The acoustic loss versus temperature curves for a selection of of the runs are shown in Fig. 8.

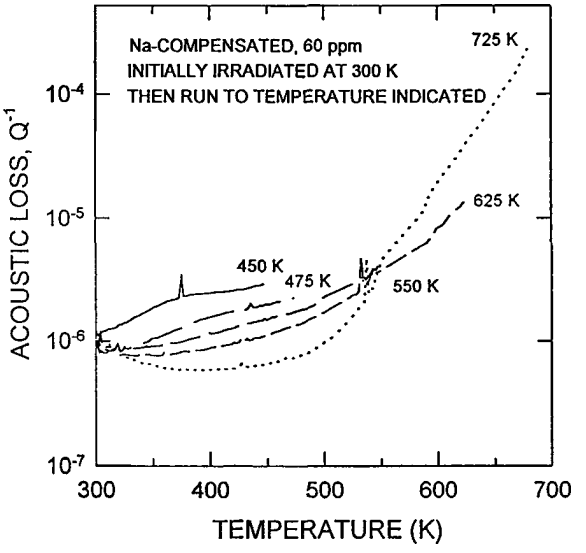


Figure 8. The acoustic loss data collected during the anneal runs on the irradiated 60 ppm sample are shown.

The data for the intial run to 450 K show that the Al-OH loss peak discussed above is present. As irradiated, this sample has about 12 ppm Al-OH centers and the loss is comparable to that of the 12 ppm H-compensated sample. Annealing is a dynamic process, and if we were to hold the irradiated sample at a fixed temperature such as 500 K we would see a steady decrease in loss until it reached the value given by the curve for the 725 K run.

Figure 9 shows the anneal of the Al-hole center and the Al-OH center as determined by the optical absorption scans. These anneal curves are in good agreement with our earlier results.[12,14] The Al-Na center recovery which was estimated from the optical results for the Al-hole and Al-OH center also agrees

with our early results. The Al-Na center does not fully recover until the sample is heated to above 700 K. The acoustic loss at 600 K as measured during the 625 K anneal run is considerably lower than that measured during the final 725 K run. This suggests that most of the alkalis participating in the thermally activated high-temperature acoustic loss are released from the aluminum site. The situation may be different in natural quartz since significant numbers of alkalis are also trapped at other point defects.

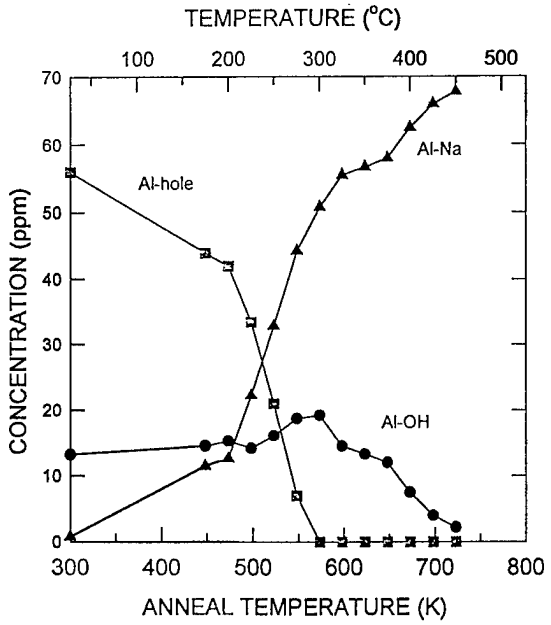


Figure 9. The anneal of the Al-OH and Al-hole centers as determined from the optical absorption is shown. The Al-Na recovery is estimated from these data.

Conclusions and Summary

The acoustic loss of Li-compensated, Na-compensated and H-compensated (swept) AT-cut resonator blanks with aluminum concentrations ranging from 0.2 ppm to 60 ppm has been measured from 300 K to nearly 700 K. A loss peak near 600 K was observed in the swept samples. This loss peak increases linearly with the aluminum concentration and, therefore, is due to the Al-OH center. Both Li- and Na-compensated samples showed a thermally activated loss at high temperatures. The high temperature loss was greater in the Li-compensated samples than in the Na-compensated samples. For both lithium and sodium the loss increased as the square-root of the aluminum concentration. This result shows that the aluminum site is the major source of the alkalis.

Acknowledgments

Ms. Ferris was supported in part by the National Science Foundation's RCMS program. The resonator blanks were fabricated by Piezo Crystal and K-W Mfg.

References

1. S. Kaitz, "Extended pressure and temperature operation of BT-cut pressure transducers," in *Proc. 38th Ann. Symp. on Freq. Control*, IEEE, pp. 245-251, May 1984.
2. J. A. Kusters and G. S. Kaitz, "Characteristics of natural, swept natural, cultured X- and Z-growth material in high temperature, high stress applications," *Proc. 39th Ann. Symp. on Freq. Control*, IEEE, pp. 223-229, May 1985.
3. R. A. Ward and E. P. EerNisse, "A reduced hysteresis, extended range quartz pressure transducer," *Proc. 41st Ann. Symp. on Freq. Control*, IEEE, pp. 344-349, May 1987.
4. E. P. EerNisse, "A resonator temperature transducer with no activity dips,", *Proc. 40th Ann. Symp. on Freq. Control*, IEEE, pp. 216-223, May 1986.
5. D. B. Fraser, "Impurities and anelasticity in single crystal cultured quartz," *Physical Acoustics*, W. P. Mason, Ed. New York, Academic, 1968, vol. V. chap. 2, pp. 59-110.
6. D. B. Fraser, "Anelastic effects of alkali ions in crystalline quartz," *J. Appl. Phys.*, vol. 26, pp. 2913-2918, 1964.
7. J. C. King, "Electrolysis of synthetic quartz; Effect upon resonator performance," in *IEE (London)*, vol. B109, Suppl. 22, pp. 295-301, 1962.
8. J. J. Martin, "Electrodiffusion (sweeping) of ions in quartz--a review," *IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 35, pp. 288-296, 1988.
9. D. R. Koehler, "Radiation-induced conductivity and high temperature Q changes in quartz resonators," *Proc. 35th Ann. Symp. on Freq. Control*, USAERADCOM, pp. 322-329, May 1981.
10. H. G. Lipson, A. Kahan, R. N. Brown, and F. K. Euler, "High temperature resonance loss and infrared characterization of quartz," *Proc. 35th Ann. Symp. on Freq. Control*, ASAERADCOM, pp. 329-334, May 1981.

11. A. D. Franklin, "Statistical thermodynamics of point defects in crystals," *Point Defects in Solids*, J. H. Crawford and L. M. Slifkin, Eds. New York, Plenum, 1972, vol. 1, ch. 1, pp. 1-103.
12. L. E. Halliburton, N. Koumvakalis, M. E. Markes, and J. J. Martin, "Radiation-effects in crystalline SiO_2: The role of aluminum,", *J. Appl. Phys.*, vol. 57, pp. 3565-3674, 1981.
13. A. R. Lopez, J. D. West, and J. J. Martin, "Electro-diffusion of ions in alpha-quartz," *Materials Research Society Symposia Proceedings,* vol. 60, pp. 451-457, 1986.
14. J. J. Martin, "Aluminum-related acoustic loss in AT-cut quartz crystals," *J. Appl. Phys.*, vol 56, pp. 2536-2540, 1984.
15. J. J. Martin, A. R. Lopez, A. F. Armington, and J. F. Balascio, "Evaluation of resonators fabricated from high quality quarzt," *Proc. 42nd Ann. Symp. on Freq. Control*," IEEE, pp. 162-168, 1988.
16. J. J. Martin, Ho B. Hwang, H. Bahadur, and G. A. Berman, "Room temperature acoustic loss in AT-cut quartz, crystals," *J. Appl. Phys.* vol. 65, pp. 4666-4671, 1989.
17. J. J. Martin, "Radiation-induced frequency offsets and acoustic loss in AT-cut quartz crystals," *J. Appl. Phys.*, vol. 68, pp. 5095-5104, 1990.
18. J. J. Martin, Ho B. Hwang and T. M. Wilson, "A radiation growth study of acoustic loss related defects in alpha-quartz," *Proc. 40th Ann. Symp. on Freq. Control*," IEEE, pp. 32-38, 1986.
19. W. A. Sibley, J. J. Martin, M. C. Wintersgill, and J. D. Brown, "The effect of radiation on the OH^- infrared absorption of quartz crystals," , *J. Appl. Phys.*, vol. 50, pp. 5449-5452, (1980).

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

GENERATION OF MICRO-DOMAINS IN AT-CUT QUARTZ BY THERMAL PROCESSING AND THE EFFECT ON RESONATOR MODES

L. Weisenbach, S.J. Martin, G.C. Frye, T.V. Bohuszewicz and D.H. Doughty

Sandia National Laboratories, Albuquerque, New Mexico, 87185

Abstract

The fabrication of acoustic sensors with sol-gel selective coatings requires the deposition of thin films on quartz resonators using solution chemistry techniques. Oxide films are spin-cast, then heat treated. A variety of film compositions are deposited, requiring a variety of firing schedules. Network analysis of the untreated AT-cut quartz devices revealed a resonant frequency of 5.0 MHz, corresponding to a pure thickness-shear mode resonance. For devices that were rapidly fired to 400°C and rapidly air cooled, network analysis showed the shear-mode response at 5.0 MHz disappeared, while a predominantly compressional mode response at 7.3 MHz emerged. Structural analysis explored the crystal structure changes induced by processing which resulted in this new mode.

Introduction

The effects of thermal processing on the resonant frequency and mode displacement of a quartz crystal micro balance (QCM) are investigated in this paper. The QCM is utilized in the fabrication of chemical sensors designed to detect metal ions in an aqueous solution. A sol-gel coating is deposited on the surface of a QCM, then fired in a furnace to form and consolidate an oxide film. The firing temperature affects the composition, structure and porosity of the sol-gel film and the optimum firing temperature varies with composition. For titanium dioxide films, a heat treatment to an elevated temperature (400°C) is required. In this paper, the effects and limits of thermal processing extremes on the resonator characteristics were explored.

Sol-gel coatings are spun onto the surfaces of QCMs to fabricate the sensors. After the films are spin-cast, the sensors are
fired for 20 minutes in a pre-heated furnace at temperatures of 150 or 400°C, depending upon the coating composition. After firing, the sample is removed and air cooled on an unheated block. In this processing scheme, the sensor structure was rapidly heated and then rapidly cooled, or quenched. This processing cycle is standard for the sol-gel coatings, which were originally designed for other applications. The quartz resonators used were designed to oscillate at 5.0 MHz. After the rapid processing to 400°C, the dominant resonant frequency of the quartz resonator had shifted from 5.0 MHz to 7.3 MHz. Sensors rapidly processed to 150°C still resonated at 5.0 MHz.

Experimental

The structure of the unfired and fired crystals was probed using X-ray diffraction rocking curve analysis performed with a Blake Industries Double Crystal X-Ray Diffraction System employing a four bounce Si monochromator with a resolution of eight arc seconds. A single Bragg reflection, corresponding to the 101 plane of the quartz crystal, was analyzed. Scanning Electron Microscopy (SEM) of the resonator surfaces was performed on a JEOL 6300V microscope. The surfaces were examined both with the electrode still on the surface, and also with the electrode removed. The resonators with the electrode removed were lightly coated with a thin layer of a gold palladium alloy to compensate for sample charging under the electron beam. The surface of the resonators were also examined using cross polarized light in an optical microscope.

Transmission electron microscopy was performed on plane-view and cross-sectional samples of the quenched crystal. The plane view sample was prepared by mechanically thinning the sample on one side to a thickness of approximately 10 μm, then the sample was thinned to perforation in an ion mill. The cross-sectional sample was prepared by making a sandwich by gluing the quartz crystal together with pieces of a Si wafer to obtain adequate thickness; then the cross-section was thinned and perforated as the plane view sample. The samples were lightly coated with a layer of amorphous carbon to avoid specimen charging under the electron beam. The analysis was performed with a JEM 2000FX microscope.

Results

The chemical sensors are designed to detect metal ions in solution. The resonance of the crystal is

0-7803-1945-1/94 $3.00 © 1994 IEEE

modified by the contacting liquid. Figure 1a and b show schematic drawings of an uncoated resonator and the shear displacement profile of the quartz and a contacting liquid. As illustrated in Figure 1b, the shear oscillation of the AT quartz causes entrainment of the contacting liquid. Changes in the liquid characteristics affect the resonant frequency and crystal damping[1].

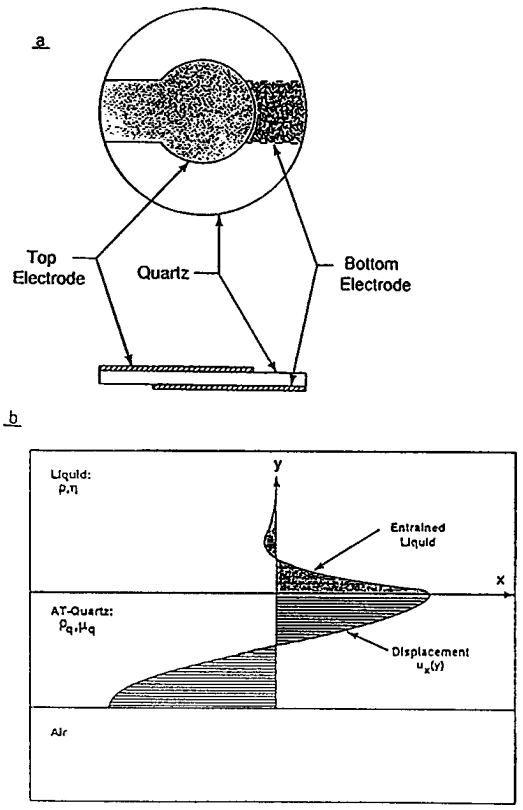


Figure 1. *Schematic of the resonator (a) and plot of the liquid entrainment induced by surface oscillation (b). Schematic view of the test case (c).*

Figure 1c is a schematic drawing of the sensor configuration used to measure the sensor response. The coated resonator is placed into the test case and sealed; an uncoated reference quartz crystal is also sealed in a test case. An ion-containing solution is passed over the oscillating sensor and reference crystal, and the frequency shifts of both are recorded. The resonant frequency of both crystals will be lowered due to liquid entrainment. A selective thin film deposited onto the resonator surface will preferentially react with certain ions in solution, altering the resonant frequency of the coated crystal by an additional, measurable amount. Figure 2a shows a plot of the typical response measured for a sol-gel coated device, oscillating at 5.0 MHz for various metal ions. The effect of the interaction of the ion solution with the thin film is to shift the resonant frequency by approximately 40 kHz.

To investigate the emergence of the 7.3 MHz mode, uncoated quartz resonators were fired for various firing times in a furnace preheated to 400°C. The magnitude of the electrical admittance (|Y|) vs. frequency for the series of samples is shown in Figure 3. The scale of the expanded ordinate is arbitrary to permit the curves to be included in the same plot. The only peak present in the spectrum of the unfired resonator is at 5.0 MHz. After 5 minutes in the pre-heated furnace, a small peak is present at 7.3 MHz, and the magnitude of the 5.0 MHz peak is decreased. The size of the 7.3 MHz peak increases with longer firing times, while the intensity of the 5.0 MHz peak decreases until, after a 20 minute bake, it is completely absent.

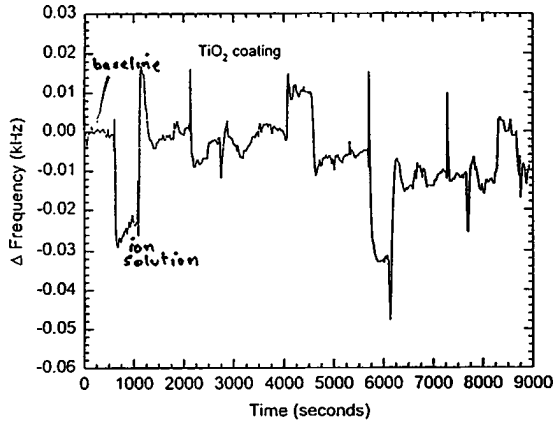


Figure 2. *Plot of the sensor response to metal ion solutions*

The reproducible and progressive nature of the emergence of the 7.3 MHz suggests the cause is a structural change in the crystal. The longer the firing time, the greater the extent of the damage. The motional resistance or dissipation in the QCM also increased with firing time, by two orders of magnitude. This result is obscured in the figure due to the arbitrarily expanded scale. The increased motional resistance is also an indication of crystal damage.

At the fundamental resonance, the velocity (v) of the acoustic wave propagating across the crystal thickness is related to the frequency, f, and crystal thickness, h, by the expression $v = 2hf$. When $f = 5.0$ MHz, the wave velocity is 3.3×10^5 cm/s, consistent with the shear wave velocity in quartz. When the resonant frequency is 7.3 MHz, the velocity is 4.8×10^5 cm/s, close to the compressional wave velocity. The

change in resonant frequency from 5.0 to 7.3 MHz indicates a conversion from pure shear displacement to dominantly compressional displacement.

The displacements associated with the untreated 5.0 MHz resonance and the treatment-induced 7.3 MHz resonance were investigated by measuring device response resulting from liquid contact. A series of alcohols, with varying viscosities and densities were placed in contact with the quartz crystal. Viscous entrainment of a contacting liquid (see Figure 1b) by shear displacement causes motional resistance changes proportional to $(\rho\eta)^{1/2}$, where ρ and η are liquid density and viscosity. The motional resistance arising from surface-normal displacement, if present, is proportional to ρ and independent of η.

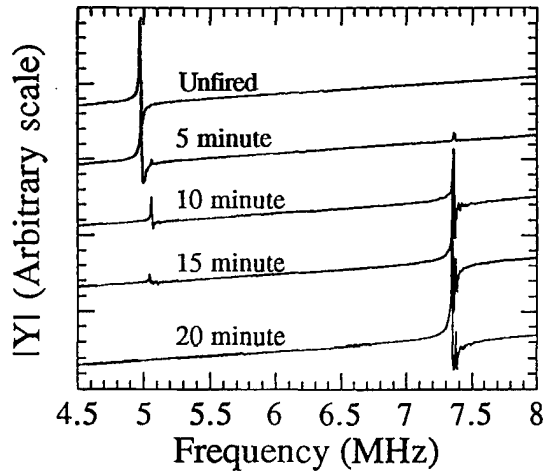


Figure 3. *The effect of firing time at 400°C on the resonant frequency of a quartz resonator.*

As shown in Figure 4, the motional resistance of the unfired (5.0 MHz) device is proportional to $(\rho\eta)^{1/2}$, indicating a pure shear surface displacement. The fired device (7.3 MHz) resistance also has a component proportional to $(\rho\eta)^{1/2}$, indicating some shear displacement; however, there is a large offset, indicating surface-normal displacement as well. Apparently, the heat treatment has changed the quartz properties so that the resultant displacement is a *combination* of shear and compressional displacements. Since the slope of the fired (7.3 MHz) is less than the unfired (5.0 MHz) device, the shear component has diminished.

To investigate the nature of the structural change suggested by the network analysis, X-ray rocking curve analysis was performed on the unbaked and fired and quenched QCMs. The QCMs were etched to remove the metal electrodes; the results are shown in Figures 5a and b. The pattern obtained for the unbaked sample (Figure 5a) shows a single sharp peak with a full width half max (FWHM) thickness of approximately 10 arc seconds or 0.1°, which is at the limit of the

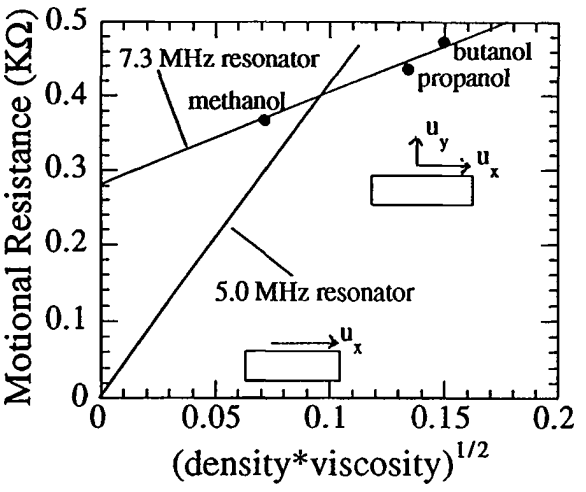


Figure 4. *Plot of the network analysis of the as-received (5.0 MHz) and quenched quartz resonators (7.3 MHz) measured with various alcohols in contact with the surface.*

instrument's resolution. The extreme narrowness of the peak indicates the high order and perfection of the as-received crystal. The incident beam energy is listed in the figure for each sample. The x-ray pattern shown in Figure 5b was acquired for the fired and quenched crystal and is very different from the pattern of the unfired crystal. A higher incident power was required to evaluate this sample and the number of counts reaching the detector was greatly reduced. The peak is much broader, and secondary peaks appear to be present in the pattern. The FWHM of the broadened peak is 360 arc seconds or 0.36°. The higher incident energy required, as well as the broadening of the peak, suggest the order of the single crystal structure has been disturbed by the rapid processing.

The above results document that rapid thermal processing causes crystal damage, even though neither SEM, TEM nor optical microscopy revealed surface cracks or twin formation in the fired and quenched sample. For comparison, quartz crystals were heated and cooled at varying rates, then characterized using a network analyzer. For the crystal which was heated and cooled at a slow rate, the resonant frequency remained at 5.0 MHz, with no evidence of crystal damage as indicated by a 7.3 MHz response. In contrast, samples heated slowly then air quenched, as well as samples heated rapidly then slowly cooled, had a 7.3 MHz response indicating the importance of heating rate, as

opposed to the temperature reached in causing crystal damage.

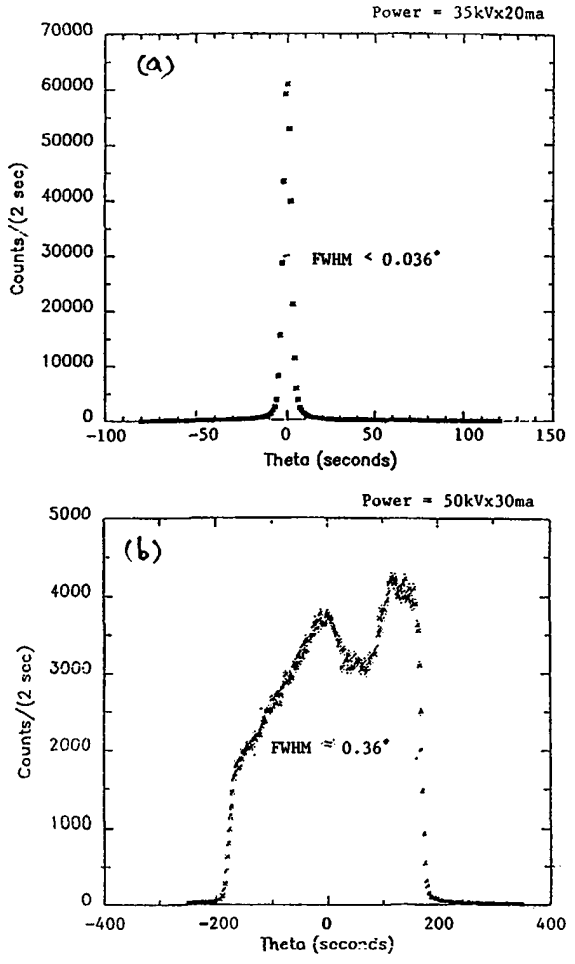


Figure 5. *X-ray rocking curve results for AT-cut quartz, as received (a) and fired at 400°C for 20 minutes (b).*

Discussion

Crystal damage is thought to arise from shear stresses induced by thermal gradients created across the crystal during rapid heating and/or cooling. In the heat treatment, the samples were placed individually into a preheated furnace while sitting on edge in a slot of an alumina boat. In this configuration, radiative heat transfer is the primary heating mechanism for the crystal.

Figure 6 schematically illustrates the thermal gradient created within the crystal and its effect. As the sample is heated or cooled, the surface will be heated or cooled faster than the center, creating a temperature differential (Fig. 6a), and a shear stress within the crystal results (Fig. 6b). The thermal expansion coefficient of AT cut quartz is 1.4×10^{-5}[2], thus the stress within the crystal due to the thermal gradient will be quite high. When the level of strain within the crystal exceeds the damage threshold, permanent changes (damage) to the crystal structure result (Fig. 6c). The damage threshold is the energy required to break bonds within the material, or cause a rotation of the tetrahedra. The pure shear mode of the quartz is a result of the perfect ordering of the crystal and the alignment of the particular plane. When this order is disrupted, a compressive mode emerges.

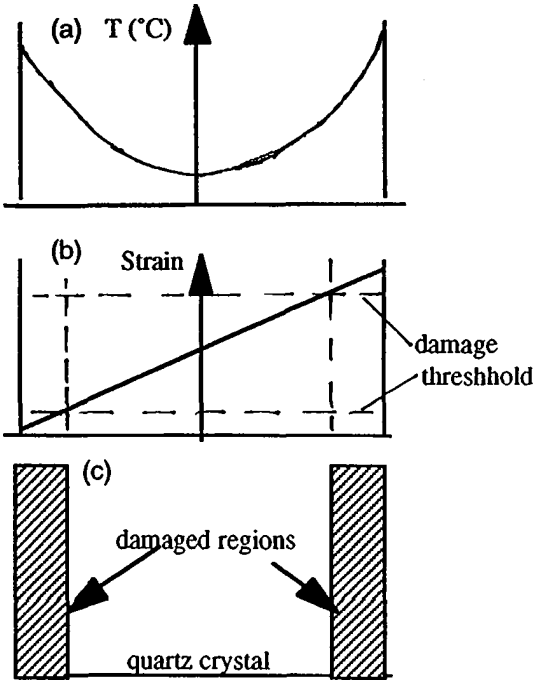


Figure 6. *Schematic representation of the effect of thermal strain on the quartz crystal.*

The liquid test results indicate that heat treatment has induced conversion from pure shear displacement to displacement that includes a surface-normal component. The reproducible change in resonant frequency from 5 to 7.3 MHz is consistent with the conversion from shear waves to higher velocity compressional waves. These results suggest that the crystal structure has been altered or damaged by the heat treatment.

Structural changes induced by heat treatment were examined using x-ray analysis. The broadening of the x-ray rocking curve peak can be explained by a number of phenomenon within the crystal structure, including the presence of micro-cracking at the surface, the existence of twins in the crystal and the presence of amorphous or disordered regions within the crystal. Analysis of the crystal surface by SEM, TEM and optical microscopy did not reveal the presence of micro-cracking. The existence of subsidiary peaks in the rocking curve pattern might be due to sub-grains within

the single crystal[3]. Sub-grains in the single crystal α-quartz may indicate the presence of twinned micro-domains. There are a number of twin types that occur in α-quartz. Dauphné twins are formed α-quartz when the orientation or tilt of the silica tetrahedra in the basal plane rotates in response to an applied stress[4]. This transition occurs without the breaking of bonds, and thus no reconstruction of the crystal is required. The evidence of the subsidiary peaks, however, is not conclusive, as various experimental conditions may cause subsidiary peaks in the pattern. The presence of twinning within the crystal could not be confirmed by either TEM (for micro-domains) or optical microscopy (for macro-domains).

Another mechanism that would relieve strain of the crystal is the formation of amorphous regions, the local disordering of the crystal structure. The presence of amorphous regions is difficult to verify. Evidence of amorphous regions (suggested by a broadening of the peaks in the pattern) is not discernible in an x-ray diffraction pattern until the volume percent of the amorphous material is greater than 5%. Amorphous regions within the crystal would be visible by viewing the lattice fringes of the crystal in the TEM, but this analysis could not be done due to the instability of the sample under the electron beam.

Though the exact nature of the crystal damage could not be ascertained, it is clearly demonstrated that rapid thermal processing of the quartz crystal in the sensor fabrication process has a deleterious effect on the structure of the crystal and leads to changes in the acoustic mode excited. Crystal damage due to thermally-induced stresses are avoided if the sensor are heated and cooled at slower rates.

Summary

The rapid heating and cooling of quartz resonators result in the generation of a surface normal displacement component. The emergence of the surface normal component is caused by damage to the crystal microstructure, which results from thermal gradients induced during the rapid thermal processing. Slower heating and cooling rates avoid the generation of this damage to the crystal structure, as indicated by the lack of the surface normal component.

Acknowledgments

The authors wish to thank Barry M. Ritchie for SEM analysis, Mike Eatough and John L. Reno for the x-ray diffraction analysis and discussions of the results obtained, Charles Hill for TEM analysis, L. Casaus for performing network analysis of the crystals and A.A. Russell. This work was performed at Sandia National Laboratories and supported by the U.S. Department of Energy under contract No. DE-AC04-94AL85000.

References

[1] S.J. Martin, V.E. Granstaff, and G.C. Frye, "Characterization of a Quartz Crystal Microbalance with Simultaneous Mass and Liquid Loading," Analytical Chemistry, vol. 63, pp. 2272-2281, 1991.

[2] A. Ballato, "Bulk Wave Resonators and Transducers," Short Course Notes, Ultrasonics Symposium, Tucson, AZ, 1992.

[3] B.D. Cullity, Elements of X-Ray Diffraction, 2nd Edition, Reading: Addison-Wesley Publishing Co., Inc., 1978, p. 277.

[4] C.S. Hurlbut, Jr., and C. Klein, Manual of Mineralogy, 19th Edition, New York: John Wiley & Sons, 1976, p.101.

DESIGN AND PERFORMANCE OF CMOS MICROMECHANICAL RESONATOR OSCILLATORS

Clark T.-C. Nguyen and Roger T. Howe

Berkeley Sensor & Actuator Center
Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory
University of California at Berkeley
Berkeley, California 94720

Abstract

A completely monolithic high-Q oscillator, fabricated via a combined CMOS plus surface micromachining technology, is described, for which the oscillation frequency is controlled by a polysilicon micromechanical resonator to achieve high stability. It is shown that the closed-loop, steady-state oscillation amplitude of this oscillator can be controlled through the dc-bias voltage applied to the capacitively driven and sensed μresonator. Brownian motion and mass loading phenomena are shown to have a greater influence on short-term stability in this micro-scale.

Introduction

Crystal oscillators are widely used to generate precision frequency standards for complex integrated circuits. With the current trend to include increasing amounts of a total system on a single silicon chip, designers have begun to include the oscillator function, without the crystal, on the silicon die. A fully monolithic high-Q oscillator, which includes the "crystal" element as well as sustaining CMOS electronics on-chip, is thus desirable.

Such an oscillator has recently been demonstrated [1], which utilizes a surface-micromachined, polycrystalline silicon resonator [2] frequency-setting element and CMOS electronics to sustain oscillation, all fabricated onto a single silicon chip (Fig. 1). The cross-section of the combined CMOS plus surface micromachining technology used to fabricate this oscillator [1,3] is shown in Fig. 2.

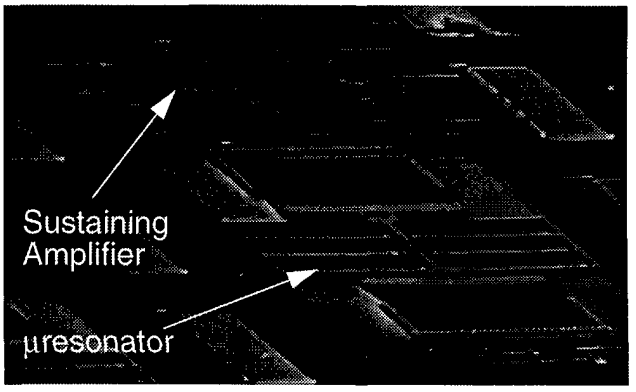


Fig. 1: SEM of the integrated CMOS μresonator oscillator.

With Q's of over 80,000 [4] under vacuum and center frequency temperature coefficients in the range of -10 ppm/°C (several times less with nulling techniques) [5], polysilicon micromechanical resonators can serve reasonably well as miniaturized substitutes for crystals in a variety of high-Q oscillator and filtering applications. As the high-Q element is miniaturized, however, such phenomena as Brownian motion and mass loading noise [6] begin to have greater influence on oscillator short-term stability and may limit the ultimate stability of micro-scale resonators unless design strategies which minimize these effects are implemented. After an initial focus on resonator transducer and oscillator design, this paper will address some of these performance limiting effects.

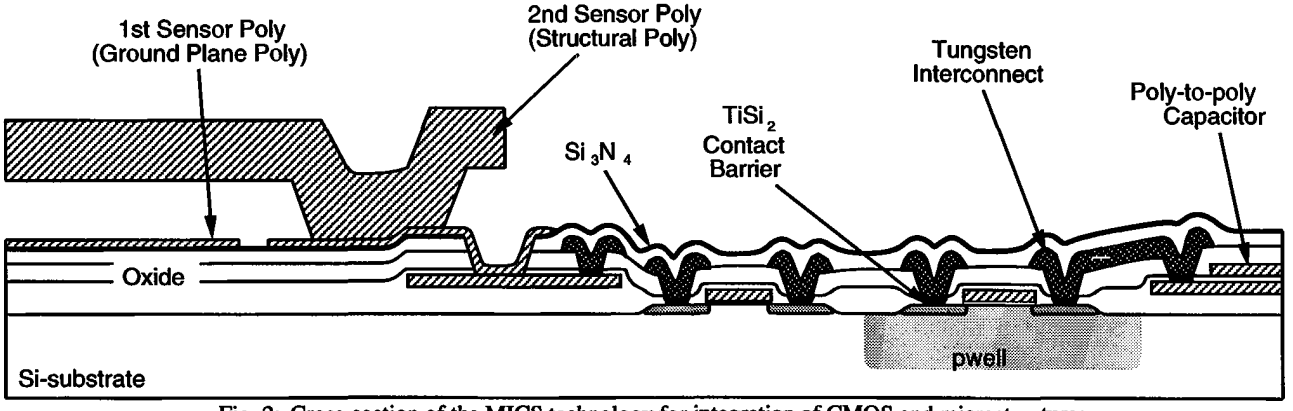


Fig. 2: Cross-section of the MICS technology for integration of CMOS and microstructures.

0-7803-1945-1/94 $3.00 © 1994 IEEE

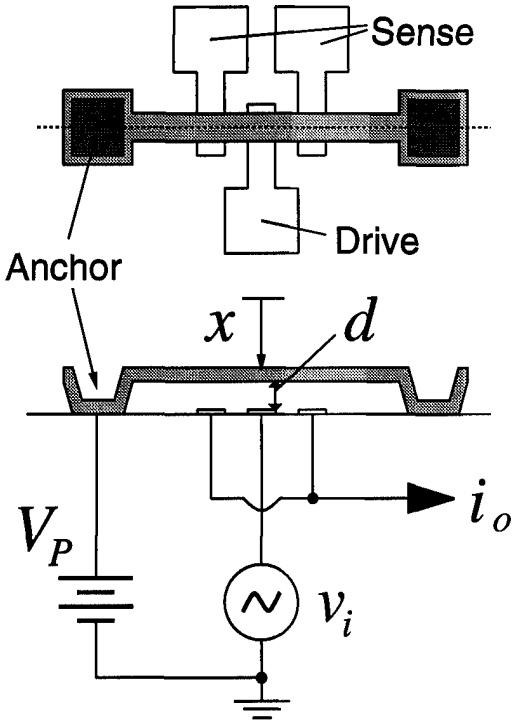


Fig. 3: Overhead and cross-sectional view of a vertical parallel-plate capacitively driven resonator with typical applied bias and excitation voltages.

Resonator Design

To simplify the task of integrating CMOS with micromechanics, capacitive excitation and detection is utilized for the μresonators in this work. A variety of capacitive topologies are available in this technology, and each will dictate the frequency tuning range and stability of the μresonator.

Figure 3 shows the cross-section of a parallel-plate capacitively driven μcantilever resonator in a typical bias and excitation configuration. [7]. Here, an ac voltage v_i electrostatically drives the cantilever. The dc-bias voltage V_P amplifies the resulting force components at the frequency of v_i. In this scheme, the beam-to-electrode capacitance is nonlinearly dependent upon beam displacement, and thus, the change in capacitance vs. displacement, $\partial C/\partial x$, is a strong function of displacement:

$$\frac{\partial C}{\partial x} = \frac{C_o}{d}\left(1-\frac{x}{d}\right)^2 , \qquad (1)$$

where C_o is the static beam-to-electrode capacitance. Using (1), the components of force f_d at the input frequency acting on the beam of Fig. 3 are

$$f_d|_{\omega_o} = \frac{1}{2}(V_P - v_i)^2 \frac{\partial C}{\partial x}\bigg|_{\omega_o} = -V_P\frac{C_o}{d}v_i + V_P^2\frac{C_o}{d^2}x \ . \qquad (2)$$

The second term of (2) implies an electrical spring constant, $k_e=V_P^2(C_o/d^2)$, which adds to the mechanical spring constant k_m and makes the center frequency f_o a function of the dc-bias voltage V_P:

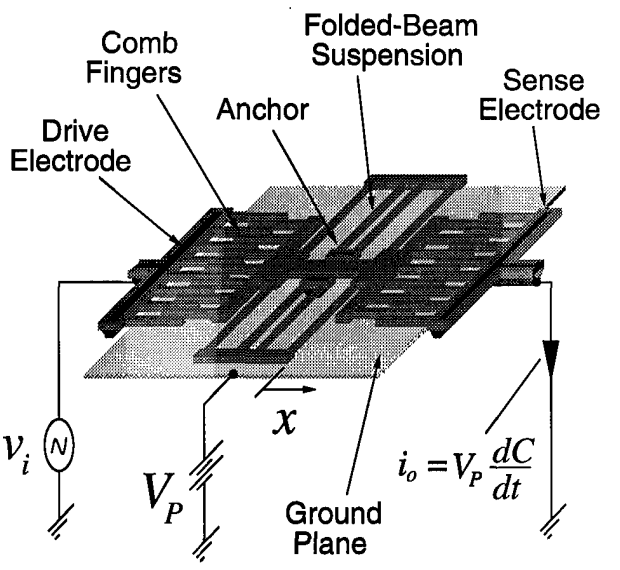


Fig. 4: Perspective view of a two-port, folded-beam, lateral comb-driven resonator with typical applied bias and excitation voltages. All areas of the resonator and electrodes are suspended 2 μm above the substrate, except for the darkly shaded areas, which are the anchor points.

$$f_o = f_o\left(1-\frac{V_P^2 C_o}{k_m d^2}\right)^{1/2} . \qquad (3)$$

The above provides a convenient means for voltage control of the center frequency, making parallel-plate driven resonators useful for VCO applications. A -3600 ppm/V fractional frequency change is typical for a 20 kHz μresonator with a nominal V_P=10 V. However, (3) also suggests that oscillators referenced to parallel-plate capacitively driven μresonators are less stable against power supply variations, due to electronic noise or temperature. If V_P=10 V is supplied by a Zener diode reference, which typically varies 250 mV over a 0°-100°C range, the corresponding fractional frequency variation for a 20 kHz μresonator is 864 ppm over this temperature range. If a bandgap reference is used (3 mV variation over a 0°-100°C range), the $\Delta f/f_o$ variation is 10 ppm.

To eliminate this component of frequency instability, the electrode-to-resonator capacitance must be made to vary linearly with resonator displacement. In this work, this is achieved by using interdigitated-comb finger drive and sense capacitors [2]. Figure 4 shows a μresonator which utilizes interdigitated-comb finger transduction in a typical bias and excitation configuration. The μresonator consists of a shuttle mass, with fingers on opposite sides, suspended 2 μm above the substrate by folded flexures, which are anchored to the substrate at two central points. The shuttle mass is free to move in the direction indicated, parallel to the plane of the silicon substrate. Folding the suspending beams as shown provides two main advantages: first, post-

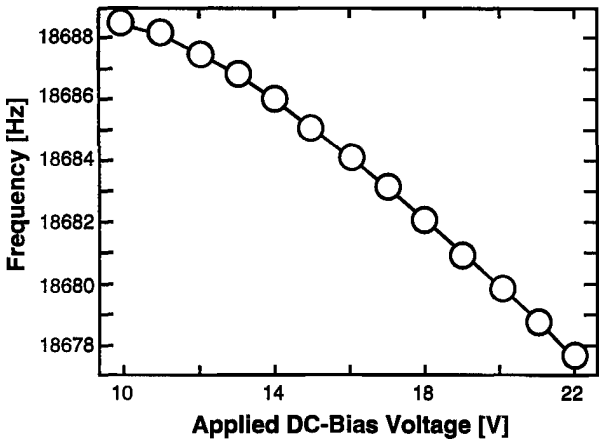


Fig. 5: Plot of center frequency vs. dc-bias voltage V_P for a comb-driven μresonator.

fabrication residual stress is relieved if all beams expand or contract by the same amount; and second, spring stiffening nonlinearity in the suspension is reduced, since the folding truss is free to move in a direction perpendicular to the resonator motion.

The drive and sense capacitors consist of overlap capacitance between the interdigitated shuttle and electrode fingers. As the shuttle moves, these capacitors vary linearly with displacement. Thus, $\partial C/\partial x$ is a constant, making the drive force f_d at the input frequency independent of x:

$$f_d|_{\omega_o} = V_P \frac{\partial C}{\partial x} v_i . \qquad (4)$$

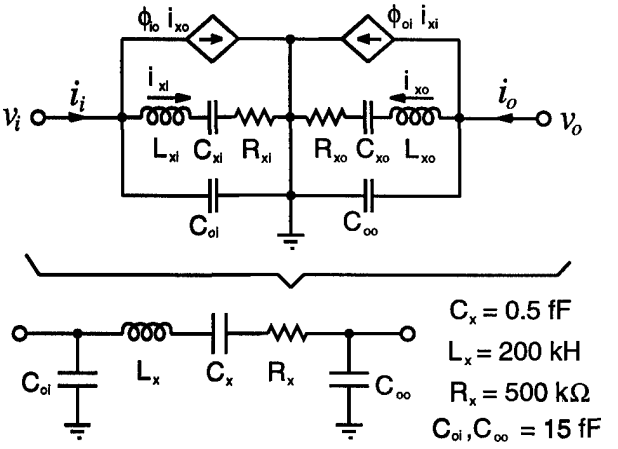
The electrical spring constant k_e is, thus, ideally nonexistent, and the resonator center frequency is independent of V_P.

For actual comb-driven μresonators, nonidealities do not permit absolute cancellation of k_e, and some variation of frequency with V_P is observed. Figure 5 shows a plot of center frequency vs. dc-bias for a micromachined comb-driven resonator. The frequency variation is about -54 ppm/V, corresponding to 14 ppm and 0.2 ppm fractional frequency variations over a 0°-100°C range for a Zener diode and a bandgap reference, respectively.

Oscillator Design

The equivalent circuit for the two-port μmechanical resonator of Fig. 4, shown transformed to an equivalent LCR representation, is presented in Fig. 6 [7,8]. Due to the use of weak capacitive electromechanical transduction, the motional element values are quite different from those for quartz crystal units (which typically have R_x=50Ω, C_x=0.04 pF, L_x=0.25 H), and this dictates differing strategies in the design of μresonator oscillators versus macroscopic crystal oscillators. The detailed, transistor-level circuit design and operation of this oscillator has already been discussed elsewhere [1]. The focus of the present discussion centers on issues of amplitude limiting.

Figure 7 shows a system-level schematic describing the basic architecture used for this oscillator. Since the



$$R_{xn} = \sqrt{km}\left[QV_P^2(\partial C/\partial x)_n^2\right]^{-1} \qquad C_{xn} = k^{-1}V_P^2(\partial C/\partial x)_n^2$$

$$L_{xn} = m\left[V_P^2(\partial C/\partial x)_n^2\right]^{-1} \qquad \phi_{mn} = \frac{(\partial C/\partial x)_m}{(\partial C/\partial x)_n}$$

Fig. 6: Equivalent circuit for a two-port μresonator showing the transformation to the more convenient LCR form. In the equations, k is the system spring constant and $(\partial C/\partial x)_n$ is the change in capacitance per displacement at port n of the μresonator.

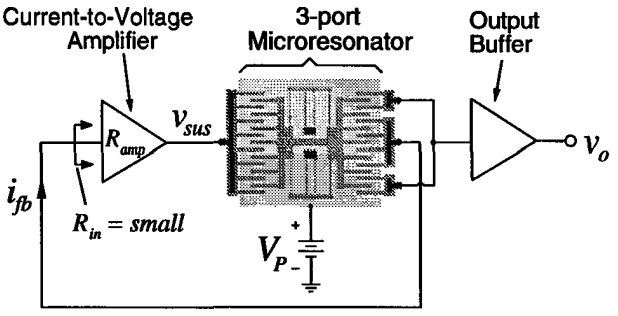


Fig. 7: System level schematic for the μresonator oscillator.

motional resistance of the μresonator is large (Fig. 6), a series resonant oscillator architecture is utilized to minimize Q-loading [1]. As shown, the system consists of a three-port micromechanical resonator, for which two ports are embedded in a positive feedback loop with a sustaining transresistance amplifier, while a third port is directed to an output buffer. The use of a third port effectively isolates the sustaining feedback loop from variations in output loading. Conceptually, the sustaining amplifier and μmechanical resonator comprise negative and positive resistances, respectively. During start-up, the negative resistance of the amplifier R_{amp} is larger in magnitude than the positive resistance of the resonator R_x, and oscillation results. Oscillation builds up until either some form of nonlinearity or a designed automatic-level control circuit alters either or both resistors so that, R_{amp}=R_x, at which point the oscillation amplitude limits.

For oscillators controlled by quartz crystals, the nonlinearity usually appears in the sustaining circuit, where transistors enter the triode region at large voltage amplitudes,

reducing effective device transconductances until the loop gain drops to unity. Limiting due to crystal nonlinearity is rare, since quartz crystal units display very little nonlinearity over normal oscillator operating voltage ranges [9].

On the other hand, even though comb-driven, folded-beam μmechanical resonators are only slightly less linear than crystals [1], limiting due to nonlinearity in flexural-mode μresonators is quite practical through adjustment of the dc-bias voltage V_P. As seen from the equations of Fig. 6, the values of the motional circuit elements representing the capacitively driven μmechanical resonator are strongly dependent upon the dc-bias voltage V_P applied to the resonator. In particular, the value of motional resistance R_x is inversely proportional to the square of V_P, and thus, it can be set to just under R_{amp} at the start of oscillation by proper selection of V_P. As oscillation builds up, stiffening nonlinearities in the resonator springs then increase the effective R_x of the μresonator until R_x=R_{amp}, when the loop gain equals one and the amplitude limits. The steady-state amplitude of oscillation is thus a function of the initial separation between R_x and R_{amp}, which is in turn a function of V_P.

To quantify this limiting process, we first write an expression describing the spring nonlinearity. Since at resonance, the force is amplified by the Q, we have

$$Qf = k_1 x + k_2 x^2 + k_3 x^3 + \ldots, \tag{5}$$

where k_1 is the small displacement system spring constant, and k_2 and k_3 model the spring nonlinearity. A series reversion then yields

$$x = b_1 Qf + b_2 (Qf)^2 + b_3 (Qf)^3 + \ldots, \tag{6}$$

where $b_1 = k_1^{-1}$, $b_2 = -k_2 k_1^{-3}$, and $b_3 = 2k_2^2 k_1^{-5} - k_3 k_1^{-4}$. Retaining only those components at resonance, the phasor form for displacement X is then given by

$$X = b_1 QF + \frac{3}{4} b_3 Q^3 F^3. \tag{7}$$

Inserting expressions for current I_x as a function of X and force F in terms of input voltage V_i into (7) [8], assuming a Duffing nonlinearity in the resonator springs (k_2=0), and differentiating, we have

$$\frac{\partial I_x}{\partial V_i} = \frac{Q}{k_1}\omega_o V_P^2\left(\frac{\partial C}{\partial x}\right)^2 - \frac{9}{4}\frac{k_3 Q^3}{k_1^4}\omega_o V_P^4\left(\frac{\partial C}{\partial x}\right)^4 V_i^2 = \frac{1}{R_{ss}} \tag{8}$$

Equation (8) shows that as the amplitude of oscillation V_i grows, the small-signal series resistance R_{ss} increases until it equals the transresistance of the sustaining amplifier, R_{amp}, at which point the loop gain is unity, and V_i settles at a steady-state value. From (8), the steady-state V_i is clearly controllable through the resonator dc-bias voltage V_P.

Fabrication

The enabling technology for the fully monolithic high-Q oscillator combines planar CMOS processing with surface micromachining [3]. The technologies are combined in a modular fashion, in which the CMOS processing and surface micromachining are done in separate process modules, with no intermixing of CMOS or micromachining steps. This Modular Integration of CMOS and microStructures (MICS) process has the advantage in that it allows the use of nearly any CMOS process with a variety of surface micromachining processes.

In order to avoid problems with microstructure topography, which commonly includes step heights of 2 to 3 μm, the CMOS module is fabricated before the microstructure module. Although this solves topography problems, it introduces constraints on the CMOS. Specifically, the metallization and contacts for the electronics must be able to survive post-CMOS micromachining processing with temperatures up to 835°C. Aluminum interconnect, the industry standard, cannot survive these temperatures. For this reason, tungsten with $TiSi_2$ contact barriers is used as interconnect for this process.

A cross-sectional outline of the MICS process sequence is presented in Fig. 8. The fabrication process begins with standard CMOS up to and including the contact cut for the first metallization (Fig. 8(a)). At this point, a thin film of titanium is sputter deposited onto the wafer surface, and then rapid-thermal annealed (RTA) for 30 seconds at 600°C in a nitrogen ambient to form $TiSi_2$ at points where titanium contacts silicon. Unreacted titanium is then etched away using a 3:1 NH_4OH:H_2O_2 solution, and another RTA is performed for 10 seconds at 1000°C. At this point, the cross-section appears as in Fig. 8(b). A 6000 Å film of tungsten is then sputter deposited and patterned to form the single-level interconnect (Fig. 8(c)). Subsequent low-pressure chemical vapor depositions (LPCVD) of 5000 Å of low-temperature oxide (LTO) and 1500 Å of silicon-rich nitride, at 450°C and 835°C, respectively, serve to passivate the metal (Fig. 8(d)).

Vias are then plasma etched through the nitride and underlying oxide to expose gate polysilicon runners that were formed during CMOS processing (Fig. 8(e)). These runners serve as an intermediate conductive level that joins the CMOS tungsten interconnect with the structural polysilicon interconnect. Direct contact between the first structural (ground plane) polysilicon and tungsten metal never occurs. This originally was a precaution to prevent contamination of the polysilicon deposition system by tungsten.

Next, 3000 Å of *in situ* phosphorous-doped LPCVD polysilicon is deposited at 610°C and patterned to define the interconnect and ground plane polysilicon for the microstructures (Fig. 8(f)). This is followed by a 2 μm LPCVD deposition (450°C) of phosphosilicate glass (PSG) that serves as a sacrificial layer to be removed when releasing the mechanical structures. Next, a contact cut in the PSG defines the anchor points for resonators and electrodes (Fig. 8(g)), and 2 μm of *in situ* phosphorous-doped LPCVD

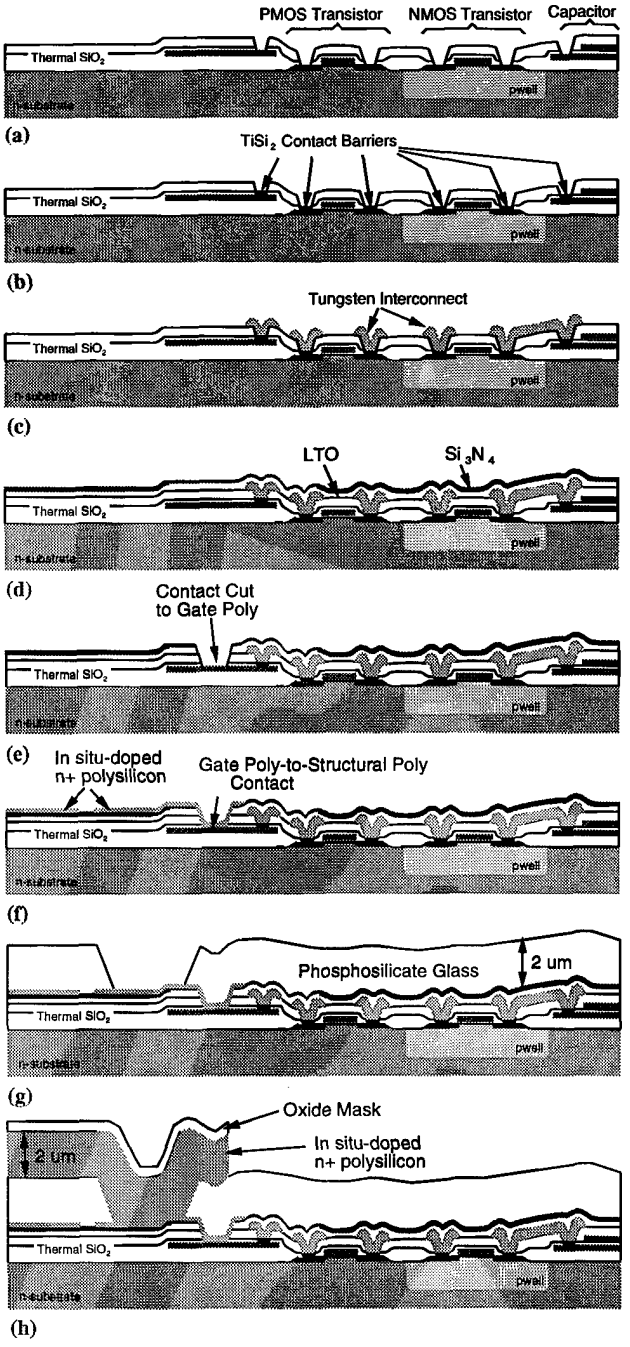


Fig. 8: Cross-sectional process flow for the CMOS plus microstructures fabrication technology. (a)-(c) constitute the metallization steps; (d)-(f) present the circuit-to-structure interface; and (g)-(h) show the micromachining steps. The final cross-section is presented Fig. 1.

polysilicon is deposited at 610°C to serve as the structural material.

The structural material is then capped with a 5000 Å film of LTO, which deposits conformally onto the polysilicon. A thin layer of photoresist is applied and patterned with a single mask that defines resonator geometries, including

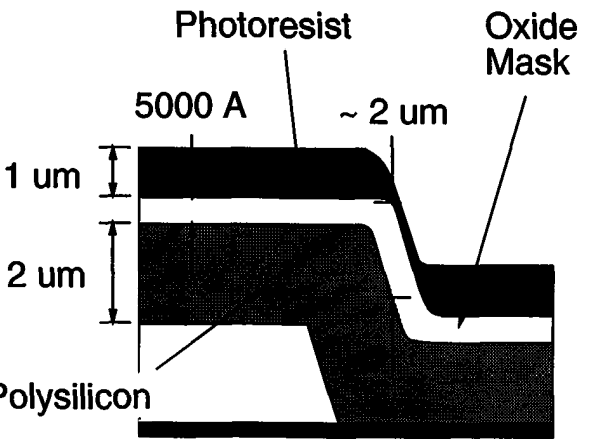


Fig. 9: Cross-section illustrating how a conformal oxide mask can reduce photoresist thickness requirements over large steps, thus, enhancing lithographic resolution. The oxide is much thicker vertically in the regions where photoresist is thin.

interdigitated-comb fingers. The oxide cap is first plasma etched using a CF_4/C_2F_6-based chemistry, which is very selective towards oxide, but which etches silicon very slowly. The oxide then serves as a hard mask for the ensuing plasma etch that patterns the structural polysilicon. This etch is done using a Cl_2-based chemistry that attacks silicon, but etches oxide very slowly. At this point, the structural polysilicon is fully patterned, and the cross-section of Fig. 8(h) results.

Use of this oxide mask greatly enhances the resolution with which finger gaps may be defined. Without the oxide mask, a double or triple layer of photoresist would be required to insure proper coverage of large steps, which occur around the anchors of the resonators and electrodes. With an oxide mask, however, the photoresist is no longer required to cover large steps, since now oxide protects the underlying polysilicon. During the oxide mask etch, the oxide along the step slopes may be attacked, since the photoresist is very thin along the steep slopes. However, as illustrated in Fig. 9, the vertical thickness of the conformal oxide layer is much larger along the these slopes, so oxide will remain to protect polysilicon even after long anisotropic overetches during oxide mask patterning. Thus, a much thinner film of photoresist may be used when an oxide mask is present, which leads to improved lithographic resolution. High resolution is extremely important for oscillator or signal processing applications of capacitively transduced resonators, since the degree of electromechanical coupling achievable via capacitive transduction is directly related to the gap spacings between interdigitated fingers.

Continuing with the process flow, a stress anneal is performed via RTA for 1 minute at 950°C, followed by a sequence of etches to expose the conductive backside of the silicon wafer. Finally, the wafer is dipped in 5:1 buffered hydrofluoric acid to remove the sacrificial PSG and free the microstructures. The wafers are dried using a supercritical carbon dioxide technique, which prevents sticking of the

structures to the substrate by eliminating surface tension forces during drying [13]. The final cross-section is shown in Fig. 1.

Although quite different from the traditional 1 hour at 1050°C furnace stress anneal [2], which cannot be used here due to the presence of CMOS electronics, the rapid thermal stress anneal performs comparably. However, due to the heavy phosphorous concentration in the polycrystalline silicon, the residual compressive stress in the polysilicon films of this work is still quite large and is difficult to anneal away. Thus, stress-relaxing designs, such as cantilevers or folded-beam resonators, are required for the current process. The aforementioned stress problems may be alleviated in the future by lowering the phosphorous content in the μresonators, or by using *in situ* boron-doped polysilicon. The latter solution has the additional advantage of substantially reducing the polysilicon deposition rate.

Short-term Frequency Stability

Superposed Electronic Noise

The phase noise power due to superposed electronic noise from the sustaining amplifier may be predicted theoretically using a procedure similar to that in [10]. Assuming a linear oscillator, and thus, neglecting $1/f$ mixed noise, the equation for the relative oscillator phase noise density N_{op} to carrier C power ratio at a deviation f_m from carrier frequency f_o is

$$\left.\frac{N_{op}}{C}\right|_{\delta f = f_m} = \frac{\left(\bar{i}_a^2/\Delta f\right)R_{amp}^2}{R_L R_{in} C}\frac{1}{8Q^2}\left(\frac{f_o}{f_m}\right)^2. \qquad (9)$$

where R_{in} and R_L are the input and load resistances, respectively.

For the case of an oscillator utilizing a micro-scale resonator, the noise current power $\bar{i}_a^2$ above must include both contributions from the amplifier and from Brownian motion of the miniaturized resonator. Using the thermal equilibrium arguments of [11], the Brownian noise displacement at resonance is given by

$$\frac{\bar{x}_n^2}{\Delta f} = \frac{4Q\sqrt{m}k_BT}{k^{3/2}} \propto \sqrt{m}\,, \qquad (10)$$

where k is the (small displacement) spring constant, k_B is the Boltzmann constant, and T is temperature. Equation (10) predicts that as a resonator is further miniaturized, the noise displacement power due to Brownian motion will decrease, which at first glance would seem to improve noise performance. However, capacitive transduction detects the velocity of the resonator, not its displacement. Thus, the output current i_x of the μresonator corresponds to velocity, and the effective noise current due to Brownian motion is given by

$$\frac{\bar{i}_x^2}{\Delta f} = \frac{4k_BT}{R_x} \propto \frac{1}{\sqrt{m}}\,, \qquad (11)$$

which increases as resonator size decreases. For the oscillator design of this work, the electronic noise contribution from Brownian motion is usually equal to or less than that from the sustaining amplifier.

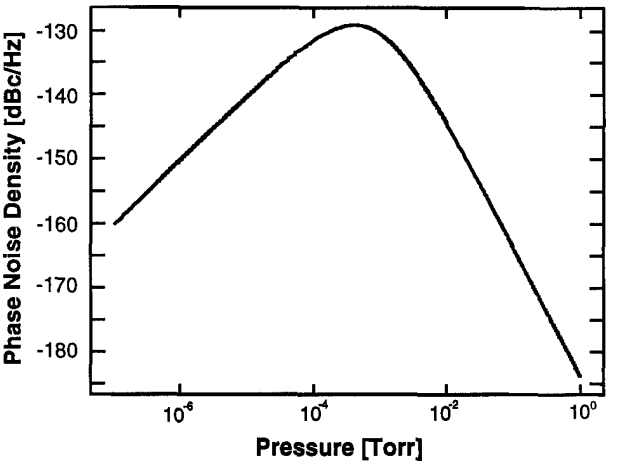


Fig. 10:Predicted plot of mass loading-derived phase noise density 100 Hz off the carrier vs. pressure for a 14.4 kHz μresonator oscillator in an ambient of gas molecules with the molecular weight of nitrogen.

Mass Loading Noise

In addition to superposed electronic noise, any physical phenomenon which causes instantaneous frequency deviations in the resonator will contribute to the total phase noise power. Given that the typical mass of a μmechanical resonator is on the order of 10^{-11} kg, mass loading noise is expected to make a sizable contribution. Mass loading noise [6] arises from instantaneous differences in the rates of adsorption and desorption of contaminant molecules to and from the resonator surface, which cause mass fluctuations, and consequently, frequency fluctuations. Some of the factors which determine the magnitude of mass loading noise include the adsorption/desorption rate of contaminant molecules, contaminant molecule size and weight, pressure, and temperature.

An expression which estimates the phase noise density due to mass loading noise has been proposed by Yong and Vig [6], and is repeated, here:

$$S_\phi(f) = \frac{8r_o r_1 (\Delta f)^2/N}{(r_o + r_1)^3 + 4\pi^2 f^2 (r_o + r_1)} \cdot \frac{1}{f^2}, \qquad (12)$$

where r_o is the mean rate of arrival of contaminant molecules at a resonator site, r_i is the desorption rate of molecules from the surface, and N is the total number of sites on the resonator surface at which adsorption or desorption can occur. Equation (12) assumes a sticking probability of one if an adsorption site is uncontaminated and zero if contaminated, so the magnitude of phase noise predicted will be higher or lower than the actual value, depending upon the actual sticking probabilities for the molecules involved. The qualitative trends predicted by (12), however, are useful.

Using (12) and accounting for the pressure dependence of r_o [6], the phase noise density due to mass loading for the 14.4 kHz μresonator of Fig. 1 can be plotted as a function of

pressure. Figure 10 presents such a plot for phase noise at a 100 Hz deviation from the carrier, where a contaminant molecule with the molecular weight of nitrogen has been assumed. The desorption energy was assumed to be 12 kcal/mol. The mass loading-derived phase noise density is largest at an intermediate value of pressure and smallest at the higher and lower pressures. At the peak (around 0.5 mTorr), the predicted phase noise density is -130 dBc/Hz, which is higher than that predicted for macroscopic quartz resonators of comparable frequency.

As the μresonator frequency increases, its mass generally decreases, and the phase noise contribution from mass loading is expected to become even more significant. The mass of a 10.7 MHz tuning fork μresonator is on the order of only 10^{-12} kg, and for this design the predicted phase noise density due to mass loading at the peak of the pressure curve is −105 dBc/Hz for 100 Hz deviation from the carrier and −142 dBc/Hz for 5 kHz deviation from the carrier.

For μmechanical resonators with Q's in the range of 50,000 to 500,000, viscous gas damping [12] ceases to be the dominant energy dissipation mechanism at pressures in the range of 0.1 to 1 mTorr, where intrinsic material damping mechanisms become dominant, and the Q of the resonator is maximized. As seen from Fig. 10, the phase noise density due to mass loading may still be large at this pressure value, and even lower pressures are required to alleviate this noise source. Thus, by setting an upper limit on operation pressure given a required phase noise density level, mass loading phenomena may ultimately dictate the design of μmechanical resonator oscillators.

On-chip vacuum encapsulation techniques have been previously investigated which provide vacuums with pressures below 300 mTorr [16], or perhaps better [14,15]. Encapsulation strategies which use gettering elements to remove residual gases may potentially provide the even lower pressure ranges (10^{-7} to 10^{-6} Torr) requested by Fig. 10.

Experimental Results

The fabricated oscillator was bonded up in a dual-in-line package and tested under a variable pressure vacuum probe station. Figure 11 shows a typical oscilloscope plot for a 16.5 kHz version of this oscillator. The amplitude of oscillation was visibly controllable through adjustment of V_P, consistent with the discussion of Section III. A measured plot of steady-state oscillator output voltage versus μresonator dc-bias is presented in Fig. 12. Over most of the range of V_P, the oscillation was visibly and measurably clean. However, over about a 5 V range some chaotic behavior of the resonator was visible under microscope. This indicates that in a finite range of dc-bias voltage, the oscillator may be exhibiting second-order nonautonomous behavior when limiting through resonator nonlinearity [17]. This phenomenon is currently under investigation.

In addition, accurate measurements of the phase noise of this oscillator are in progress.

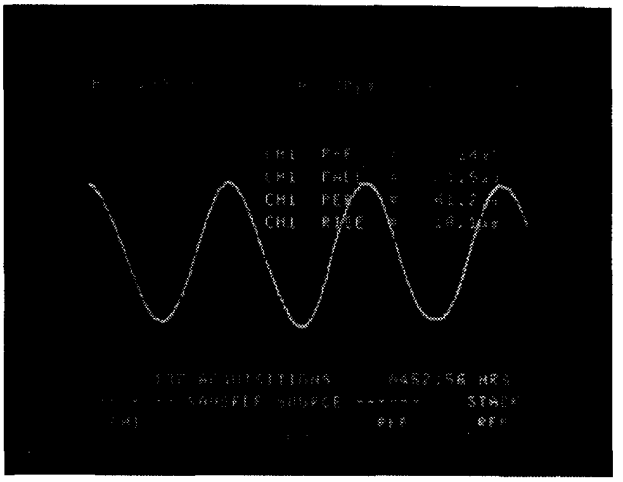


Fig. 11:Oscilloscope waveform for a μresonator oscillator.

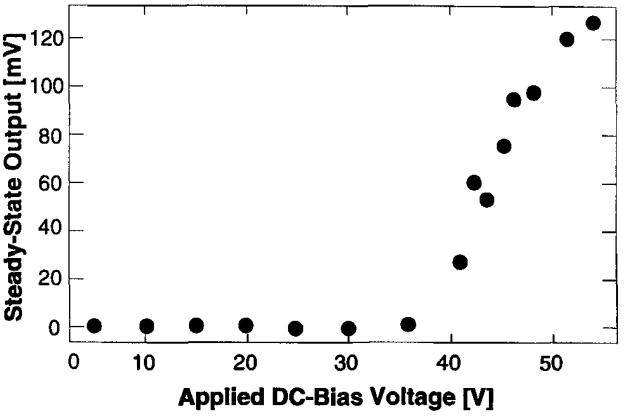


Fig. 12:Measured dependence of oscillator steady-state voltage output amplitude as a function of μresonator dc-bias.

Conclusions

Completely monolithic, highly stable, high-Q oscillators utilizing surface-micromachined polysilicon mechanical resonators have been designed, fabricated, and tested with particular attention to amplitude control and phase noise performance. Due to the novelty of the process and the devices, conservative measures were taken for the designs, and oscillators up to only 100 kHz were fabricated. Designs up to a few megaHertz are feasible using folded-beam resonator designs, and higher frequencies (tens of MHz) should be feasible using more advanced designs aimed at maximizing resonator quality factor, which may otherwise degrade with increasing frequency. Both material and architectural improvements should be possible to increase μresonator Q.

The consequences associated with miniaturization of high-Q elements were addressed via this oscillator. Brownian motion and mass loading were identified as phenomena which become increasingly important contributors to phase noise as resonator dimensions shrink. According to theory, mass loading-induced phase noise can be substantially

reduced by operating the miniature μmechanical resonator under very low pressures. For this reason, integrated vacuum encapsulation techniques may play an important role in the future.

Acknowledgements. The authors would like to thank Shenqing Fang for assistance in CMOS fabrication, as well as Katalin Voros and the staff of the Berkeley Microfabrication Laboratory for process support. In addition, we are grateful for discussions with John Vig of the U.S. Army Electronics Technology and Devices Laboratory, whose insights on mass loading contributed much to this work. This research was supported by the Berkeley Sensor & Actuator Center (BSAC).

References

[1] C. T.-C. Nguyen and R. T. Howe, "CMOS micromechanical resonator oscillator," *Technical Digest*, IEEE International Electron Devices Meeting, Washington, D.C., Dec. 5-8, 1993, pp. 199-202.

[2] W. C. Tang, C. T.-C. Nguyen, and R. T. Howe, "Laterally driven polysilicon resonant microstructures," *Sensors and Actuators*, vol. 20, pp. 25-32, 1989.

[3] W. Yun, R. T. Howe, and P. R. Gray, "Surface micromachined, digitally force-balanced accelerometer with integrated CMOS detection circuitry, *Technical Digest*, IEEE Solid-State Sensor & Actuator Workshop, Hilton Head Island, South Carolina, June 22-25, 1992, pp. 126-131.

[4] C. T.-C. Nguyen and R. T. Howe, "Quality factor control for micromechanical resonators," *Technical Digest*, IEEE International Electron Devices Meeting, San Francisco, California, Dec. 13-16, 1992, pp. 505-508.

[5] C. T.-C. Nguyen and R. T. Howe, "Microresonator frequency control and stabilization using an integrated micro oven," *Technical Digest*, 7th International Conference on Solid-State Sensors and Actuators (Transducers'93), Yokohama, Japan, June 1993, pp. 1040-1043.

[6] Y. K. Yong and J. R. Vig, "Resonator surface contamination—a cause of frequency fluctuations?" *IEEE Trans. Ultrason. Ferroelec. Freq. Contr.*, vol. 36, no. 4, pp. 452-458, March 1989.

[7] R. T. Howe and R. S. Muller, "Resonant microbridge vapor sensor," *IEEE Trans. Electron Devices*, vol. ED-33, pp. 499-506, 1986.

[8] C. T.-C. Nguyen, "Electromechanical characterization of microresonators for circuit applications," M.S. Report, Dept. of Electrical Engineering and Computer Sciences, University of California at Berkeley, April 1991.

[9] F. L. Walls and J.-J. Gagnepain, "Environmental sensitivities of quartz oscillators," *IEEE Trans. Ultrason. Ferroelec. Freq. Contr.*, vol. 39, no. 2, pp. 241-249, March 1992.

[10] W. P. Robins, *Phase Noise in Signal Sources*. London: Peter Peregrinus Ltd., 1982.

[11] T. B. Gabrielson, "Mechanical-thermal noise in micromachined acoustic and vibration sensors, *IEEE Trans. Electron Devices*, vol. 4, no. 5, pp. 903-909, May 1993.

[12] R. A. Buser, "Theoretical and experimental investigations of silicon single crystal resonant structures," Ph.D. Thesis, Inst. of Microtechnology, University of Neuchatel, CH-2000 Neuchatel, Switzerland, July 1989.

[13] G. T. Mulhern, D. S. Soane, and R. T. Howe, "Supercritical carbon dioxide drying of microstructures," 7th International Conference on Solid-State Sensors and Actuators (Transducers'93), Yokohama, Japan, June 1993, pp. 296-299.

[14] H. Guckel, *et al*, "The application of fine-grained tensile polysilicon to mechanically resonant transducers," *Sensors and Actuators*, **A21-23**, pp. 346-351 (1990).

[15] H. Guckel, *et al*, "Polysilicon resonant microbeam technology for high performance sensor applications," *Technical Digest*, IEEE Solid-State Sensor & Actuator Workshop, Hilton Head Island, South Carolina, June 22-25, 1992., pp. 153-156.

[16] L. lin, *et al*, "Vacuum-encapsulated lateral microresonators," *Technical Digest*, 7th International Conference on Solid-State Sensors and Actuators (Transducers'93), Yokohama, Japan, June 1993, pp. 270-273.

[17] T. S. Parker and L. O. Chua, "Chaos: a tutorial for engineers," *Proceedings of the IEEE*, vol. 75, no. 8, pp. 982-1008.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

MICROMACHINED THIN FILM BULK ACOUSTIC RESONATORS

R. Ruby, P. Merchant
Hewlett-Packard Laboratories,
3500 Deer Creek Rd.
Palo Alto, CA. 94303

Abstract-- We have fabricated thin film bulk acoustic resonators (FBARS) with measured Q's of over 1000 and resonant frequencies as low as 1.5 GHz and as high as 7.5 GHz. The device, as currently fabricated, consists of the piezoelectric material aluminum nitride (AlN) sandwiched between electrodes all of which lie on a thin low-stress silicon nitride (Si_xN_y) membrane. Integrated on the membrane are small microheaters for frequency tuning and/or temperature stabilization. We have observed frequency shifts of 50 to 80 ppm per degree C depending on relative material thicknesses. Maximum temperature excursions over 580 C could be achieved using the microheaters We have also observed frequency shifts of 5 to 10 ppm per volt depending on harmonic.

Introduction

Recently, there has been a growing excitement in rf frequency applications motivated in part by the "wireless" bandwagon. Applications that are beginning to impact the market are inexpensive GPS transponders, integrated radios with palmtop computers, and of course cellular telephony.

Thin Film Bulk Acoustic Resonators, or FBARS, hold the promise of being extremely small and ultimately integrable onto active substrates. FBARs can be made as small as a few ten's of microns on a side (and on the order of 1 micron thick) to a few 100's of microns on a side. This gives 5 to 8 orders of magnitude decrease in volume as compared to ceramic resonators and metal cavity resonators. Furthermore, there appears to be no intrinsic reason why these devices cannot be integrated with electronic circuitries. One can envision a simple low-temperature process that can accept any substrate (for example, a wafer with BiCMOS circuits) and allow FBARS to be integrated on top.

Process Overview & Highlights

The FBAR device is supported on a thin low-stress Si_xN_y membrane that lies over a back-etched cavity in the silicon wafer. The silicon nitride membrane is about 0.5 μm thick and lies in the acoustic path of the FBAR element. We use aluminum nitride (AlN) as our piezoelectric material.

We incorporated several innovations into our process including a proprietary silicon etch process to insure high yields of testable wafers using standard processing techniques. One drawback to our etch process is that it sometimes leaves a silicon residue (looks like "cigarette ash") on the backside of the silicon nitride membrane.

The interconnects from pad to FBAR device were not optimized in our first-generation mask set and introduced 20 to 40 ohms of parasitic resistance in series with the FBAR device. This defect was corrected in the next mask revision and results from both mask sets will be reviewed.

Figure 1 shows a cross-section of an FBAR device.

Aluminum Nitride Deposition and Patterning

The aluminum nitride films were reactively sputter deposited using a load locked system with a base pressure in the low 1×10^{-8} Torr range. The two inch diameter (99.999% pure) aluminum target was located two inches from the substrate. The rf power was 400 W with a deposition rate of about 200 A/min. The substrate was not heated and typically stayed below 100 C. The pressure was 5 mTorr and the ratio of nitrogen to argon was about 1:1.

The deposited films were characterized by ellipsometry (every film) and X-ray diffraction (some films). Typically, 20 to 60 ellipsometric measurements across the 4" dia. wafer were taken. Since the film uniformity resulting from using such a small target varied by about 25% from the center to the edge, these measurements allowed one to plot the values of ψ vs Δ (a measure of the ellipticity of reflected light of a circularly polarized light incident on a substrate), over the full range of the ellipsometry curve giving a very accurate fit to the film index of refraction, n_f, as well as the underlying electrode's optical constants, n_s and k_s. The AlN film measured index of refraction is 2.065 with ± 0.002 uncertainty. This number was determined from multiple measurements on 19 wafers deposited under a variety of conditions.

X-ray diffraction was used to determine the phase of the materials (θ-2θ scans) and the grain orienta-

0-7803-1945-1/94 $3.00 © 1994 IEEE

tions (rocking curves). One sees only the AlN (0002) peak in the θ-2θ scans. No other phases or orientations of this material were found. The rocking curve for the AlN peak shows a FWHM of about 3.3 degrees. An SEM cross-section of the same film from which the X-ray data was taken shows columnar structure and dense packing of the grains, consistent with the X-ray data.

Some investigations of the sputtering parameter space were made prior to choosing the sputtering conditions described above. As a starting point, we chose to operate in the rf mode since the target self-bias tracks the film deposition rate. This allowed us to use the self-bias parameter to track the hysteresis typically found in reactive sputtering processes in a simple way as a direct readout from the power supply. We began by working just below the knee in this curve where the target has been nearly fully poisoned by the reactive N_2. We have been able to achieve films with low stresses (< ± 2 x 10^9 dynes/cm2).

As a matter of practical concern, we have never seen a short (due to pinholes or otherwise) in any of our devices, and we have measured over 300 devices. Also, we have used a sloped profile etch when patterning and opening up the via holes through the AlN. So far, we have never seen an open due to insufficient metal coverage over a via hole.

Measurement Technique

All FBAR devices were measured in reflection (S_{11}), using an HP 8510C network analyzer and a Cascade microwave probe. The probe and network analyzer were calibrated using a Cascade calibration standard. Typical data, in our case, consists of the real and imaginary terms of S_{11} for 801 points over some frequency span. The 8510 was controlled by a Basic program running under RMB-UX (Rocky Mountain Basic on UNIX). The data, extracted from the 8510 and stored as an ASCII file on a local workstation, was then moved over, via the network, to an HP 735 workstation running Mathematica.

Each filename was made up of a unique descriptor concatenated with a timestamp. Mathematica is then used to create a relational database, where basic parameters (such as loaded Q, k_t^2, series and parallel resonant frequencies, ...) are extracted from each device data file. This data can be then re-assimilated and displayed in some convenient format (for example, wafer maps showing the best FOM, (Figure of Merit), at each die and the area of that particular device).

The loaded series Q (or parallel) were measured using the equation [1]

$$Q_l^s = \frac{f}{2} \bullet \frac{d\phi_z}{df}\bigg|_{f_s}$$

where ϕ_z is the argument of the complex impedance and f_s, f_p are the series and parallel frequency resonant peaks. Due to the sparseness of our data directly at the resonance, Q_l^{series} is actually a bit higher than we report.

The ratio of mechanical to electrical energy, k_t^2 (also referred to as the coupling constant), was first measured using the equation [1]

$$k_t^2 = \phi_r/\mathrm{Tan}(\phi_r); \qquad \phi_r = (\pi/2)(f_s/f_p)$$

Later, we realized that this equation could be replaced by the empirically found equation

$$k_t^2 \approx 4.8\frac{f_p - f_s}{f_p + f_s}$$

This simple equation was quite accurate for values of $k_t^2 < 7\%$. The figure of merit (FOM) used for these devices is defined as the product of $Q_l^s * k_t^2$.

For devices fabricated with the original mask set, most of the resonant circles (both impedance circles and S-parameter measurements) were completely below the real axis. For these devices, the measured Q was determined by the half-power points found on the Smith Chart. The k_t^2 was only qualitatively determined from best fits to the data from a full blown simulation (using the Mason model [2]) of the device for the measured thickness of each metal electrode, the silicon nitride membrane and the AlN piezoelectric layer.

Results & Discussion

Figures 2 and 3 show measured data taken from two different devices on the same wafer (using the revised mask set). This particular wafer suffered from acute concentrations of silicon "ash". It was also the first wafer using the revised mask set. This ash acts as mass loading that leads to both local variations in resonant frequencies and depression of the measured Q's.

The wafer from which data shown in Fig.'s 2 & 3 had device-to-device frequency variations of up to 1% within a given 1 cm^2 die. An overall trend of 10 % variation from center to edge is seen on several wafers with and without the "ash" and is indicative of the variation in AlN thickness due to our sputter deposition geometry. The average loaded Q's of this wafer were Q^s = 200 for the series resonance and Q^p = 320 for the parallel resonance, with a maximum Q_l of 500 for Q^s and 900 for Q^p.

Devices tested on wafers showing little or no "ash" had measured Q's on the order of 1200 to 1300. These devices had 20 to 40 Ohms of parasitic resistance in series with the resonator. We are confident that the loaded Q will go up when this parasitic resistance is reduced and all residue of the "ash" is eliminated.

We have made devices with the active piezoelectric film thickness's as thin as 4300 A and as thick as 8000A and have seen fundamental resonances as low as 1.7 GHz and third harmonics as high as 7.5 GHz. Thus, the whole 1 to 10 GHz frequency band has the potential of being covered by these devices.

The maximum "fQ" product for the series resonance was 9.3 x 10^{12} for a wafer showing little or no ash, and 2.0 x 10^{12} for a later wafer with "ash" and with the revised mask set.

Using Mathematica, scatter plots of various parameters were constructed using the data as measured over 82 devices sampled across the 4" die. It was immediately seen that the loaded Q's were "evenly" distributed from "0" to 500 (due to the "ash" as described above). However, the k_t^2 values were more tightly distributed about the average value of 2.4% (with a few points at or near 5%).

We believe that we are not quite optimized for AlN deposition and now are trying to correlate FBAR performance with deposition conditions.

Frequency Tuning (Using Thermal and Voltage Tuning Techniques)

Early on, we recognized the need for active frequency tuning capability for these devices. We explored temperature and DC voltage "in-situ" tuning. One of the fundamental properties of any processed FBAR device is its TCF, the thermal coefficient of frequency ($\delta f/f$ vs. T). We are interested in several aspects: 1) how much power is needed to obtain some ΔT; 2) The potential thermal time constant of these devices; and 3) to see what various blends of metal electrode, Si_xN_y, AlN thickness's that could either maximize the TCF for tuning purposes or minimize TCF for stabilization reasons.

The heater resistors are fabricated at the same time and at the same mask level as the bottom electrode of the FBAR device. This heater layer is conveniently sandwiched between the Si_xN_y supporting layer below and the AlN piezoelectric layer above, preventing the resistors from being exposed to oxygen while running at elevated temperatures. The heater meanders around the FBAR electrode, while still contained within the membrane region. Log magnitude frequency plots of these devices with and without heaters suggested that the proximity of the heaters did not impact the electrical properties of the FBAR.

We have measured voltage as a function of current for our heaters and used this data to plot resistance (V/I) versus power (V*I) and have found this to be extremely linear for power inputs up to 250 mW and corresponding calculated temperatures of about 580 degrees. At these extreme temperatures, the membrane is quite distorted. Eventually, cracks in the AlN form and oxygen reacts with the heater forming an open.

We did not actually measure the temperature directly, instead, we measured the change in the heater resistance and calculated the temperature from the values of the TCR (thermal coefficient of resistance) of the heater.

The ratio of input power to calculated temperature (units of thermal conductance) ranged from .4 to .8 mW/K for a limited set of geometries. This suggests that one could get a much wider range of values if needed. The calculated heat capacity of these membranes are on the order of 10^{-7} Joules/K, suggesting thermal time constants on the order of 100 to 200 μsec. Of course, there is a direct trade-off in response time and power needed to achieve a particular temperature. Faster thermal time constants would require more power for a given temperature excursion.

We made various devices with different ratios of electrode thickness to piezoelectric thickness (keeping the Si_xN_y thickness constant), to see the effect on the TCF of these devices. Although, results are somewhat preliminary, we have been able to change TCF from 45 ppm to 78 ppm.

We also measured the frequency shift of our FBAR resonators as a function of applied dc voltage. The maximum frequency shift (per volt) was about 10 ppm for the 2nd harmonic and significantly smaller for the 1st harmonic. We saw some (~5%) nonlinearity in frequency shift as a function of applied dc voltage. This measurement has not yet been repeated for the devices on the new mask set.

Conclusions

This work represents an early investigative phase studying the applicability and usefulness of FBAR devices. We have demonstrated that temperature tuning is possible and have obtained respectable values for the quality factor, Q, and the coupling coefficient, k_t^2, terms. We are currently working on correlating various properties of the AlN films and of the FBAR devices with AlN deposition conditions.

Acknowledgements

The Authors would like to thank Gladys Koke, Diane Burriesci, and Yogesh Desai for their processing contributions, Bob Taber and Leslie Field for technical discussions, and Rolf Jaeger, Jim Hollenhorst and Gary Baldwin for their kind support.

References

1) K. Lakin, G. Kline, K. McCarron, "High-Q Microwave Acoustic Resonators and Filters", IEEE Trans. Microwave Theory & Tech., v41(12), 2139

(1993)

2) J. F. Rosenbaum, Bulk Acoustic Wave Theory and Devices., Norwood, MA: Artech House, Inc. 1988

Fig. 2. S_{11} device measurement with FOM > 10

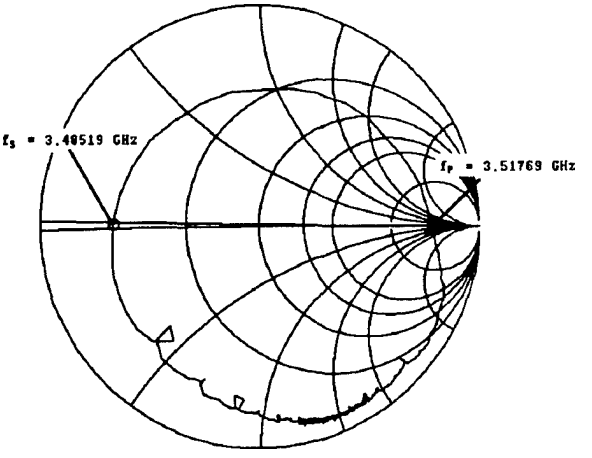


Fig. 3. S_{11} device measurement with $k_t^2 > 5\%$

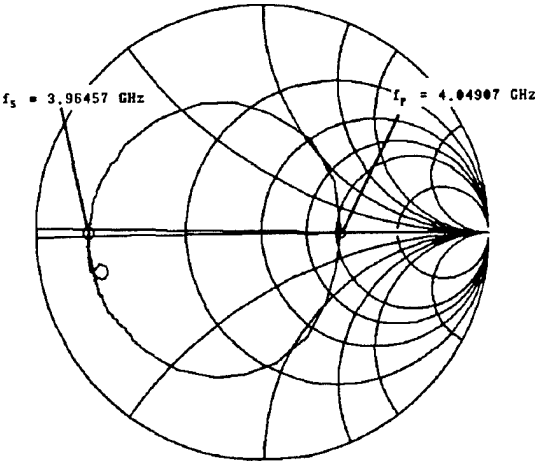
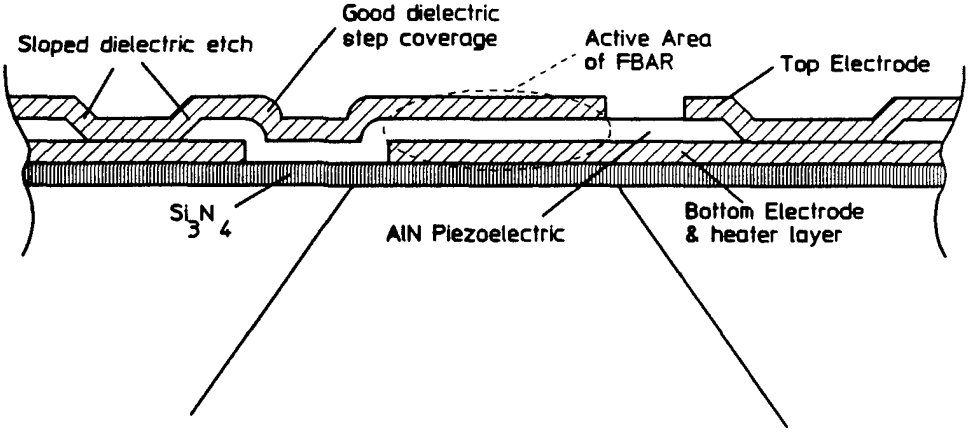


Fig. 1. Profile of an FBAR Device



1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DYNAMIC DEFLECTION AND CAPACITANCE OF MICROMACHINED PIEZOELECTRIC BENDERS

JAN G.SMITS and WAI-SHING CHOI, Department of Electrical Engineering, Boston University, 44 Cummington Street, Boston, MA 02215

Abstract

Micromachined piezoelectric heterogeneous bimorphs have been fabricated using the techniques of I.C. fabrication. The voltage across the piezoelectric element of a heterogeneous bender causes it to contract or expand while the passive element is compressed or elongated and responds with a counterforce, thus building up a moment, which bends the structure.

This dynamic moment has been calculated, and the corresponding Euler - Bernoulli equation has been written with a time-dependent voltage as a source term in the inhomogeneous part of the equation, thus becoming a Rayleigh equation. The bending resistance has been determined using the transformation of area method. Once the solution of this equation is found the deformation can be calculated and the resulting dielectric displacement can be determined. From the latter we can determine the charge that flows into the electrodes, by integrating over the surface area. The relation between the charge on the electrodes and the driving voltage is the capacitance C. The capacitance has the expected resonance behavior: the bimorph goes through resonance at frequencies which are the solutions of the implicit relation $\cos\Omega L \cosh\Omega L = -1$. The deflection of the bimorph has also been calculated and follows a similar behavior.

Introduction

Piezoelectric bimorphs are useful as actuators or sensors, as they exhibit large displacement per applied volt, larger than any other piezoelectric construction. This large effect is reached at the cost of sacrificing force for displacement, so as force generator they are not strong, but they are geared toward being a good deflector. The equation for the deflection is derived along the lines as set out in Timoshenko's paper on the bending of a bimetallic [1]. A full account of bending, rotation and volume displacement and electrode charge due to an applied voltage and also due to an applied force, a moment, and a pressure has been given in [2], while the dynamic behavior is presented in [3]. An electrical network equivalent for bimorphs is detailed in [4].

Fabrication

The bimorphs consist of a sputtered piezoelectric ZnO layer sandwiched between AuCr electrodes, on a sputtered nonpiezoelectric Si_3N_4 layer. The ZnO film was around 1μm thick as was the Si_3N_4film, while the AuCr films consisted of 1000Å of Cr and 2000Å of Au. After the deposition of these films photoresist was spun on and a moat around the beam was opened to etch a "U" shaped window in the Si_3N_4, which was then etched in HF, until the etch stopped at the silicon substrate. Then the wafers were immersed in the ZnO etch bath to remove the sacrificial ZnO layer of which the deposition preceded the deposition of the previously mentioned films. This ZnO etch bath consisted of an acetic-phosphoric acid mixture.

The deflection has been measured as a function of the AC voltage as well as of the DC biasing voltage and has been compared with the theoretical prediction of the deflection behavior. As it turns out there is a second bending effect present: the ZnO is conductive, which results in Joule heating of the beam. As the structure is highly asymetric with Au on the top and Si_3N_4on the bottom, with respective thermal expansion coefficients of 18 and 0.5 ppm, one can expect considerable *bimetallic* bending due to this thermal effect. This effect also shielded the experimental observation of the frequency dependence of the capacitance, as the resistance was in parallel with this capacitance. Due to the quadratic nature of Joule heating, resonance at half of the natural resonance frequency was found as well. The DC deflection versus voltage curves could well be represented by a linear term for the piezoelectric effect, a quadratic term for the Joule heating, and a fourth order term, for which no good physical effect could be identified, but which corresponds to a saturation effect. The coefficient of the linear term was found to correspond to a piezoelectric coefficient of $d_{31} = 5.2 \times 10^{-12}$, equal to the accepted bulk value of ZnO. As the measured deflections are quite large, up to 1mm, with voltage sensitivities up to 300μm per volt for a 3mm long bimorph, it could well be that the Euler-Bernoulli equation is no longer valid, as it presumes that the deflections remain "small", hence it may be that the fourth order term is a mechanical nonlinear effect. Corresponding to the fourth term there was a resonance peak at one quarter of the fundamental mode of the beam.

0-7803-1945-1/94 $3.00 © 1994 IEEE

Centroid and Bending Stiffness

A beam of constant width w consisting of more than one material, each having a different Young's modulus can be represented by another beam of the same total thickness, and of the same Young's modulus by modulating the widths of the components. A stiffness difference in the original beam shows up as a width difference in the new beam.

The profile of a bimorph which consists of a composite of Si_3N_4, Cr, Au, ZnO, Au and Cr again, can be calculated using the Young's moduli of those materials, which are (in the order of the first four) 131 GPa, 324 GPa, 42.7 GPa, 133 GPa. The ratio of these Young's moduli with respect to that of ZnO is 0.985, 2.44, 0.321, 1. This is the ratio of the widths of the composite's cross-section as shown in figure 1. The thickness of each of the layers is respectively, (starting with Si_3N_4) $1\mu, 0.1\mu, 0.2\mu, 1\mu, 0.1\mu, 0.2\mu$.

The centroid position of the beam can be determined using the equation for centroids [5]:

$$\overline{y} = \frac{\int_A y dA}{\int_A dA} \tag{1}$$

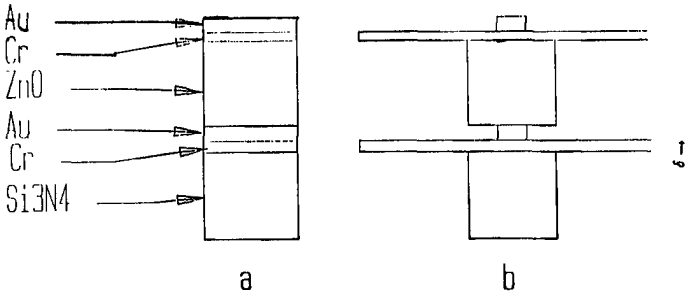


FIGURE 1, Cross section (a) and transformed cros section (b) of the beam.

which turns out to be at a distance of 1.28μ from tl bottom of the bimorph, almost in the Gold-ZnO inte face. Using this value we can determine the moment inertia of the beam around the centroid using the the expression

$$I = \int_A y^2 dA \tag{2}$$

where the center of the coordinate system has been chosen at the centroid position. For this particular film we found that the bending stiffness EI_{eff} is 3.8×10^{-11} Nm^2. The mass of an infinitesimally thin slice of the beam with length dx is given by $\rho A_{eff} dx = \Sigma_i w \rho_i t_i dx$, where t_i is the thickness of the i-th layer. Using values for the specific densities of 3000, 7140, 19320, 5600, in the order of the first four layers, we find:$\rho A_{eff} = 3.55 \times 10^{-6}$, which results in a term $EI_{eff}/\rho A_{eff} = 1.08 \times 10^{-5}$. This calibrates Ω as $\Omega = \sqrt{\omega}/3.29 \times 10^{-3}$. We use this to get a value for the factor $a = \sqrt{EI_{eff}/\rho A_{eff}} = 3.29 \times 10^{-3}$. This can be used in the expression for the deflection and capacitance [3]:

$$\delta = \frac{3 d_{31} V \sin \Omega L \sinh \Omega L}{4h^2 \Omega^2 (1 + \cos \Omega L \cosh \Omega L)} \tag{3}$$

$$C = \frac{wL}{2h}(\epsilon_{33}^T - \frac{d_{31}^2}{s}) + \frac{3w}{8h} \frac{d_{31}^2}{s_{11}^E \Omega} \frac{\cosh \Omega L \sin \Omega L + \cos \Omega L \sinh \Omega L}{1 + \cos \Omega L \cosh \Omega L} \tag{4}$$

where we will make the appropriate substitutions, as mentioned above, as well as

$$s = \Sigma_i s_{11i} t_i / h \tag{5}$$

We have also measured the deflection of bimorphs as a function of frequency and we have plotted the measured deflection together with the theoretical prediction in figure 2.

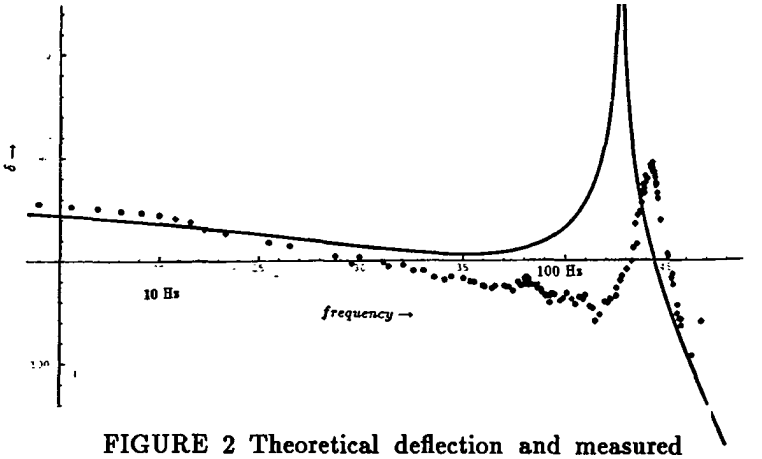


FIGURE 2 Theoretical deflection and measured deflection of a ZnO on Si_3N_4bimorph.

As can be seen in figure 2, we have good qualitative agreement between the measured and the theoretical deflections. The resonance of the bimorph was calculated using the thicknesses of the different layers, as they were measured with a Scanning Electron Microscope, however, the accuracy of a thickness measurement is not that great, if the thicknesses are of the order of a few thousand Ångstrom, and as the resonance frequency depends on the six thicknesses and the six Young's moduli, and the six specific densities of the constitutive materials, in all eighteen variables, we can see that the agreement is quite remarkable.

Capacitance

The capacitance has been calculated using expression (4) in which we have substituted for the ratio d_{31}^2/s_{11}^E the appropriate expression, taking into account that the piezoelectric element is loaded by the five other layers, which results in a static part of the capacitance as:

$$C_{stat} = \frac{wL}{h}(\epsilon_{33}^T - \frac{d_{31}^2}{\Sigma_i^6 s_i t_i / h}) \qquad (6)$$

In the dynamic part the s_{11}^E has been removed with its moment of inertia and replaced with the bending stiffness as taken from (2). The theoretical frequency behavior of the capacitance is plotted in figure 3.

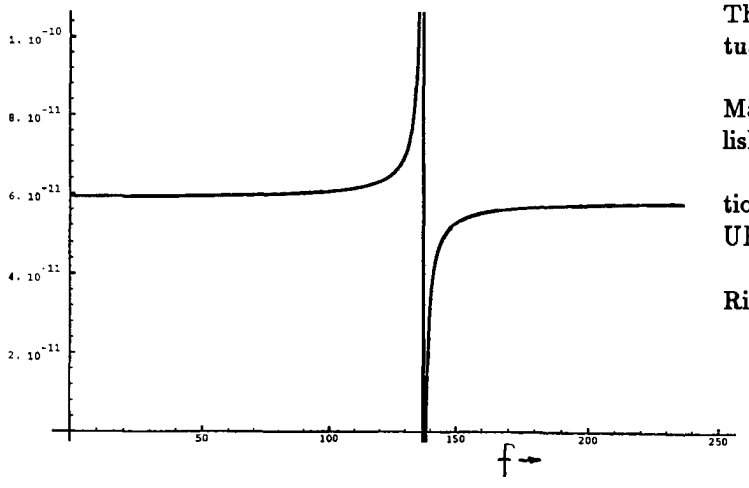


Figure 3, The capacitance of the bimorph as a function of frequency.

We have not been able to confirm the capacitance experimentally, as the piezoelectric ZnO turned out to be conductive, and the parallel resistor it forms overshadowed the capacitance. From figure 3 we see that the resonance of the capacitance occurs at the beam resonance of around 138 Hz, but that the antiresonance is quite close, at 139 Hz.

Conclusions

We have shown that the resonance behavior of the deflection of a cantilever multilayer bimorph follows from the Euler -Bernoulli-Rayleigh equations. We calculated it and compared it with the experimentally observed deflections of a Si_3N_4, Au, Cr, ZnO, Au, Cr, cantilevered bimorph. Resonance was observed at 141 Hz and calculated at 138 Hz. The capacitance has the same resonance frequency, and an antiresonance frequency of 139 Hz.

REFERENCES

1. S. Timoshenko, Reference in Strength of Materials, Krieger Publishing, Malabar Florida, 1955, Refering to his article in J. Opt. Soc. Am. Vol 11, p233, 1925.

2. Jan G. Smits, Piezoelectric Micropump with Three Valves Working Peristaltically , Sensors and Actuators, A21-A23 pp. 203-206, 1990.

3. J.G.Smits and A.Ballato, Dynamic Amittance Matrix of Piezoelectric Cantilever Bimorphs, to be published in Journal of Microelectromechanical Systems.

4. A.Ballato and J.G.Smits, Network Representation for Piezoelectric Bimorphs, IEEE Transactions on UFFC, Vol 38, 1991, pp. 595-602.

5. D.Q.Fletcher, Mechanics of Materials, Holt, Rinehart and Winston, New York, 1985, p412.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A STUDY OF QUARTZ TUNING FORK RESONATORS IN THE OVERTONE MODE

Eishi Momosaki,Takeo Kaneko and Tatsuya Shimoda

Seiko Epson Corporation
3-3-5 Owa,Suwa-shi,Nagano-ken,392 Japan

Abstract

Quartz tuning fork resonators in the overtone mode have been studied in less detail than those in the fundamental mode. We will discuss here the characteristics and optimal design to put them into practical use.

1.Introduction

Recently quartz crystal resonators with frequencies of hundreds of kilo-hertz are being used widely in the field of pagers and wireless instruments. Quartz tuning fork resonators in the overtone mode are suitable for such applications because of their smaller size. However, quartz tuning fork resonators in the overtone mode have been studied in less detail than those in the fundamental mode. This is because the resonators in the fundamental mode have been used in wristwatches for a long time. It is necessary for quartz tuning fork resonators in the overtone mode to be carefully designed from the following viewpoint: total vibrational system design including the mounting leads and capsules, since the vibration in the base portion of quartz tuning forks is transmitted to the mounting leads and the capsules as a result of the larger amplitude in the base portion.

We will discuss here the characteristics and optimal design from the above mentioned point of view.

2.Influence of vibrational leakage

2-1.Evaluation of vibrational leakage

Figure 1 shows a simplified view of a sample of a quartz tuning fork resonator unit in the overtone mode. Resonant frequency is about 500 kHz. Diameter of supporting lead is 200μm, the length in the capsule 600μm and the length of attached area to a quartz tuning fork 500μm. The deviations of resonant frequency and crystal impedance(CI) are defined as the data of the clamping test when the capsule is clamped by a paper clip in the direction of z-z' in Fig.1. Such clamping test data are useful to evaluate the influence of the vibrational leakage of quartz tuning fork resonators in the overtone mode. Figure 2(a) shows the experimental data of the resonant frequency deviations in the cases of two different types of capsules. Capsule A is better designed and

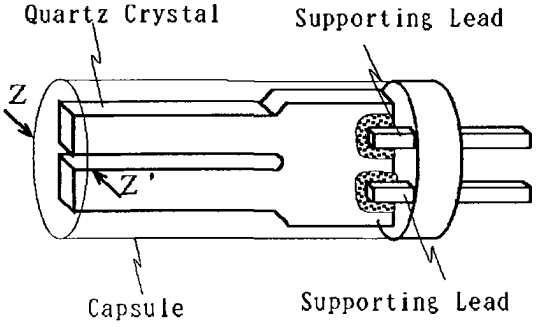


Fig.1 View of a quartz tuning fork resonator.

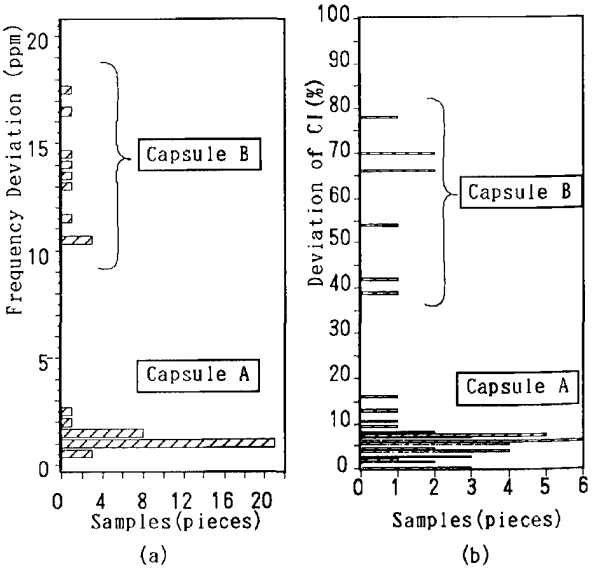


Fig.2 Experimental data of (a)the resonant frequency deviations and (b)CI deviations in two different types of capsules.

0-7803-1945-1/94 $3.00 © 1994 IEEE

has a diameter 1.3 times larger than that of capsule B. Figure 2(b) shows the data of CI deviation in the same clamping test. It is considered that the differences of the characteristics are caused by the differences of the mechanical boundary conditions for the capsules in the total vibrational system including the quartz tuning forks, mounting leads and capsules as shown in Fig.1. This means that the vibration of the quartz tuning fork is transmitted to the mounting leads and the capsules. In this way, the vibrational leakage influences the resonant frequency and CI depending on the clamping conditions.

2-2.Influence on temperature characteristics

One can often observe unwanted temperature characteristics of the resonant frequency and CI in quartz tuning fork resonators in the overtone mode as shown in Fig.3 and 4. There occur abrupt changes in both frequency and CI at specific temperatures; e.g., 0°C and 80°C in Fig.3 and 35°C in Fig.4.

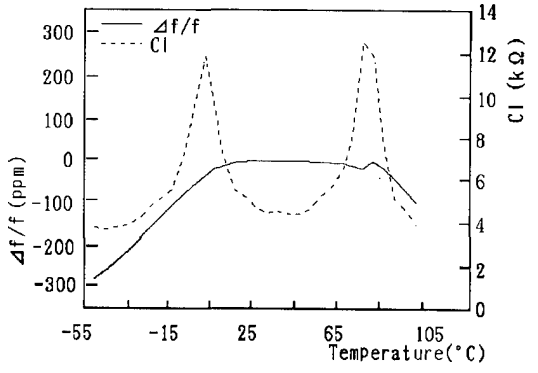


Fig.3 Measured frequency-temperature characteristics and CI-temperature characteristics of a sample resonator.

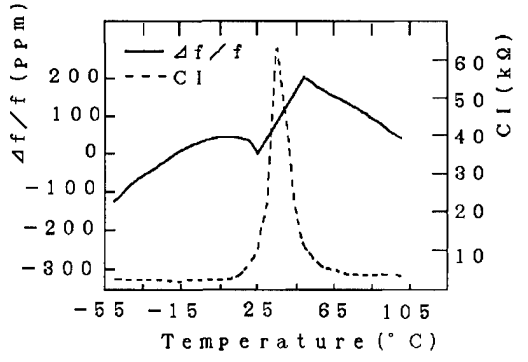


Fig.4 Measured frequency-temperature characteristics and CI-temperature characteristics of a sample resonator.

It is supposed that the phenomenon of unwanted temperature characteristics is caused by the coupling of other vibrational modes with the main vibrational mode. The coupling phenomenon between modes in a quartz resonator has been studied often and reported by Koga[1], Onoe[2], Tomikawa and Okazaki[3] from the points of view of electrical coupling and mechanical coupling mechanisms. They reported that the unwanted temperature characteristics were changed when the loading capacitor in the oscillator was changed, especially in the case of electrical coupling of plural vibrational modes. In our case, however, there are no changes in the above mentioned characteristics in Fig.3 and Fig.4. We can therefore suppose that the dominant coupling mechanism is mechanical coupling. Furthermore, it will be considered that it is not caused by the usual mechanism in which the vibrational modes in the quartz tuning fork are mutually coupled, but instead by coupling the main mode with the vibrations of the supporting leads and capsules.

First, we intend to explain the above mentioned phenomena using the equivalent circuit model as shown in Fig.5. It consists of the arms of the main vibration of a quartz tuning fork, a vibration of supporting leads,and a vibration of a capsule corresponding to the mechanical structures shown in Fig.1. A peak of a sudden change of the frequency and also that of the crystal impedance can be solved by use of this model. The circuit may be simplified such that only the part in the dotted line is used when a mode coupling with a main mode is specified among several coupling modes. In that case, port 3-3' may be shorted when the boundary condition at port 3-3' is mechanically free. For this condition, a peak of a sudden change of the frequency *DF* and also that of a crystal impedance *DR* can be written as follows.

$$x=1-\frac{\omega}{\omega_2}=\alpha\cdot\Delta T, \tag{1}$$

$$C_K''=\frac{2C_2x}{(4+\frac{1}{Q_2^2})x^2-\frac{2}{Q_2^2}x+\frac{1}{Q_2^2}}, \tag{2}$$

$$\frac{\Delta f}{f}\fallingdotseq-\frac{1}{2}\frac{C_1}{C_K}\frac{C_K''}{C_K}, \tag{3}$$

$$\frac{\Delta R}{R}\fallingdotseq\frac{R_2}{R_1}\frac{C_2^2}{C_K^2}\frac{C_K''}{2C_2x}, \tag{4}$$

$$Df\fallingdotseq\frac{1}{4}\frac{C_1C_2Q_2}{C_K^2}, \tag{5}$$

$$DR\fallingdotseq\frac{R_2}{R_1}\frac{C_2^2}{C_K^2}Q_2^2. \tag{6}$$

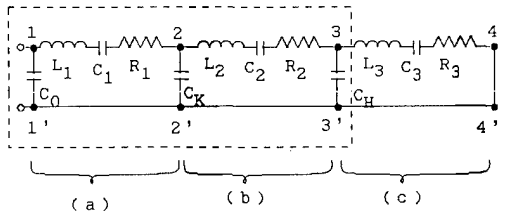


Fig.5 Equivalent circuit for a quartz tuning fork resonator in the overtone mode. (a)Circuit for the main vibrarion of a quartz tuning fork. (b)Circuit for vibration of supporting leads. (c)Circuit for vibration of a capsule.

In eq.(1), ω , ω_2 , a , ΔT are the angular resonant frequency of the quartz tuning fork, the angular resonant frequency of the other structures, the temperature coefficient and the deviation of temperature, respectively. In this equation, it is also assumed that ω will change less than ω_2 when the temperature is changed and that the temperature deviation ΔT is measured from the temperature point at which $\omega_2=\omega$. In the other equations, C_1 , C_2 , R_1 , R_2 , Q_1 , Q_2 and C_K are the values corresponding to the devices in Fig.5. C_K'' is approximately the capacitance deviation from C_K for the load at port 2-2'. In eq.(3), $\Delta f/f$ is the frequency deviation of the synthesized frequency f in the total vibrational system from the resonant frequency of the main mode of the quartz tuning fork. $\Delta R/R$ is the deviation of the crystal impedance R of the above mentioned total vibrational system compared with that of the quartz tuning fork.

DF, *DR* and Q_2 are derived from the experimental data in Fig.6 and Fig.7.

The value a cannot be measured directly because it is the first temperature coefficient of a vibration of mechanical structures which are not able to be excited electrically. We adopted our own way to measure the value as shown in Fig.8. We prepare the samples with slightly different resonant frequencies, and measure the frequency temperature characteristics of the main vibrational mode of the quartz tuning fork resonators, each of which has abrupt changes of the characteristics. Then we get the line in Fig.9 which connects the points of the abrupt changes. We obtain a=-105ppm/°C, because it is the coefficient of the line. As a result, we show all of the equivalent circuit parameters in Fig.10 when R_2 equals 4 M ohm. Using these parameters, the calculated temperature characteristics are shown in Fig.11 as the rigid line, in which the experimental values are also shown as the circles corresponding to the specimen shown in Fig.4.

Based on this analysis using the equivalent circuit, we can regard the abnormal phenomena of the frequency-temperature characteristics of the quartz tuning fork in the overtone mode as being caused by the frequency-temperature characteristics of the mechanical vibration of the structures coupling with the vibration of the quartz tuning fork. CI-temperature characteristics can be also explained as well.

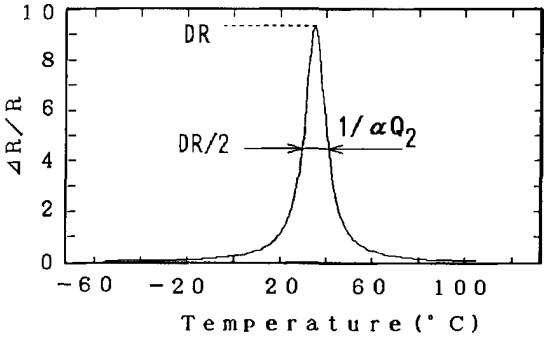


Fig.6 Meaning of Q_2, DR and α in a CI chart.

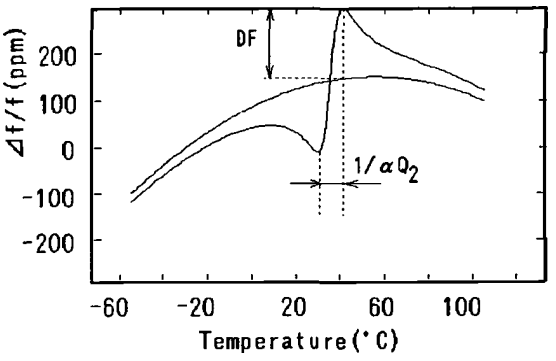


Fig.7 Meaning of Q2, DF and α in a frequency chart.

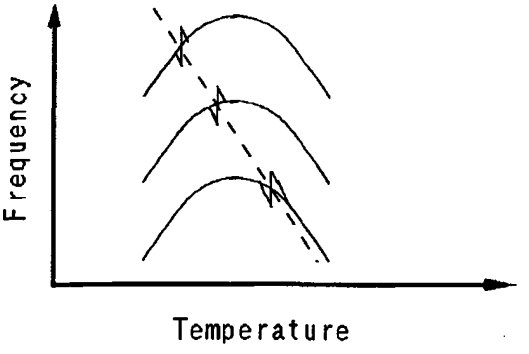


Fig.8 Frequency temperature characteristics for a variety of samples with different frequencies.

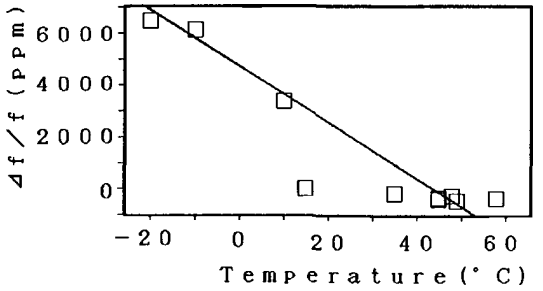


Fig.9 Resonant frequencies at which there occur abrupt changes.

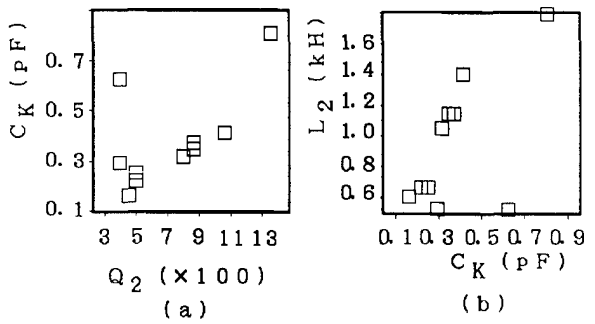


Fig.10 (a)C_K vs Q_2 derived from experimental values of DF, DR, and α. (b)L_2 vs C_K derived from experimental values of DF, DR and α.

Second, the relation between the result of the clamping tests and those of the temperature characteristics will be clarified.

The resonant frequency and CI are measured when the capsule is clamped with a paper clip in the direction of Z-Z' as shown in Fig.1. After removing the paper clip, the resonant frequency and CI are again measured. The mass m of the paper clip was 3.3 g. In this way we obtained the changes of the resonant frequency and CI. We were able to evaluate the abnormal phenomena of the above mentioned temperature characteristics using these changes of frequency and CI. Clamping the capsule with the paper clip corresponds to a mechanical load m at port 3-3' in Fig.5. In electrical terms this means that inductance L has been inserted at the port 3-3'. We obtain the curves in Fig.12, which show the changes of the frequency and CI, by using eqs.(3) and (4). In the circles A, B and C we show the relations between the two frequencies ω_1 and ω_2 of the resonant frequency of the quartz tuning fork and the mechanical structures. In A, ω_1 is larger than ω_2 at a room temperature of 25°C. In B, ω_1 is nearly equal to ω_2 at a room temperature of 25°C. In C, ω_1 is smaller than ω_2. T_0 defined as the temperature at which the two frequencies ω_1 and ω_2 are nearly equal. Since temperature T_0 is lower than the room temperature value of 25°C in A, frequency ω in eq.(3) is located in the vicinity of A on the graph in Fig.12. When the capsule is clamped with the paper clip, ω_2 decreases corresponding to the insertion of inductance L. This is equivalent to the temperature increasing from room temperature in Fig.12 when inductance L is not inserted, because the temperature coefficient of ω_2 is negative. Therefore, the state of point A on the graph in Fig.12 is moved to the state of point A' when clamping with the paper clip. The states of points B and C are moved to the points of B' and C', respectively as well. In Fig.12 the upper figure shows the frequency deviation and the lower figure shows the CI deviation. The points A, B and C in both upper and lower figures of Figure 12 correspond to the circle diagrams A, B and C.

Figure 13 shows the equivalent temperature changes which are estimated from the point of view Fig.12 using the experimental data of CI change. Figure 14 shows the equivalent temperature changes which are calculated using the experimental data of frequency change.

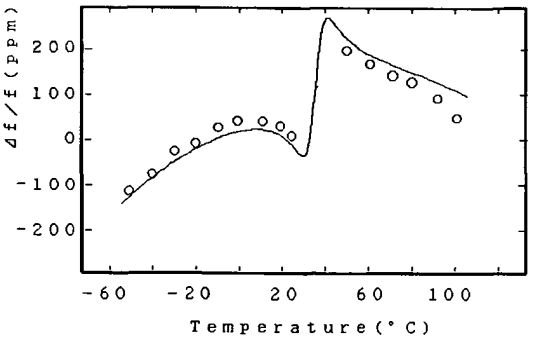


Fig.11 Calculated(solid line) and experimental(circles) frequency temperature characteristics.

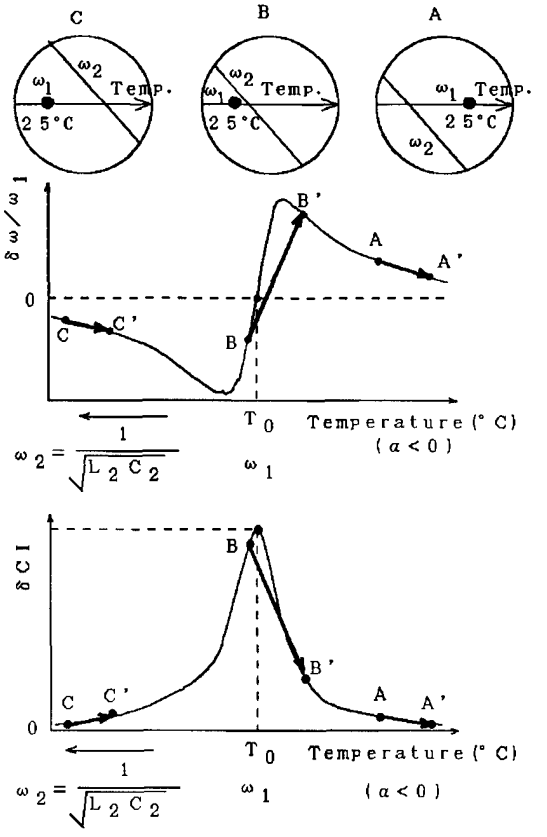


Fig.12 Frequency-temperature characteristics and CI-temperature characteristics for three cases ($\omega_1>\omega_2$, $\omega_1=\omega_2$, $\omega_1<\omega_2$) and their relation to the effects of clamping with a paper clip and temperature changes.

As a result of these figures, we find that temperature changes of 40 to 60°C are equivalent to the mass changes caused by the paper clips. This corresponds to a 1% change of inductance L_2 in Fig.2. The mass changes, as the boundary conditions caused by the clamping test, are equivalent to the temperature changes.

Thus, we clarified that there was a relationship between the data of the clamping test and the data of the temperature characteristics.

3.Optimal design for preventing influence

3-1.Selection of capsules

Three dimensional modal analysis of the capsules was done using three dimensional FEM in order to get the optimal capsule selection. Figure 15 shows the mode charts of the vibrations for four types of the capsules. Figure 16 illustrates the typical vibrational modes. Using the mode charts, the capsule was well selected not to have any vibrational modes in the vicinity of the designed frequency.

3-2.Design of supporting leads

We calculated the vibrational mode shapes for the tuning forks with supporting leads. Several kinds of vibrational modes in which the lead portion was mainly displaced were observed. Figure 17 shows a mode chart for the lead length. The resonant frequency of the mode strongly relating to the lead length(lead mode) decreases as the lead length decreases. Because this lead mode is considered to be a spurious mode, the design should be done to avoid the region where the lead mode frequency is near the targeted frequency. Next, the length of the leads was designed to minimize the influence of the vibrational leakage from the quartz tuning fork to the capsule of the resonator. Figure 18 shows the lead length dependence on the vibrational leakage for a 307 kHz quartz tuning fork resonator. In this figure, the deviation of the frequency caused by the vibrational leakage was about -50 ppm at a lead length of zero but it became almost zero when the lead length was set to 400 μm.

It is well known that the CI of a contour mode resonator is reduced when it is mounted with a lead having an optimal length. This optimal length is considered to be 1/4 of the wavelength of the dominant sound wave propagating through it. In the calculation for our 307 kHz tuning fork resonators, the wave lengths of the longitudinal, transversal and flexural waves are 13.2 mm, 8.2 mm and 2.0 mm, respectively. The optimal lead length 400 μm obtained experimentally is close to 1/4 of the wavelength of the flexural wave. It can be concluded that the vibrational energy of the quartz tuning fork leaks through the lead dominantly by the flexural wave and that the length of leads

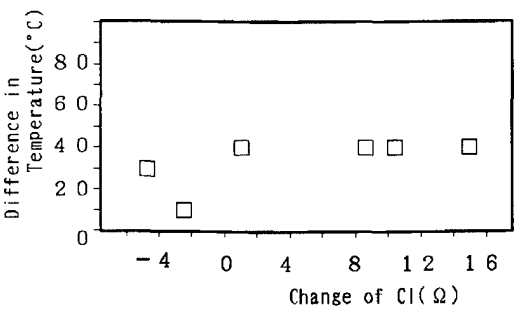


Fig.13 Equivalent differences in temperature vs change of CI.

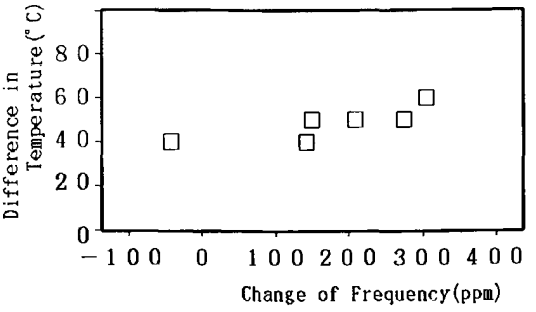


Fig.14 Equivalent differences in temperature vs change of frequency.

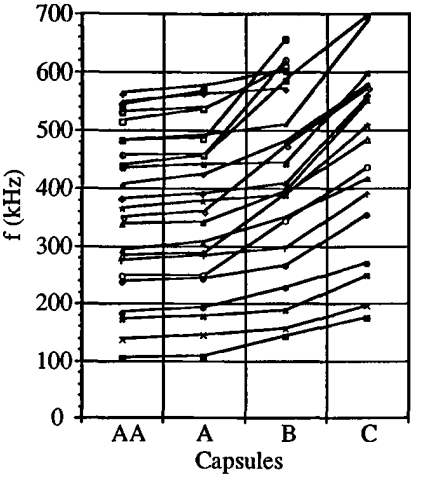


Fig.15 Mode charts for capsules.

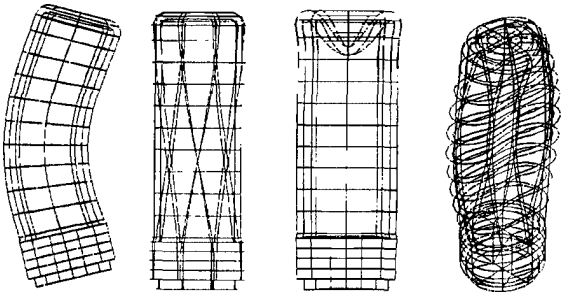


Fig.16 Examples of vibrational modes.

should be set to 1/4 of the wave length of the flexural wave for minimization of the influence of the vibrational leakage.

3-3.Design of quartz tuning forks

In order to minimize the vibrational leakage from a quartz tuning fork to the supporting leads, the base portion of it was designed optimally based on the result of FEM analysis. Figure 17 shows the mode chart calculated by FEM. In this figure, we selected an adequate region for the length of the base portion from the point of view of avoiding the spurious modes. Then, we had to design the length so that the vibrational leakage was minimized. Figure 20 shows the stress of the lead at the end portion as calculated by FEM. The experimental data of CI is also shown in Fig.20. The experimental data of the test samples show good accordance with the results of the analysis. Based on the analysis,we achieved an optimal design of the base portion of the quartz tuning fork with the least vibrational leakage.

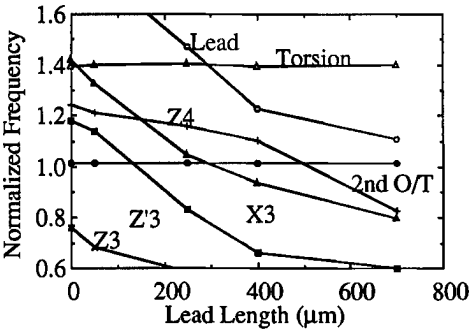


Fig.17 Mode charts for lead length.

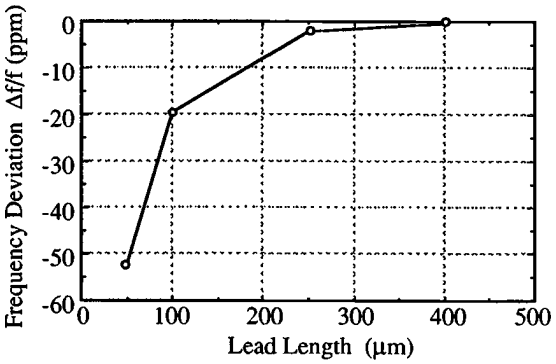


Fig.18 Frequency deviation with respect to lead length.

4.Conclusions

We clarified the mechanism of the unfavorable characteristics of the quartz tuning fork resonators in the overtone mode by considering the influence of the vibrational leakage.

The abnormal phenomena of the frequency-temperature characteristics and CI-temperature characteristics of the quartz tuning fork in the overtone mode can be regarded as being caused by the frequency-temperature characteristics of the mechanical vibration of the structures coupling with the vibration of the quartz tuning fork.

The relation between the results of the clamping tests and those of the temperature characteristics also was clarified.

Consequently, based on the above mentioned considerations and FEM analysis of the total vibrational system including supporting leads and capsules, we can design quartz tuning resonators in the overtone mode with excellent characteristics without the influence of vibrational leakage.

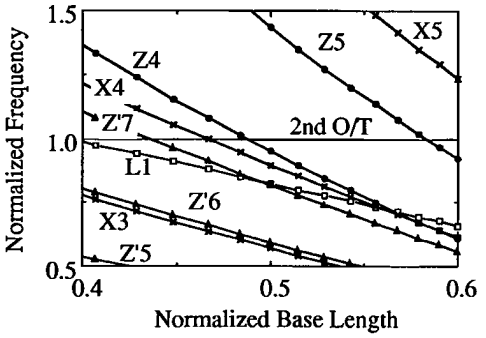


Fig.19 Mode charts for base portion of a quartz tuning fork.

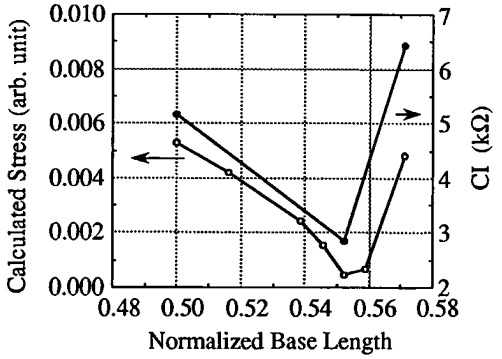


Fig.20 Stress at an edge of a lead with respect to base length.

Acknowledgements

The authors would like to thank the members of the Quartz Device Division for their support. We especially appreciate Mr. S. Miyasaka and Mr. M. Nagai for preparation and measurement of the samples.

References

[1]I.Koga, "Anomalous Vibration(Activity Dip) in AT-Cut(R1) Plates", Trans.IECE Jpn. ,vol.52-A, No.6, 1969,

pp229-240.
[2]M.Onoe, "Analysis and Detection of Unwanted Mode in Quartz Crystal Units for Oscillators",J. IECE Jpn., vol.47, 1964, pp38-47.
[3]Y.Tomikawa and M.Okazaki, "Characteristics of a Quartz Crystal Tuning Fork with Shortened Arm Length for High Frequencies", Proc.5th Symp. Ultrasonic Electronics, Tokyo, 1984, Jpn. J. Appl. Phys. 24, 1985, Suppl. 24-1, pp136-138.
[4]H.Sekimoto, Y.Watanabe, T.Taniguchi and M.Nakazawa, "Analysis of Spurious Characteristics of Rectangular AT-Cut Quartz Resonators Using Plate Equations and One-Dimensional Finite Element Method", Trans. IEE Jpn., vol.111-C, 1991, pp372-377.
[5]H.Fukuyo, A.Yokoyama, N.Ooura and S.Nonaka, "The Thickness-Shear Vibrations of Circular Biconvex AT-Cut Plate", Trans. IEE Jpn. vol.94-A ,No.6, 1974, pp257-264.
[6]H.Sekimoto, Y.Watanabe, Y.Kitada and Y.Oomura, "Effect of Z'-dimensions on Spurious Mode Patterns in Rectangular AT-cut Quartz Plates", Technical Report of IECE Jpn., US93-14, CPM93-28, 1993, pp15-22.
[7]W.P.Mason, Electromechanical Transducers and Wave Filters, D.Van Nostrand Co., New York, 1948, 2nd ed., pp.399-412.
[8]M.Onoe, "General Equivalent Circuit of a Piezoelectric Transducer Vibrating in Thickness Modes", Trans. IECE Jpn., vol.55-A, 1972, pp239-244

Analysis of A Piezoelectric Composite Microcantilever For Accelerometer Applications

J.T. Stewart

US Army Research Laboratory
Electronics and Power Sources Directorate
AMSRL-EP
Fort Monmouth, NJ 07703-5601

Abstract

Analysis of a composite microcantilever is performed with specific application to resonant accelerometers. The structures of interest are composite plates consisting of Silicon with piezoelectric layers such as Zinc Oxide or Quartz. A classical laminate theory approach is applied to the problems of static and dynamic behavior of layered piezoelectric plates under cylindrical bending assumptions. These methods are suitable for the study of thin plates (a/b > 20). A first order plate expansion is used which includes the effects of shear deformation and rotatory inertia. The particular solution under uniform mechanical and electrical loading for the static bending problem is presented in closed form. Using the static and dynamic solutions, the acceleration sensitivity of the composite cantilever is estimated using a perturbation method.

1 Introduction

Analysis of a composite microcantilever is performed with specific application to resonant accelerometers. Numerous examples of resonant structures exist in the area of microelectromechanical systems (MEMS) for sensing applications[8]. These devices include accelerometers [9,10], pressure/force sensors[11], and resonant strain gauges[12]. All these sensors share one thing in common: they measure a mechanical quantity by observing the change in resonant frequency of a small vibrating beam or plate. The present study is concerned with layered composite structures which are constructed of piezoelectric materials on silicon. Figures (1) and (2) show the basic problems of interest.

A classical laminate theory approach [7] is taken in the analysis which has the advantage that the composite plate can be viewed as a whole and complicated layer-wise solutions are not introduced. The disadvantage of this approach is that the microscopic composite behavior (i.e.: traction continuity at layer interfaces) is not addressed, and while the macroscopic plate behavior can be made accurate, extending the results to microscopic composite behavior should be avoided. The laminate theory applied here is accurate for thin plates with a/b > 20.

The approximate equations for the composite plate are developed by employing Mindlin's plate expansion[1] and truncating the series to a first order approximation. A set of shear correction factors are applied to improve the accuracy of the equations. The shear correction factors are determined by comparing the exact cutoff frequencies for the layered plate to those predicted by the first order expansion. Dynamic and static solutions are obtained for cantilever plates under the assumption of cylindrical bending. In these solutions all displacement components representing flexure, shear, length extension, thickness extension, and the zero and first order u_3 components are retained. The general particular solution for the cylindrical bending of a composite plate under uniform mechanical and electrical loading is presented in closed form with the specifics listed in appendix A.

To calculate the sensitivity of the composite cantilever to normal acceleration, Tiersten's methods [4] are employed with the plate expansion which results in a set of non-linear equations from which a perturbation in frequency is obtained. This analysis is applied to a composite cantilever consisting of a Quartz-Si-Quartz sandwich plate with a proof mass located at the free end.

2 Mindlin's Approximate Plate Equations For A Layered Plate

Equations of Motion and Boundary Conditions

Mindlin's approximate equations of motion for a layered piezoelectric plate are developed by assuming displacement and potential fields in the form of infinite power series in the thickness coordinate.When this is done the three dimensional equations of motion are rendered as an infinite sequence of two dimensional partial differential

0-7803-1945-1/94 $3.00 © 1994 IEEE

equations independent of x_2. In this framework, stresses and electric displacements are rendered as resultants and equilibrium is satisfied with respect to these quantities. In general, equilibrium will not hold for actual stresses and electric displacements when a truncated series is used. Figure (3) shows the orientation of a plate with pertinent quantities labeled. The thickness of the plate is denoted as 2b and is assumed to be constant throughout. The area in the x_1-x_3 plane occupied by the plate is denoted as A and the curve formed by intersecting the boundary of A with the middle plane of the plate ($x_2 = 0$) will be denoted by C. Figure (4) shows a cross-sectional view of the layered plate. The quantities $y_1, y_2, \ldots y_{N-1}$ denote the locations of the layer interfaces and the quantities $y_0 = -b$ and $y_N = b$ denote the location of the bottom and top surfaces, respectively, of the plate. A point (x_1, x_2, x_3) is an element of the r^{th} layer if $(x_1, x_3) \in A$ and $x_2 \in [y_{r-1}, y_r]$.

The plate equations are obtained by assuming the following form of the displacements and electric potential:

$$u_j = \sum_{n=0}^{\infty} u_j^{(n)} \eta^n \qquad \phi = \sum_{n=0}^{\infty} \phi^{(n)} \eta^n \tag{1}$$

where u_j and ϕ are the components of displacement and electric potential, respectively, and $\eta = x_2/b$ is a dimensionless thickness coordinate. Equation (1) is used in the variational form of the three dimensional equations of motion and charge equations of electrostatics, and after simplifying the thickness portion of the integral, the following two dimensional equations are obtained:

$$T_{ij,i}^{(n)} - \frac{n}{b} T_{2j}^{(n-1)} + \frac{1}{b} F_j^{(n)} = \sum_{m=0}^{\infty} \rho^{(m+n)} \ddot{u}_j^{(m)} \tag{2}$$

$$D_{i,i}^{(n)} - \frac{n}{b} D_2^{(n-1)} + \frac{1}{b} D^{(n)} = 0 \tag{3}$$

where:

$$T_{ij}^{(n)} = \sum_{r=1}^{N} \int_{\eta_{r-1}}^{\eta_r} T_{ij}^{[r]} \eta^n d\eta \tag{4}$$

and

$$D_i^{(n)} = \sum_{r=1}^{N} \int_{\eta_{r-1}}^{\eta_r} D_i^{[r]} \eta^n d\eta \tag{5}$$

are, respectively, the n^{th} order stress and electric displacement resultants with $T_{ij}^{[r]}$ and $D_i^{[r]}$ denoting the components of stress and electric displacement for the r^{th} layer.

$$\rho^{(n)} = \sum_{r=1}^{N} \left[\frac{\eta_r^{n+1} - \eta_{r-1}^{n+1}}{n+1} \right] \rho^{[r]} \tag{6}$$

represents a weighted average of layer densities with $\rho^{[r]}$ denoting the density of the r^{th} layer, and

$$F_j^{(n)} = T_{2j}(b) - (-1)^n T_{2j}(-b) + b a_j \rho^{(n)} \tag{7}$$

$$D^{(n)} = D_2(b) - (-1)^n D_2(-b) \tag{8}$$

where $F_j^{(n)}$ are the components of the n^{th} order surface traction and body force resultant, with a_j denoting the j^{th} component of acceleration, and $D^{(n)}$ denotes the n^{th} order surface charge resultant.

Using equation (1) in the three dimensional strain-displacement and electric field-potential relations yields infinite power series for the strain tensor, s_{ij}, and electric field vector, E_i:

$$s_{ij} = \sum_{n=0}^{\infty} s_{ij}^{(n)} \eta^n \qquad E_i = \sum_{n=0}^{\infty} E_i^{(n)} \eta^n \tag{9}$$

$$s_{ij}^{(n)} = \frac{1}{2} \left[u_{i,j}^{(n)} + u_{j,i}^{(n)} + \left(\frac{n+1}{b} \right) \left(u_i^{(n+1)} \delta_{2j} + u_j^{(n+1)} \delta_{2i} \right) \right] \tag{10}$$

$$E_i^{(n)} = -\phi_{,i}^{(n)} - \left(\frac{n+1}{b} \right) \delta_{2i} \phi^{(n+1)} \tag{11}$$

where $s_{ij}^{(n)}$ are the n^{th} order components of strain and $E_i^{(n)}$ are the n^{th} order components of electric field. From the surface portion of the variational integral, the natural and forced boundary conditions for the plate problem are deduced as:

$$\bar{T}_j = n_i T_{ij} \qquad \bar{\sigma} = n_i D_i \qquad \text{on A} \tag{12}$$

$$\bar{T}_j^{(n)} = n_a T_{aj}^{(n)} \qquad \bar{\sigma}^{(n)} = n_a D_a^{(n)} \qquad (13)$$
$$a = 1, 3 \qquad \text{on C}$$

$$u_j = \bar{u}_j \qquad \phi = \bar{\phi} \qquad \text{on A} \qquad (14)$$

$$u_j^{(n)} = \bar{u}_j^{(n)} \qquad \phi^{(n)} = \bar{\phi}^{(n)} \qquad \text{on C} \qquad (15)$$

where the barred quantities $\bar{T}_j$, $\bar{\sigma}$, $\bar{T}_j^{(n)}$, $\bar{\sigma}^{(n)}$, $\bar{u}_j$, $\bar{\phi}$, $\bar{u}_j^{(n)}$, and $\bar{\phi}^{(n)}$ represent quantities which are specified on the plate's boundary. The quantities $\bar{T}_j^{(n)}$ and $\bar{\sigma}^{(n)}$ are defined as the edge traction and edge charge resultants and are given as:

$$T_j^{(n)} = \int_{-1}^{1} T_j \eta^n d\eta \qquad \sigma^{(n)} = \int_{-1}^{1} \sigma \eta^n d\eta \qquad (16)$$

First Order Truncation

The general infinite system of equations implied by equations (2) and (3) is truncated to a first order approximation by setting to zero all components of displacement and electric potential that are greater than one:

$$u_j^{(n)} = 0 \qquad \phi^{(n)} = 0 \qquad n > 1 \qquad (17)$$

When this is done the zero and first order stress and electric displacement resultants can be expressed in terms of the remaining displacement and potential terms as:

$$T_{ij}^{(0)} = \bar{c}_{ijkl}^{(0)} u_{k,l}^{(0)} + \frac{1}{b}\bar{c}_{ijk2}^{(0)} u_k^{(1)} + \bar{e}_{kij}^{(0)} \phi_{,k}^{(0)} \qquad (18)$$
$$+ \frac{1}{b}\bar{e}_{2ij}^{(0)} \phi^{(1)} + \bar{c}_{ijkl}^{(1)} u_{k,l}^{(1)} + \bar{e}_{kij}^{(1)} \phi_{,k}^{(1)}$$

$$T_{ij}^{(1)} = \bar{c}_{ijkl}^{(1)} u_{k,l}^{(0)} + \frac{1}{b}\bar{c}_{ijk2}^{(1)} u_k^{(1)} + e_{kij}^{(1)} \phi_{,k}^{(0)} \qquad (19)$$
$$+ \frac{1}{b} e_{2ij}^{(1)} \phi^{(1)} + c_{ijkl}^{(2)} u_{k,l}^{(1)} + e_{kij}^{(2)} \phi_{,k}^{(1)}$$

$$D_i^{(0)} = \bar{e}_{ikl}^{(0)} u_{k,l}^{(0)} + \frac{1}{b}\bar{e}_{ik2}^{(0)} u_k^{(1)} - \varepsilon_{ik}^{(0)} \phi_{,k}^{(0)} \qquad (20)$$
$$- \frac{1}{b}\varepsilon_{i2}^{(0)} \phi^{(1)} + e_{ikl}^{(1)} u_{k,l}^{(1)} - \varepsilon_{ik}^{(1)} \phi_{,k}^{(1)}$$

$$D_i^{(1)} = \bar{e}_{ikl}^{(1)} u_{k,l}^{(0)} + \frac{1}{b}\bar{e}_{ik2}^{(1)} u_k^{(1)} - \varepsilon_{ik}^{(1)} \phi_{,k}^{(0)} \qquad (21)$$
$$- \frac{1}{b}\varepsilon_{i2}^{(1)} \phi^{(1)} + e_{ikl}^{(2)} u_{k,l}^{(1)} - \varepsilon_{ik}^{(2)} \phi_{,k}^{(1)}$$

The quantities $c_{ijkl}^{(n)}$, $e_{kij}^{(n)}$, and $\varepsilon_{ij}^{(n)}$ appearing in equations (18) through (21) represent effective plate material constants for the composite which are weighted averages of the constants for each layer and are defined as:

$$c_{ijkl}^{(n)} = \sum_{r=1}^{N} \left[\frac{\eta_r^{n+1} - \eta_{r-1}^{n+1}}{n+1}\right] c_{ijkl}^{[r]} \qquad (22)$$

$$e_{kij}^{(n)} = \sum_{r=1}^{N} \left[\frac{\eta_r^{n+1} - \eta_{r-1}^{n+1}}{n+1}\right] e_{kij}^{[r]} \qquad (23)$$

$$\varepsilon_{ij}^{(n)} = \sum_{r=1}^{N} \left[\frac{\eta_r^{n+1} - \eta_{r-1}^{n+1}}{n+1}\right] \varepsilon_{ij}^{[r]} \qquad (24)$$

where $c_{ijkl}^{[r]}$, $e_{kij}^{[r]}$, and $\varepsilon_{ij}^{[r]}$ represent, respectively, the elastic stiffness, piezoelectric stress constants, and dielectric permittivity constants for the r^{th} layer. Using the strain-displacement and electric field-potential relations (10) and (11) with the constitutive relations (18) through (21) in the zero and first order equations of motion and electrostatics obtained from equations (2) and (3) yields the general zero and first order differential equations:

$$\bar{c}_{ijkl}^{(0)} u_{k,li}^{(0)} + \bar{c}_{ijkl}^{(1)} u_{k,li}^{(1)} + \bar{e}_{kij}^{(1)} \phi_{,ki}^{(1)} \qquad (25)$$
$$+ \frac{1}{b}\bar{c}_{ijk2}^{(0)} u_{k,i}^{(1)} + \frac{1}{b}\bar{e}_{2ij}^{(0)} \phi_{,i}^{(1)} + \frac{1}{b} F_j^{(0)}$$
$$= \rho^{(0)} \ddot{u}_j^{(0)} + \rho^{(1)} \ddot{u}_j^{(1)}$$

$$\bar{e}_{ikl}^{(0)} u_{k,li}^{(0)} + e_{ikl}^{(1)} u_{k,li}^{(1)} - \varepsilon_{ik}^{(0)} \phi_{,ki}^{(0)} \qquad (26)$$
$$- \varepsilon_{ik}^{(1)} \phi_{,ki}^{(1)} + \frac{1}{b}\bar{e}_{ik2}^{(0)} u_{k,i}^{(1)} - \frac{1}{b}\varepsilon_{i2}^{(0)} \phi_{,i}^{(1)}$$
$$+ \frac{1}{b} D^{(0)} = 0$$

$$\bar{c}^{(1)}_{ijkl} u^{(0)}_{k,li} + c^{(2)}_{ijkl} u^{(1)}_{k,li} + e^{(1)}_{kij} \phi^{(0)}_{,ki} \tag{27}$$
$$+ e^{(2)}_{kij} \phi^{(1)}_{,ki} - \frac{1}{b} \bar{c}^{(0)}_{2jkl} u^{(0)}_{k,l} + \frac{1}{b} (\bar{c}^{(1)}_{ljk2} - \bar{c}^{(1)}_{2jkl}) u^{(1)}_{k,l}$$
$$- \frac{1}{b} \bar{e}^{(0)}_{k2j} \phi^{(0)}_{,k} + \frac{1}{b} \left(e^{(1)}_{2kj} - \bar{e}^{(1)}_{k2j} \right) \phi^{(1)}_{,k}$$
$$- \frac{1}{b^2} \bar{c}^{(0)}_{2jk2} u^{(1)}_{k} - \frac{1}{b^2} \bar{e}^{(0)}_{22j} \phi^{(1)} + \frac{1}{b} F^{(1)}_{j}$$
$$= \rho^{(1)} \ddot{u}^{(0)}_{j} + \rho^{(2)} \ddot{u}^{(1)}_{j}$$

$$\bar{e}^{(1)}_{ikl} u^{(0)}_{k,li} + e^{(2)}_{ikl} u^{(1)}_{k,li} - \varepsilon^{(1)}_{ik} \phi^{(0)}_{,ki} \tag{28}$$
$$- \varepsilon^{(2)}_{ik} \phi^{(1)}_{,ki} - \frac{1}{b} \bar{e}^{(0)}_{2kl} u^{(0)}_{k,l} + \frac{1}{b} \left(\bar{e}^{(1)}_{ik2} - e^{(1)}_{2ki} \right) u^{(1)}_{k,i}$$
$$+ \frac{1}{b} \varepsilon^{(0)}_{2k} \phi^{(0)}_{,k} - \frac{1}{b^2} e^{(0)}_{2k2} u^{(1)}_{k} + \frac{1}{b^2} \varepsilon^{(0)}_{22} \phi^{(1)}$$
$$+ \frac{1}{b} D^{(1)} = 0$$

The barred quantities in equations (18) through (21) and (25) through (28) represent quantities to which have been applied shear correction factors κ_1, κ_2, and κ_3 according to the following rule:

$$\bar{c}^{(0)}_{ijkl} = \kappa^{\mu}_{i+j-2} \kappa^{\nu}_{k+l-2} c^{(0)}_{ijkl} \tag{29}$$

$$\bar{c}^{(1)}_{ijkl} = \frac{1}{2} (\kappa^{\mu}_{i+j-2} c^{(1)}_{ijkl} + \kappa^{\nu}_{k+l-2} c^{(1)}_{ijkl}) \tag{30}$$

$$\bar{e}^{(0)}_{kij} = \kappa^{\mu}_{i+j-2} e^{(0)}_{kij} \qquad \bar{e}^{(1)}_{kij} = \kappa^{\mu}_{i+j-2} e^{(1)}_{kij} \tag{31}$$

where:

$$\mu = \cos^2\left(\frac{ij\pi}{2} \right) \qquad \nu = \cos^2\left(\frac{kl\pi}{2} \right) \tag{32}$$

In equations (29) through (31) there is no implied sum on i,j,k, and l. These correction factors are determined for the composite by using a method similar to that which is discussed in ref [6]. The exact cutoff frequencies are obtained from an exact solution for an infinite layered plate using a transfer matrix analysis [5] and the correction factors are computed by comparing them to those obtained from the plate equations.

In addition to the differential equations of motion, the condition for orthogonality of eigen-solutions is obtained in the usual way as:

$$\int_A (\rho^{(0)} u^{(0)m}_{j} u^{(0)n}_{j} \tag{33}$$
$$+ \rho^{(1)} [u^{(0)m}_{j} u^{(1)n}_{j} + u^{(1)m}_{j} u^{(0)n}_{j}]$$
$$+ \rho^{(2)} u^{(1)m}_{j} u^{(1)n}_{j}) = N^2_{(m)} \delta_{mn}$$

Where $u^{(0)n}_{j}$, $u^{(0)m}_{j}$, $u^{(1)n}_{j}$, and $u^{(1)m}_{j}$ denote the zero and first order displacement components of the n^{th} and m^{th} eigen-solutions with $N_{(m)}$ the normalizing constant for the m^{th} eigen-mode.

3 Dynamic Solutions

General Straight Crested Wave Solution

The dynamic solutions to the first order plate equations under cylindrical bending are obtained by assuming the following forms for the displacements and electric potentials:

$$u^{(0)}_{j} = A^{(0)}_{j} e^{i(\xi x_1 - \omega t)} \tag{34}$$

$$\phi^{(0)} = A^{(0)}_{4} e^{i(\xi x_1 - \omega t)} \tag{35}$$

$$u^{(1)}_{j} = -iA^{(1)}_{j} e^{i(\xi x_1 - \omega t)} \tag{36}$$

$$\phi^{(1)} = -iA^{(1)}_{4} e^{i(\xi x_1 - \omega t)} \tag{37}$$

Where ξ is a wave number along x_1 and ω is the angular frequency. Using equations (34) through (37) in the general differential equations (25) through (28) yields an 8X8 matrix equation with terms which are polynomials in ξ multiplying a column vector of the terms $A^{(0)}_{j}$, $A^{(1)}_{j}$, $A^{(0)}_{4}$, and $A^{(1)}_{4}$. This equation can be organized into the following form:

$$[\xi^2 \underset{\sim}{G}_2 + \xi \underset{\sim}{G}_1 + \underset{\sim}{G}_0 - \omega^2 \underset{\sim}{G}_\rho] \underset{\sim}{A} = \underset{\sim}{0} \tag{38}$$

or:

$$\underset{\sim}{G} \underset{\sim}{A} = \underset{\sim}{0} \tag{39}$$

where:

$$\underset{\sim}{A} = \left[A^{(0)}_{j} \; A^{(1)}_{j} \; A^{(0)}_{4} \; A^{(1)}_{4} \right]^T \tag{40}$$

and the matrices $\underset{\sim}{G}_2$, $\underset{\sim}{G}_1$, $\underset{\sim}{G}_0$, and $\underset{\sim}{G}_\rho$ are matrices of

plate material constants.

For given values of ξ and ω, the terms $A_4^{(0)}$, and $A_4^{(1)}$ are statically condensed from equation (39) as follows: Solve the 8th row of (39) for $A_4^{(1)}$ to give:

$$A_4^{(1)} = G'_1 A_1^{(0)} + G'_2 A_2^{(0)} + G'_3 A_3^{(0)} + G'_4 A_1^{(1)} + G'_5 A_2^{(1)} + G'_6 A_3^{(1)} + G'_7 A_4^{(0)} \quad (41)$$

where:

$$G'_i = -\frac{G_{8i}}{G_{88}} \qquad i = 1, 2, \ldots, 7 \quad (42)$$

which results in the reduced system:

$$\underset{\sim}{G'}\underset{\sim}{A'} = \underset{\sim}{0} \quad (43)$$

where $\underset{\sim}{A'}$ is a seven component vector made up of the components of $\underset{\sim}{A}$, less the term $A_4^{(1)}$ and:

$$G'_{ij} = G_{ij} - \frac{G_{i8}G_{8j}}{G_{88}} \quad (44)$$

This procedure is repeated on the reduced matrix to eliminate $A_4^{(0)}$ to give:

$$G''_i = -\frac{G'_{7i}}{G'_{77}} \qquad i = 1, 2, \ldots, 6 \quad (45)$$

and

$$\underset{\sim}{G''}\underset{\sim}{A''} = \underset{\sim}{0} \quad (46)$$

where:

$$G''_{ij} = G'_{ij} - \frac{G'_{i7}G'_{7j}}{G'_{77}} \quad (47)$$

is a 6X6 matrix representing a piezoelectrically stiffened elastic problem and A'' is a six component vector containing the displacement components. For a non-trivial solution to equation (46) to exist, the determinant of $\underset{\sim}{G''}$ must vanish:

$$\left|\underset{\sim}{G''}\right| = 0 \quad (48)$$

Equation (48) is the dispersion relation for the piezoelectically stiffened problem. This equation represents a 12th order polynomial in even powers of ξ. This polynomial will admit 12 roots, ξ_q, q = 1,2,...,12, and 12 corresponding amplitude ratios, $\alpha_{jq}^{(0)}, \alpha_{jq}^{(1)}, q = 1, 2, \ldots, 12$.

With this, the displacement components can be represented as a linear combination of 12 partial wave solutions (omitting the $e^{i\omega t}$ term):

$$u_j^{(0)} = \sum_{q=1}^{12} B_q \alpha_{jq}^{(0)} e^{i\xi_q x_1} \quad (49)$$

$$u_j^{(1)} = -i \sum_{q=1}^{12} B_q \alpha_{jq}^{(1)} e^{i\xi_q x_1} \quad (50)$$

With given solutions (49) and (50) the electric potential components can be reconstructed as:

$$\phi^{(0)} = \sum_{q=1}^{12} B_q \alpha_{4q}^{(0)} e^{i\xi_q x_1} \quad (51)$$

$$\phi^{(1)} = -i \sum_{q=1}^{12} B_q \alpha_{4q}^{(1)} e^{i\xi_q x_1} \quad (52)$$

where:

$$\alpha_{4q}^{(0)} = \sum_{j=1}^{3} [\alpha_{jq}^{(0)} G''_{jq} + \alpha_{jq}^{(1)} G''_{(j+3)q}] \quad (53)$$

and

$$\alpha_{4q}^{(1)} = \sum_{j=1}^{3} [\alpha_{jq}^{(0)} G'_{jq} + \alpha_{jq}^{(1)} G'_{(j+3)q}] + \alpha_{4q}^{(0)} G'_{7q} \quad (54)$$

Using these equations in the constitutive relations (18) through (21) gives expressions for the x_1-normal components of stress traction and electric charge as:

$$T_{1j}^{(0)} = \sum_{q=1}^{12} B_q d_{jq}^{(0)} e^{i\xi_q x_1} \quad (55)$$

$$T_{1j}^{(1)} = \sum_{q=1}^{12} B_q d_{jq}^{(1)} e^{i\xi_q x_1} \quad (56)$$

$$D_1^{(0)} = \sum_{q=1}^{12} B_q d_{4q}^{(0)} e^{i\xi_q x_1} \quad (57)$$

$$D_1^{(1)} = \sum_{q=1}^{12} B_q d_{4q}^{(1)} e^{i\xi_q x_1} \quad (58)$$

where $d_{jq}^{(0)}$ and $d_{jq}^{(1)}$ are amplitude ratios for stress and electric displacement which are functions of the wave numbers, displacement and potential amplitude ratios, and the composite plate material constants.

Free Vibrations Of A Cantilever

The boundary conditions for a cantilever with a mechanically clamped end at $x_1 = -a$ and a mechanically free end at $x_1 = +a$ are given as follows:

$$u_j^{(n)} = 0 \qquad x_1 = -a \quad (59)$$

$$T_{1j}^{(n)} = 0 \qquad x_1 = a \quad (60)$$

When a rigid proof mass is present at $x_1 = +a$ the conditions (60) are modified as:

$$\begin{aligned} bT_{1j}^{(0)} + M\ddot{u}_j^{(0)} &= 0 \qquad x_1 = a \\ b^2 T_{1j}^{(1)} + \delta_{1j} M(\bar{x}_1 \ddot{u}_2^{(0)} - \bar{x}_2 \ddot{u}_1^{(0)}) &= 0 \end{aligned} \quad (61)$$

where M is the mass of the proof mass per unit width and $\bar{x}_1$ and $\bar{x}_2$ represent the distances along x_1 and x_2, respectively, from the centroid of the proof mass structure to the end of the plate's middle surface ($x_1 = a$, $x_2 = 0$). It is assumed that $\bar{x}_3 = 0$. Using equations (49),(50),(55), and (56) with the boundary conditions (59) and (61) gives:

$$\sum_{q=1}^{12} B_q \alpha_{jq}^{(0)} e^{-i\frac{\pi}{2} z_q r} = 0 \quad (62)$$

$$\sum_{q=1}^{12} B_q \alpha_{jq}^{(1)} e^{-i\frac{\pi}{2} z_q r} = 0 \quad (63)$$

$$\sum_{q=1}^{12} B_q \tilde{d}_{jq}^{(0)} e^{i\frac{\pi}{2} z_q r} = 0 \quad (64)$$

$$\sum_{q=1}^{12} B_q \tilde{d}_{jq}^{(1)} e^{i\frac{\pi}{2} z_q r} = 0 \quad (65)$$

where $z_q = \xi_q / \frac{\pi}{2b}$ is a dimensionless wave number, $r = a/b$ is the aspect ratio of the plate, and the terms $\tilde{d}_{jq}^{(0)}$ and $\tilde{d}_{jq}^{(1)}$ are given as:

$$\tilde{d}_{jq}^{(0)} = d_{jq}^{(0)} - \frac{M}{b}\omega^2 \alpha_{jq}^{(0)} \quad (66)$$

and

$$\tilde{d}_{jq}^{(1)} = d_{jq}^{(1)} - \delta_{1j}\frac{M}{b^2}\omega^2\left(\bar{x}_1 \alpha_{2q}^{(0)} - \bar{x}_2 \alpha_{1q}^{(0)}\right) \quad (67)$$

Equations (62) through (65) can be arranged into a 12X12 matrix problem of the form:

$$\tilde{H}\tilde{B} = 0 \quad (68)$$

For a non-trivial solution $\tilde{B}$ to equation (68) to exist, the determinant of the coefficient matrix must vanish:

$$|\tilde{H}| = 0 \quad (69)$$

Which is the frequency equation for the cantilever problem. The values of ω which satisfy (69) are the resonances of the cantilever.

Figure (5) shows the flexural and shear mode shapes for the first three resonances of a ZnO-Si bimorph. Figure (6) shows the effect of an end mass structure on the first resonance of a Quartz-Si bimorph.

4 Static Solutions

Homogeneous Solution

The homogeneous solution for the deflection of the plate under cylindrical bending is obtained by setting the traction and charge resultants $F_j^{(0)}$, $F_j^{(1)}$, $D^{(0)}$, and $D^{(1)}$ to zero in the static differential equations ($\omega = 0$) and assuming the following form of the displacements and potentials:

$$w_j^{(0)} = A_j^{(0)} e^{z\zeta} \qquad \psi^{(0)} = A_4^{(0)} e^{z\zeta} \quad (70)$$

$$w_j^{(1)} = A_j^{(1)} e^{z\zeta} \qquad \psi^{(1)} = A_4^{(1)} e^{z\zeta} \quad (71)$$

where $\zeta = \frac{x_1}{b}$ is a dimensionless coordinate along x_1 and

z represents a constant to be determined. To differentiate the static solutions from the dynamic ones, the notation: $w_j^{(n)}$ and $\psi^{(n)}$ are used for static displacement and electric potential, respectively. Using equations (59) and (60) in the differential equations (25) through (28) results in an 8X8 matrix equation similar to equation (72) with terms which are polynomials in z. This equation can be represented as:

$$GA = 0 \tag{72}$$

A non-trivial solution to equation (61) exists if:

$$|G| = 0 \tag{73}$$

Equation (62) represents the characteristic equation which takes the form of a 16^{th} order polynomial in z. This polynomial will admit 16 roots, or characteristic values, only 6 of which are non-zero (4 for the non-piezoelectric case). For each z_q, q = 1,2,...,6, there exists an amplitude ratio, $\bar{\alpha}_{jq}^{(0)}, \bar{\alpha}_{jq}^{(1)}, q = 1, 2, \ldots 6$. With this the homogeneous solution may be written as a linear combination of functions:

$$w_j^{(0)} = \sum_{q=1}^{6} B_q \bar{\alpha}_{jq}^{(0)} e^{z_q \zeta} \tag{74}$$

$$w_j^{(1)} = \sum_{q=1}^{6} B_q \bar{\alpha}_{jq}^{(1)} e^{z_q \zeta} \tag{75}$$

$$\psi^{(0)} = \sum_{q=1}^{6} B_q \bar{\alpha}_{4q}^{(0)} e^{z_q \zeta} \tag{76}$$

$$\psi^{(1)} = \sum_{q=1}^{6} B_q \bar{\alpha}_{4q}^{(1)} e^{z_q \zeta} \tag{77}$$

Particular Solution

The particular solution for the static bending problem is constructed for the case of uniform loading, i.e.:

$$F_j^{(0)} = \text{const} \qquad F_j^{(1)} = \text{const} \tag{78}$$

$$D^{(0)} = \text{const} \qquad D^{(1)} = \text{const} \tag{79}$$

The solution sought is, in general, a 4^{th} order polynomial which can be represented as follows:

$$w_j^{(0)} = \delta_{2j}\Gamma\zeta^4 + (\Delta_j + \delta_{2j}B_7)\zeta^3 + (H_j + \Lambda_j B_7 + \delta_{2j}B_8)\zeta^2 + B_{j+8}\zeta + B_{j+11} \tag{80}$$

$$w_j^{(1)} = -4\delta_{1j}\Gamma\zeta^3 + (\Psi_j - 3\delta_{1j}B_7)\zeta^2 + (M_j + N_j B_7 - 2\delta_{1j}B_8)\zeta + \Sigma_j + \Xi_j B_7 + \Pi_j B_8 + K_{jm}B_{m+8} + \bar{K}_j B_{15} \tag{81}$$

$$\psi^{(0)} = \Delta_4\zeta^3 + (H_4 + \Lambda_4 B_7)\zeta^2 + B_{15}\zeta + B_{16} \tag{82}$$

$$\psi^{(1)} = \Psi_4\zeta^2 + (M_j + N_j B_7)\zeta + \Sigma_4 + \Xi_4 B_7 + \Pi_4 B_8 + K_{4m}B_{m+8} + \bar{K}_4 B_{15} \tag{83}$$

The constants Γ, Δ_j,...,etc. appearing in equations (69) through (72) are complicated functions of the composite plate material constants, the loads, and the plate's thickness. While these quantities are very complicated, they can be evaluated in closed form and are listed in their entirety in appendix A. The quantities $B_7, \ldots, B_{16}$ are unknown constants to be determined.

General Solution of a Cantilever Under Normal Acceleration

Equations (74) through (77) and (80) through (83) can be organized into the following 8 equations:

$$w_j^{(0)} = \sum_{q=1}^{6} B_q \bar{\alpha}_{jq}^{(0)} e^{z_q \zeta} + \sum_{q=7}^{16} B_q \beta_{jq}^{(0)}(\zeta) + \mu_j^{(0)}(\zeta) \tag{84}$$

$$w_j^{(1)} = \sum_{q=1}^{6} B_q \bar{\alpha}_{jq}^{(1)} e^{z_q \zeta} + \sum_{q=7}^{16} B_q \beta_{jq}^{(1)}(\zeta) + \mu_j^{(1)}(\zeta) \tag{85}$$

$$\psi^{(0)} = \sum_{q=1}^{6} B_q \bar{\alpha}_{4q}^{(0)} e^{z_q \zeta} + \sum_{q=7}^{16} B_q \beta_{4q}^{(0)}(\zeta) + \mu_4^{(0)}(\zeta) \tag{86}$$

$$\psi^{(1)} = \sum_{q=1}^{6} B_q \bar{\alpha}_{4q}^{(1)} e^{z_q \zeta} + \sum_{q=7}^{16} B_q \beta_{4q}^{(1)}(\zeta) \quad (87)$$

$$+ \mu_4^{(1)}(\zeta)$$

where $\beta_{jq}^{(0)}(\zeta)$, $\beta_{jq}^{(1)}(\zeta)$, $\mu_j^{(0)}(\zeta)$, and $\mu_j^{(1)}(\zeta)$, $j = 1, \ldots, 4$, $q = 7, \ldots, 16$ are, in general, polynomial functions in ζ which are derived from equations (80) through (83).

The mechanical boundary conditions for cantilever fixed at $x_1 = -a$ and with a proof mass at $x_1 = +a$ are given as:

$$u_j^{(n)} = 0 \qquad x_1 = -a \quad (88)$$

$$\begin{aligned} bT_{1j}^{(0)} &= Ma_j \qquad x_1 = a \\ b^2 T_{1j}^{(1)} &= \delta_{1j} M\left(\bar{x}_1 a_2 - \bar{x}_2 a_1\right) \end{aligned} \quad (89)$$

The electrical boundary conditions that will be considered here are:

$$\begin{aligned} \psi^{(0)} &= 0 \qquad \psi^{(1)} = 0 \qquad x_1 = -a \\ D_1^{(0)} &= 0 \qquad D_1^{(1)} = 0 \qquad x_1 = a \end{aligned} \quad (90)$$

With this, the following set of 16 equations results:

$$\sum_{q=1}^{6} B_q \bar{\alpha}_{jq}^{(0)} e^{-z_q r} + \sum_{q=7}^{16} B_q \beta_{jq}^{(0)}(-r) \quad (91)$$

$$= -\mu_j^{(0)}(-r)$$

$$\sum_{q=1}^{6} B_q \bar{\alpha}_{jq}^{(1)} e^{-z_q r} + \sum_{q=7}^{16} B_q \beta_{jq}^{(1)}(-r) \quad (92)$$

$$= -\mu_j^{(1)}(-r)$$

$$\sum_{q=1}^{6} B_q \bar{\alpha}_{4q}^{(0)} e^{-z_q r} + \sum_{q=7}^{16} B_q \beta_{4q}^{(0)}(-r) \quad (93)$$

$$= -\mu_4^{(0)}(-r)$$

$$\sum_{q=1}^{6} B_q \bar{\alpha}_{4q}^{(1)} e^{-z_q r} + \sum_{q=7}^{16} B_q \beta_{4q}^{(1)}(-r) \quad (94)$$

$$= -\mu_4^{(1)}(-r)$$

$$\sum_{q=1}^{6} B_q \bar{d}_{jq}^{(0)} e^{z_q r} + \sum_{q=7}^{16} B_q \theta_{jq}^{(0)}(r) = \quad (95)$$

$$-\chi_j^{(0)}(r) + \frac{M}{b} a_j$$

$$\sum_{q=1}^{6} B_q \bar{d}_{jq}^{(1)} e^{z_q r} + \sum_{q=7}^{16} B_q \theta_{jq}^{(1)}(r) = \quad (96)$$

$$-\chi_j^{(1)}(r) + \delta_{1j} \frac{M}{b^2}\left(\bar{x}_1 a_2 - \bar{x}_2 a_1\right)$$

$$\sum_{q=1}^{6} B_q \bar{d}_{4q}^{(0)} e^{z_q r} + \sum_{q=7}^{16} B_q \theta_{4q}^{(0)}(r) = \quad (97)$$

$$-\chi_4^{(0)}(r)$$

$$\sum_{q=1}^{6} B_q \bar{d}_{4q}^{(1)} e^{z_q r} + \sum_{q=7}^{16} B_q \theta_{4q}^{(1)}(r) = \quad (98)$$

$$-\chi_4^{(1)}(r)$$

In equations (91) through (98) the terms $\theta_{jq}^{(0)}(\zeta)$, $\theta_{jq}^{(1)}(\zeta)$, $\chi_j^{(0)}(\zeta)$, and $\chi_j^{(1)}(\zeta)$, $j = 1, \ldots, 4$, $q = 7, \ldots, 16$, are polynomial functions of ζ which are obtained by substituting the particular solution into equations (18) through (21). Equations (91) through (98) represent 16 equations in the 16 unknowns, B_q, $q = 1, \ldots, 16$ which can be represented as a 16X16 matrix equation, the solution of which yields the static deformation.

Figure (7) shows the static deflection of a ZnO-Si bimorph with an end mass structure under the action of a normal acceleration.

5 Acceleration Sensitivity

The acceleration sensitivity of the composite cantilever is determined by employing the plate expansion (1) in the linearized electroelastic equations of Baumhauer and Tiersten for small fields superposed on a bias [2]. The resulting equations are comparable to those obtained by Lee, et al [3] for homogeneous plates with initial stress. In the present development, the effects of piezoelectricity are neglected in the sensitivity calculations. This is a valid assumption since the effects of piezoelectric coupling should not have a great effect on the sensitivity. The first perturbation of the eigenvalue [4] is used to estimate the effect of normal acceleration on the fundamental flexural resonance of the cantilever. A detailed development of these methods will not be pursued here, rather the basic results are listed.

The zero and first order differential equations which are obtained can be written in the following form:

$$T^{(0)}_{L\gamma,L} + \tilde{T}^{(0)}_{L\gamma,L} = \rho_0^{(0)}\ddot{u}_\gamma^{(0)} + \rho_0^{(1)}\ddot{u}_\gamma^{(1)} \quad (99)$$

$$T^{(1)}_{L\gamma,L} - \frac{1}{b}T^{(0)}_{2\gamma} + \tilde{T}^{(1)}_{L\gamma,L} - \frac{1}{b}\tilde{T}^{(0)}_{2\gamma} = \rho_0^{(1)}\ddot{u}_\gamma^{(0)} + \rho_0^{(2)}\ddot{u}_\gamma^{(1)} \quad (100)$$

where $T^{(0)}_{L\gamma}$ and $T^{(1)}_{L\gamma}$ are linear terms which are defined by equations (18) through (21) and $\tilde{T}^{(0)}_{L\gamma}$ and $\tilde{T}^{(1)}_{L\gamma}$ represent non-linear terms which are functions of the static deformations, $w_N^{(0)}$ and $w_N^{(1)}$, under the applied loads and are given as:

$$\tilde{T}^{(0)}_{L\gamma} = \hat{c}^{(0)}_{L\gamma M\alpha}u^{(0)}_{\alpha,M} + \hat{c}^{(1)}_{L\gamma M\alpha}u^{(1)}_{\alpha,M} + \frac{1}{b}\hat{c}^{(0)}_{L\gamma 2\alpha}u^{(1)}_{\alpha} \quad (101)$$

$$\tilde{T}^{(1)}_{L\gamma} = \hat{c}^{(1)}_{L\gamma M\alpha}u^{(0)}_{\alpha,M} + \hat{c}^{(2)}_{L\gamma M\alpha}u^{(1)}_{\alpha,M} + \frac{1}{b}\hat{c}^{(1)}_{L\gamma 2\alpha}u^{(1)}_{\alpha} \quad (102)$$

where:

$$\hat{c}^{(n)}_{L\gamma M\alpha} = \Big[c^{(n)}_{L\gamma M\alpha KN} + c^{(n)}_{LMKN}\delta_{\gamma\alpha} + c^{(n)}_{L\gamma KM}\delta_{\alpha N} + c^{(n)}_{LKM\alpha}\delta_{\gamma N}\Big]w^{(0)}_{N,K} + \Big[c^{(n+1)}_{L\gamma M\alpha KN} + c^{(n+1)}_{LMKN}\delta_{\gamma\alpha} + c^{(n+1)}_{L\gamma KM}\delta_{\alpha N} + c^{(n+1)}_{LKM\alpha}\delta_{\gamma N}\Big]w^{(1)}_{N,K} + \frac{1}{b}\Big[c^{(n)}_{L\gamma M\alpha 2N} + c^{(n)}_{LM2N}\delta_{\gamma\alpha} + c^{(n)}_{L\gamma 2N}\delta_{\alpha N} + c^{(n)}_{L2M\alpha}\delta_{\gamma N}\Big]w^{(1)}_{N} \quad (103)$$

with $c^{(n)}_{L\gamma M\alpha KN}$ representing the composite plate third order elastic stiffnesses which are defined in the same way as the second order terms given by equation (22).

The first perturbation of the eigenvalue is obtained by first considering the m^{th} eigensolution at ω_m satisfying the linear equations:

$$T^{(0)m}_{L\gamma,L} + \rho_0^{(0)}\omega_m^2 u_\gamma^{(0)m} + \rho_0^{(1)}\omega_m^2 u_\gamma^{(1)m} = 0 \quad (104)$$

$$T^{(1)m}_{L\gamma,L} - \frac{1}{b}T^{(0)m}_{2\gamma} + \rho_0^{(1)}\omega_m^2 u_\gamma^{(0)m} + \rho_0^{(2)}\omega_m^2 u_\gamma^{(1)m} = 0 \quad (105)$$

and a near by perturbed solution at ω satisfying the non-linear equations:

$$T^{(0)}_{L\gamma,L} + \tilde{T}^{(0)}_{L\gamma,L} + \rho_0^{(0)}\omega^2 u_\gamma^{(0)} + \rho_0^{(1)}\omega^2 u_\gamma^{(1)} = 0 \quad (106)$$

$$T^{(1)}_{L\gamma,L} - \frac{1}{b}T^{(0)}_{2\gamma} + \tilde{T}^{(1)}_{L\gamma,L} - \frac{1}{b}\tilde{T}^{(0)}_{2\gamma} + \rho_0^{(1)}\omega^2 u_\gamma^{(0)} + \rho_0^{(2)}\omega^2 u_\gamma^{(1)} = 0 \quad (107)$$

Next, form the following products and integrate over the reference area:

$$\int_{A_0}\Big\{\Big[T^{(0)m}_{L\gamma,L} + \rho_0^{(0)}\omega_m^2 u_\gamma^{(0)m} + \rho_0^{(1)}\omega_m^2 u_\gamma^{(1)m}\Big]u_\gamma^{(0)} - \Big[T^{(0)}_{L\gamma,L} + \tilde{T}^{(0)}_{L\gamma,L} + \rho_0^{(0)}\omega^2 u_\gamma^{(0)} + \rho_0^{(1)}\omega^2 u_\gamma^{(1)}\Big]u_\gamma^{(0)m} + \Big[T^{(1)m}_{L\gamma,L} - \frac{1}{b}T^{(0)m}_{2\gamma} + \rho_0^{(1)}\omega_m^2 u_\gamma^{(0)m} + \rho_0^{(2)}\omega_m^2 u_\gamma^{(1)m}\Big]u_\gamma^{(1)} - \Big[T^{(1)}_{L\gamma,L} - \frac{1}{b}T^{(0)}_{2\gamma} + \tilde{T}^{(1)}_{L\gamma,L} - \frac{1}{b}\tilde{T}^{(0)}_{2\gamma} + \rho_0^{(1)}\omega^2 u_\gamma^{(0)} + \rho_0^{(2)}\omega^2 u_\gamma^{(1)}\Big]u_\gamma^{(1)m}\Big\}dA_0 = 0 \quad (108)$$

Equation (107), after some manipulations, can be rewritten as:

$$(\omega_m^2 - \omega^2) \int_{A_0} [\rho_0^{(0)} u_\gamma^{(0)m} u_\gamma^{(0)} + \rho_0^{(1)} (u_\gamma^{(0)} u_\gamma^{(1)m} + u_\gamma^{(0)m} u_\gamma^{(1)}) + \rho_0^{(2)} u_\gamma^{(1)m} u_\gamma^{(1)}] dA_0 = \int_{A_0} [\tilde{T}_{L\gamma,L}^{(0)} u_\gamma^{(0)m} + \tilde{T}_{L\gamma,L}^{(1)} u_\gamma^{(1)m} - \frac{1}{b} \tilde{T}_{2\gamma}^{(0)} u_\gamma^{(1)m}] dA_0 + \int_{C_0} [T_{L\gamma}^{(0)} u_\gamma^{(0)m} - T_{L\gamma}^{(0)m} u_\gamma^{(0)} + T_{L\gamma}^{(1)} u_\gamma^{(1)m} - T_{L\gamma}^{(0)m} u_\gamma^{(0)}] dC_0 \quad (109)$$

Equation (108) is simplified with the aid of the following assumptions and definitions:

$$\Delta\omega = \omega_m - \omega \qquad |\Delta\omega| \ll \omega_m$$
$$\Delta u_\gamma = u_\gamma^{(n)m} - u_\gamma^{(n)} \qquad |\Delta u_\gamma| \ll u_\gamma^{(n)m} \quad (110)$$

which leads to:

$$\Delta\omega \cong \frac{1}{2\omega_m} \int_{A_0} [\tilde{T}_{L\gamma,L}^{(0)} \bar{u}_\gamma^{(0)m} + \tilde{T}_{L\gamma,L}^{(1)} \bar{u}_\gamma^{(1)m} - \frac{1}{b} \tilde{T}_{2\gamma}^{(0)} \bar{u}_\gamma^{(1)m}] dA_0 \quad (111)$$

where:

$$\bar{u}_\gamma^{(0)} = \frac{u_\gamma^{(0)m}}{N_{(m)}} \qquad \bar{u}_\gamma^{(1)} = \frac{u_\gamma^{(1)m}}{N_{(m)}} \quad (112)$$

are the components of the normalized mode shape of the m^{th} eigen-mode. Equation (111) gives the small frequency change of the m^{th} eigen-mode under the applied static bias. With this the normal acceleration sensitivity of the cantilever is defined as:

$$S = \frac{\Delta\omega}{\omega a_2} \quad (113)$$

Figure (8) shows the effect of crystal orientation of one layer on the sensitivity of a Quartz-Si-Quartz sandwich. In this study, the Quartz crystal in layer 1 is rotated about the x_1 axis while holding the axes of the other two layers constant.

6 Conclusion

The developments presented here offer accurate approximate methods for the analysis and design of a very important class of micro-structures. The dynamic and static solutions for the composite are easily obtained without complicated layer-wise analysis. The static solution given here, while complicated, can be constructed in closed form by following the steps outlined and using the equations listed in appendix A. Static solutions of this nature are also important for the design of micro-actuators.

Presently, many of the resonant MEMS devices are constructed of pure Silicon. Investigations into the behavior of composite plates shows that a wider range of behavior can be obtained when layers of materials with differing mechanical properties are used. The macroscopic properties of the composites can be adjusted by varying the proportions and crystallographic orientations of the constituents, giving rise to more complex behavior than can be obtained from homogeneous Silicon.

Acknowledgments

This work was supported by The National Research Council under which the author is presently a post-doctoral associate.

Appendix A

Specific Definitions For Constants Appearing In The Particular Static Solution

This section lists without discussion the definitions of constants which go into the calculation of the quantities that make up the particular static solution.

The terms Γ, Δ_k, and Ψ_k, are defined as:

$$\Gamma = -\frac{bF_2^{(0)}}{24\bar{c}_{11k1}^{(1)} f_k - 24c_{1111}^{(2)} + 2\bar{c}_{11k2}^{(1)} h_k} - \frac{bF_2^{(0)}}{24e_{111}^{(1)} + 2e_{211}^{(1)} h_4} \quad (A.1)$$

$$\Delta_k = 4f_k\Gamma \qquad \Delta_4 = 4f_4\Gamma \quad (A.2)$$

$$\Psi_k = h_k\Gamma \qquad \Psi_4 = h_4\Gamma \quad (A.3)$$

in (A.1)-(A.3) the terms h_k and f_k are given as:

$$h_k = 12\gamma_{km}\left(\hat{c}_{2m11}^{(1)} - \hat{c}_{2mn1}^{(0)} f_n\right) \quad (A.4)$$

$$h_4 = \frac{1}{\Delta_\varepsilon}\Big[12\Big(\bar{e}^{(0)}_{2k1}\varepsilon^{(0)}_{11} - \bar{e}^{(0)}_{1k1}\varepsilon^{(0)}_{12}\Big)f_k + \Big(\bar{e}^{(0)}_{2k2}\varepsilon^{(0)}_{11} - \bar{e}^{(0)}_{1k2}\varepsilon^{(0)}_{12}\Big)h_k - 12\,(e^{(1)}_{211}\varepsilon^{(0)}_{11} - e^{(1)}_{111}\varepsilon^{(0)}_{12})\Big] \quad \text{(A.5)}$$

$$f_k = \delta_{ka}\nu_{ab}\Big(\hat{c}^{(1)}_{1b11} - \hat{c}^{(0)}_{1bk2}\gamma_{km}\hat{c}^{(1)}_{2m11}\Big) \quad \text{(A.6)}$$

$$f_4 = \frac{1}{\Delta_\varepsilon}\Big[\Big(\bar{e}^{(0)}_{1k1}\varepsilon^{(0)}_{22} - \bar{e}^{(0)}_{2k1}\varepsilon^{(0)}_{12}\Big)f_k + \frac{1}{12}\Big(\bar{e}^{(0)}_{1k2}\varepsilon^{(0)}_{22} - \bar{e}^{(0)}_{2k2}\varepsilon^{(0)}_{12}\Big)h_k - (e^{(1)}_{111}\varepsilon^{(0)}_{22} - e^{(1)}_{211}\varepsilon^{(0)}_{12})\Big] \quad \text{(A.7)}$$

The quantities γ_{km} and ν_{ab} are defined such that:

$$\gamma_{km}\hat{c}^{(0)}_{2mj2} = \delta_{kj} \qquad \nu_{ac}\tilde{c}^{(0)}_{1cb1} = \delta_{ab} \quad \text{(A.8)}$$

In equation (A.8), and in all subsequent equations, the subscripts a,b,c range over 1 and 3 only. The terms $\tilde{c}^{(0)}_{1ab1}$ and $\hat{c}^{(n)}_{ijkl}$ are effective elastic constants which are defined as:

$$\tilde{c}^{(0)}_{1ab1} = \hat{c}^{(0)}_{1ab1} - \hat{c}^{(0)}_{1ak2}\gamma_{km}\hat{c}^{(0)}_{2mb1} \quad \text{(A.9)}$$

$$\hat{c}^{(0)}_{ijkl} = \bar{c}^{(0)}_{ijkl} + \frac{1}{\Delta_\varepsilon}\Big[\varepsilon^{(0)}_{22}\bar{e}^{(0)}_{1ij}\bar{e}^{(0)}_{1lk} - \varepsilon^{(0)}_{12}\,(\bar{e}^{(0)}_{1ij}\bar{e}^{(0)}_{2lk} + \bar{e}^{(0)}_{2ij}\bar{e}^{(0)}_{1lk}) + \varepsilon^{(0)}_{11}\bar{e}^{(0)}_{2ij}\bar{e}^{(0)}_{2lk}\Big] \quad \text{(A.10)}$$

$$\hat{c}^{(1)}_{ij11} = \bar{c}^{(1)}_{ij11} + \frac{1}{\Delta_\varepsilon}\Big[\varepsilon^{(0)}_{22}\bar{e}^{(0)}_{1ij}e^{(1)}_{111} - \varepsilon^{(0)}_{12}\Big(\bar{e}^{(0)}_{1ij}e^{(1)}_{211} + \bar{e}^{(0)}_{2ij}e^{(1)}_{111}\Big) + \varepsilon^{(0)}_{11}\bar{e}^{(0)}_{2ij}e^{(0)}_{211}\Big] \quad \text{(A.11)}$$

where:

$$\Delta_\varepsilon = \varepsilon^{(0)}_{11}\varepsilon^{(0)}_{22} - \varepsilon^{(0)2}_{12} \quad \text{(A.12)}$$

The terms H_k, Λ_k, M_k, and N_k are given as:

$$H_k = \frac{1}{2}\delta_{ka}\nu_{ab}\Big(\hat{q}_b - \hat{c}^{(0)}_{1bk2}\gamma_{km}\hat{t}_m\Big) \quad \text{(A.13)}$$

$$H_4 = \frac{1}{\Delta_\varepsilon}\Big[\Big(\varepsilon^{(0)}_{22}\bar{e}^{(0)}_{1k1} - \varepsilon^{(0)}_{12}\bar{e}^{(0)}_{2k1}\Big)H_k + \frac{1}{2}\Big(\varepsilon^{(0)}_{22}\bar{e}^{(0)}_{1k2} - \varepsilon^{(0)}_{12}\bar{e}^{(0)}_{2k2}\Big)M_k - \frac{1}{2}(\varepsilon^{(0)}_{22}q_4 - \varepsilon^{(0)}_{12}t_4)\Big] \quad \text{(A.14)}$$

$$\Lambda_k = \frac{1}{2}\delta_{ka}\nu_{ab}\Big(\hat{r}_b - \hat{c}^{(0)}_{1bk2}\gamma_{km}\hat{v}_m\Big) \quad \text{(A.15)}$$

$$\Lambda_4 = \frac{1}{\Delta_\varepsilon}\Big[\Big(\varepsilon^{(0)}_{22}\bar{e}^{(0)}_{1k1} - \varepsilon^{(0)}_{12}\bar{e}^{(0)}_{2k1}\Big)\Lambda_k + \frac{1}{2}\Big(\varepsilon^{(0)}_{22}\bar{e}^{(0)}_{1k2} - \varepsilon^{(0)}_{12}\bar{e}^{(0)}_{2k2}\Big)N_k - \frac{1}{2}(\varepsilon^{(0)}_{22}r_4 - \varepsilon^{(0)}_{12}v_4)\Big] \quad \text{(A.16)}$$

$$M_k = \gamma_{km}\Big[\hat{t}_m - \hat{c}^{(0)}_{2mb1}\nu_{ba}\Big(\hat{q}_a - \hat{c}^{(0)}_{1ak2}\gamma_{km}\hat{t}_m\Big)\Big] \quad \text{(A.17)}$$

$$M_4 = \frac{1}{\Delta_\varepsilon}\Big[2\Big(\varepsilon^{(0)}_{11}\bar{e}^{(0)}_{2k1} - \varepsilon^{(0)}_{12}\bar{e}^{(0)}_{1k1}\Big)H_k + \Big(\varepsilon^{(0)}_{11}\bar{e}^{(0)}_{2k2} - \varepsilon^{(0)}_{12}\bar{e}^{(0)}_{1k2}\Big)M_k - (\varepsilon^{(0)}_{11}t_4 - \varepsilon^{(0)}_{12}q_4)\Big] \quad \text{(A.18)}$$

$$N_k = \gamma_{km}\Big[\hat{v}_m - \hat{c}^{(0)}_{2mb1}\nu_{ba}\Big(\hat{r}_a - \hat{c}^{(0)}_{1ak2}\gamma_{km}\hat{v}_m\Big)\Big] \quad \text{(A.19)}$$

$$N_4 = \frac{1}{\Delta_\varepsilon}\Big[2\Big(\varepsilon^{(0)}_{11}\bar{e}^{(0)}_{2k1} - \varepsilon^{(0)}_{12}\bar{e}^{(0)}_{1k1}\Big)\Lambda_k + \Big(\varepsilon^{(0)}_{11}\bar{e}^{(0)}_{2k2} - \varepsilon^{(0)}_{12}\bar{e}^{(0)}_{1k2}\Big)N_k - (\varepsilon^{(0)}_{11}v_4 - \varepsilon^{(0)}_{12}r_4)\Big] \quad \text{(A.20)}$$

In equations (A.13)-(A.20) the quantities $\hat{q}_k$, q_k, $\hat{r}_k$, r_k, $\hat{t}_k$, t_k, $\hat{v}_k$, and v_k are defined as:

$$\hat{q}_k = q_k + \frac{1}{\Delta_\varepsilon}\left[\left(\bar{e}^{(0)}_{11k}\varepsilon^{(0)}_{22} - \bar{e}^{(0)}_{21k}\varepsilon^{(0)}_{12}\right)q_4 + \left(\bar{e}^{(0)}_{21k}\varepsilon^{(0)}_{11} - \bar{e}^{(0)}_{11k}\varepsilon^{(0)}_{12}\right)t_4\right] \quad \text{(A.21)}$$

$$\hat{r}_k = r_k + \frac{1}{\Delta_\varepsilon}\left[\left(\bar{e}^{(0)}_{11k}\varepsilon^{(0)}_{22} - \bar{e}^{(0)}_{21k}\varepsilon^{(0)}_{12}\right)r_4 + \left(\bar{e}^{(0)}_{21k}\varepsilon^{(0)}_{11} - \bar{e}^{(0)}_{11k}\varepsilon^{(0)}_{12}\right)v_4\right] \quad \text{(A.22)}$$

$$\hat{t}_k = t_k + \frac{1}{\Delta_\varepsilon}\left[\left(\bar{e}^{(0)}_{12k}\varepsilon^{(0)}_{22} - \bar{e}^{(0)}_{22k}\varepsilon^{(0)}_{12}\right)q_4 + \left(\bar{e}^{(0)}_{22k}\varepsilon^{(0)}_{11} - \bar{e}^{(0)}_{12k}\varepsilon^{(0)}_{12}\right)t_4\right] \quad \text{(A.23)}$$

$$\hat{v}_k = v_k + \frac{1}{\Delta_\varepsilon}\left[\left(\bar{e}^{(0)}_{12k}\varepsilon^{(0)}_{22} - \bar{e}^{(0)}_{22k}\varepsilon^{(0)}_{12}\right)r_4 + \left(\bar{e}^{(0)}_{22k}\varepsilon^{(0)}_{11} - \bar{e}^{(0)}_{12k}\varepsilon^{(0)}_{12}\right)v_4\right] \quad \text{(A.24)}$$

$$q_j = -bF^{(0)}_j - 2\left(\bar{c}^{(1)}_{1jk1}h_k + \bar{e}^{(1)}_{11j}h_4\right)\Gamma \quad \text{(A.25)}$$

$$r_j = 6\bar{c}^{(1)}_{1j11} \quad \text{(A.26)}$$

$$t_j = 2\left[12\bar{c}^{(1)}_{1jk1}f_k - 12c^{(2)}_{1j11} + (\bar{c}^{(1)}_{1jk2} - \bar{c}^{(1)}_{2jk1})h_k + 12e^{(1)}_{11j}f_4 + \left(e^{(1)}_{21j} - \bar{e}^{(1)}_{12j}\right)h_4\right]\Gamma \quad \text{(A.27)}$$

$$v_j = 6\bar{c}^{(1)}_{2j11} \quad \text{(A.28)}$$

$$q_4 = -bD^{(0)} - 2\left(e^{(1)}_{1k1}h_k - \varepsilon^{(1)}_{11}h_4\right)\Gamma \quad \text{(A.29)}$$

$$r_4 = 6e^{(2)}_{111} \quad \text{(A.30)}$$

$$t_4 = 2\left[12\bar{e}^{(1)}_{1k1}f_k - 12e^{(2)}_{111} - 12\varepsilon^{(1)}_{11}f_4 + \left(\bar{e}^{(1)}_{1k2} - e^{(1)}_{2k1}\right)h_k\right]\Gamma \quad \text{(A.31)}$$

$$v_4 = 6e^{(1)}_{211} \quad \text{(A.32)}$$

The terms Σ_k, Ξ_k, Π_k, K_{km}, and $\bar{K}_k$ are given as:

$$\Sigma_k = \gamma'_{km}\lambda'_m \quad \text{(A.33)}$$

$$\Sigma_4 = \frac{1}{\varepsilon^{(0)}_{22}}\left(\bar{e}^{(0)}_{2k2}\gamma'_{km}\lambda'_m - \lambda_4\right) \quad \text{(A.34)}$$

$$\Xi_k = \gamma'_{km}\tau'_m \quad \text{(A.35)}$$

$$\Xi_4 = \frac{1}{\varepsilon^{(0)}_{22}}\left(\bar{e}^{(0)}_{2k2}\gamma'_{km}\tau'_m - \tau_4\right) \quad \text{(A.36)}$$

$$\Pi_k = 2\gamma'_{km}\tilde{c}^{(1)}_{2m11} \quad \text{(A.37)}$$

$$\Pi_4 = \frac{2}{\varepsilon^{(0)}_{22}}\left(\bar{e}^{(0)}_{2k2}\gamma'_{km}\tilde{c}^{(1)}_{2m11} - e^{(1)}_{211}\right) \quad \text{(A.38)}$$

$$K_{km} = -\gamma'_{km}\tilde{c}^{(0)}_{2nm1} \quad \text{(A.39)}$$

$$K_{4m} = -\frac{1}{\varepsilon^{(0)}_{22}}\left(\bar{e}^{(0)}_{2k2}\gamma'_{kn}\tilde{c}^{(0)}_{2nm1} - \bar{e}^{(0)}_{2m1}\right) \quad \text{(A.40)}$$

$$\bar{K}_k = -\gamma'_{km}\tilde{e}^{(0)}_{12m} \quad \text{(A.41)}$$

$$\bar{K}_4 = -\frac{1}{\varepsilon^{(0)}_{22}}\left(\bar{e}^{(0)}_{2k2}\gamma'_{kn}\tilde{e}^{(0)}_{12m} + \varepsilon^{(0)}_{12}\right) \quad \text{(A.42)}$$

The quantity γ'_{km} is defined such that:

$$\gamma'_{km}\tilde{c}^{(0)}_{2mj2} = \delta_{kj} \quad \text{(A.43)}$$

The terms $\tilde{c}^{(0)}_{ijkl}$, $\tilde{c}^{(1)}_{2j11}$, and $\tilde{e}^{(0)}_{12j}$ are effective constants which are defined as:

$$\tilde{c}^{(0)}_{ijkl} = \bar{c}^{(0)}_{ijkl} + \frac{\bar{e}^{(0)}_{2ij}\bar{e}^{(0)}_{2lk}}{\varepsilon^{(0)}_{22}} \quad \text{(A.44)}$$

$$\tilde{c}^{(1)}_{2j11} = \bar{c}^{(1)}_{2j11} + \frac{\bar{e}^{(0)}_{22j}e^{(1)}_{221}}{\varepsilon^{(0)}_{22}} \quad \text{(A.45)}$$

$$\tilde{e}^{(0)}_{12j} = \bar{e}^{(0)}_{12j} + \frac{\bar{e}^{(0)}_{22j}\varepsilon^{(0)}_{12}}{\varepsilon^{(0)}_{22}} \quad \text{(A.46)}$$

Finally, the terms λ'_j, λ_j, τ'_j, and τ_j are defined as:

$$\lambda'_j = \lambda_j + \frac{\bar{e}^{(0)}_{22j}}{\varepsilon^{(0)}_{22}}\lambda_4 \quad \text{(A.47)}$$

$$\tau'_j = \tau_j + \frac{\bar{e}^{(0)}_{22j}}{\varepsilon^{(0)}_{22}}\tau_4 \quad \text{(A.48)}$$

$$\lambda_j = bF_j^{(1)} + 2\bar{c}_{1jk1}^{(1)} H_k + 2c_{1jk1}^{(2)} \Gamma + 2e_{11j}^{(1)} H_4 + 2e_{11j}^{(2)} h_4 \Gamma + (\bar{c}_{1jk2}^{(1)} - \bar{c}_{2jk1}^{(1)}) M_k + \left(e_{21j}^{(1)} - \bar{e}_{12j}^{(1)}\right) M_4 \quad \text{(A.49)}$$

$$\tau_j = 2\bar{c}_{1jk1}^{(1)} \Lambda_k - 6c_{1j11}^{(2)} + 2e_{11j}^{(1)} \Lambda_4 + (\bar{c}_{1jk2}^{(1)} - \bar{c}_{2jk1}^{(1)}) N_k + \left(e_{21j}^{(1)} - \bar{e}_{12j}^{(1)}\right) N_4 \quad \text{(A.50)}$$

$$\lambda_4 = bD^{(1)} + 2\bar{e}_{1k1}^{(1)} H_k + 2e_{1k1}^{(2)} h_k \Gamma - 2\varepsilon_{11}^{(1)} H_4 - 2\varepsilon_{11}^{(2)} h_4 \Gamma + \left(\bar{e}_{1k2}^{(1)} - e_{2k1}^{(1)}\right) M_k \quad \text{(A.51)}$$

$$\tau_4 = 2\bar{e}_{1k1} \Lambda_k - 6e_{111}^{(2)} - 2\varepsilon_{11}^{(1)} \Lambda_4 + \left(\bar{e}_{1k2}^{(1)} - e_{2k1}^{(1)}\right) N_k \quad \text{(A.52)}$$

References

1.) R.D. Mindlin, "High Frequency Vibrations Of Piezoelectric Crystal Plates", Int. J. Solids and Structures, 1972, Vol. 8, pp. 895-906.

2.) J.C. Baumhauer and H.F. Tiersten, "Nonlinear Electroelastic Equations for Small Fields Superposed on a Bias", J. Acoust. Soc. Am., 54(4), 1973, pp. 1017-1034.

3.) P.C.Y. Lee, Y.S. Wang, and X. Markenscoff, "High Frequency Vibrations of Crystal Plates Under Initial Stresses", J. Acoust. Soc. Am., 57(1), Jan. 1975, pp. 95-105.

4.) H. F. Tiersten, "Perturbation Theory For Linear Electroelastic Equations for Small Fields Superposed On a Bias", J. Acoust. Soc. Am.,64(3) Sept., 1978, pp. 832-837.

5.) J.T. Stewart and Y.K. Yong, "Exact Analysis of The Propagation of Acoustic Waves In Multilayered Anisotropic Piezoelectric Plates", IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 41(3), May 1994, pp.375-390.

6.) Y.K. Yong and J.T. Stewart, "A Laminated Plate Theory For High Frequency, Piezoelectric Thin Film Resonators", J. Appl. Physics, 74(5), Sept. 1993, pp. 3028-3046.

7.) R.M. Jones, Mechanics of Composite Materials, New York, Hemisphere Publishing Corp., 1975.

8.) R.A. Buser and N.F. DeRooij, "Resonant Silicon Structures", Sensors and Actuators, 17, 1989, pp. 145-154.

9.) M.A. Meldrum, "Application of Vibrating Beam Technology to Digital Acceleration Measurement", Sensors and Actuators, A21-A23 (1990), pp. 377-380.

10.) H. Seidel and L. Csepregi, "Design Optimization For Cantilever-Type Accelerometers",Sensors and Actuators, 6(1984), pp. 81-92.

11.) F.R. Blom, S. Bouwstra, J.H.J. Fluitman, and M. Elwenspoek, "Resonating Silicon Beam Force Sensor", Sensors and Actuators, 17 (1989), pp. 513-519.

12.) K. Ikeda, H. Kuwayama, T. Kobayashi, T. Wantanabe, T. Nishikawa, T. Yoshida, K. Harada, "Silicon Pressure Sensor Integrates Resonant Strain Gauge On Diaphragm", Sensors and Actuators, A21-A23 (1990), PP. 146-150.

Figures

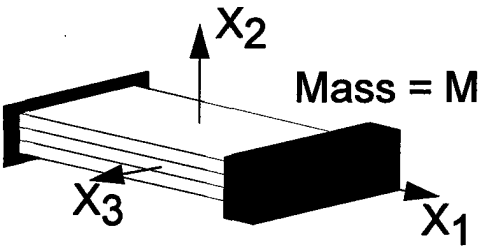


Figure 1.) Orientation of basic cantilever structure.

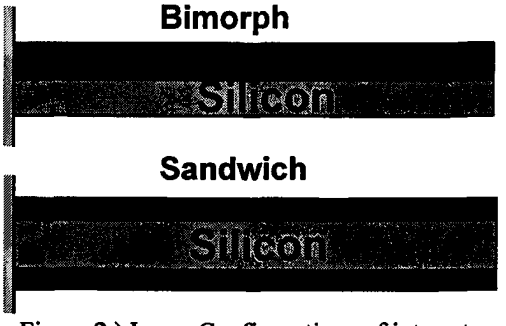


Figure 2.) Layer Configurations of interest.

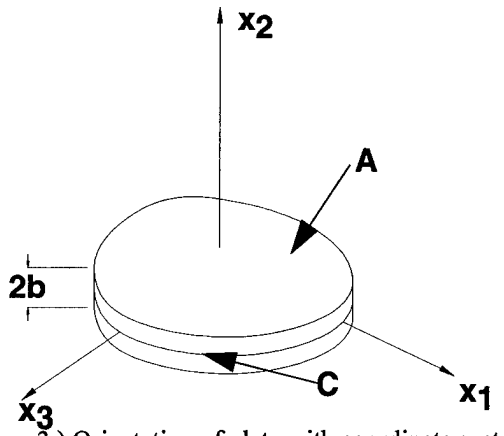


Figure 3.) Orientation of plate with coordinate system

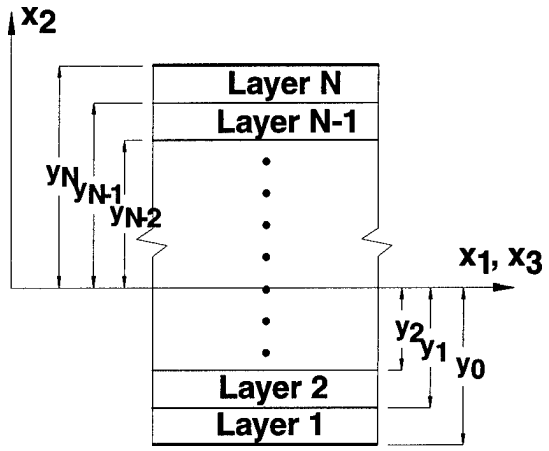


Figure 4.) Cross-sectional view of layered plate.

Figure 5.a.)

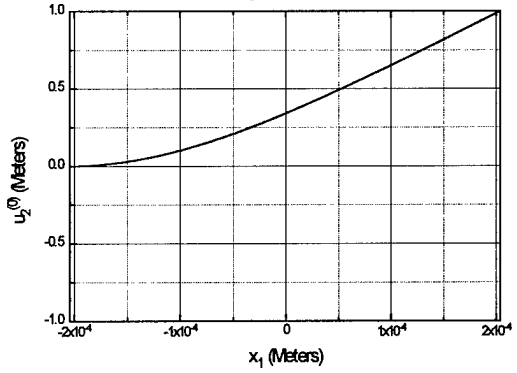


Figure 5.b.)

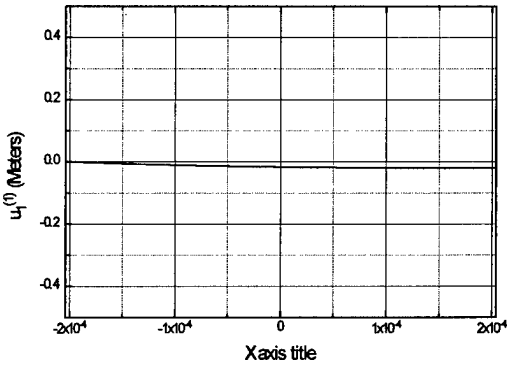


Figure 5.c.)

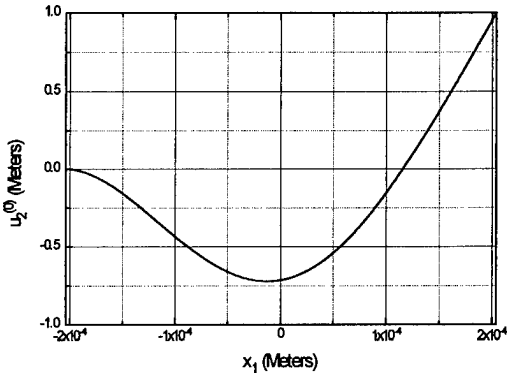


Figure 5.d.)

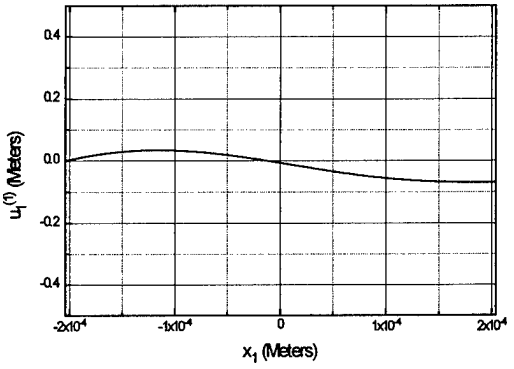


Figure 5.e.)

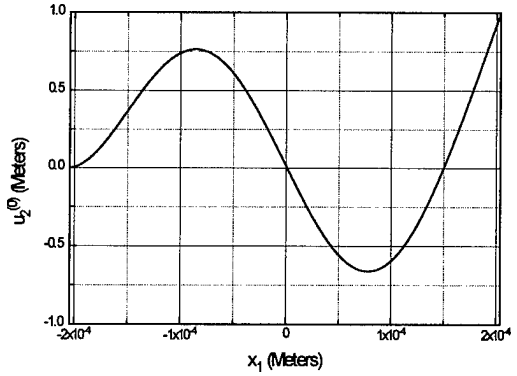
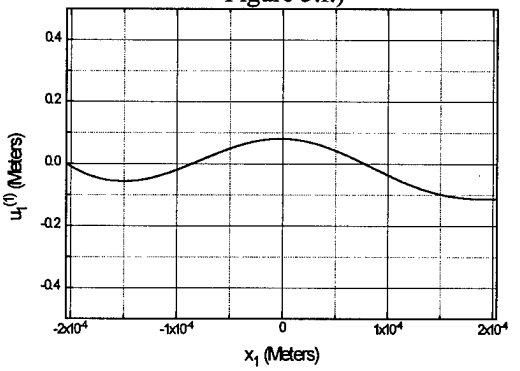
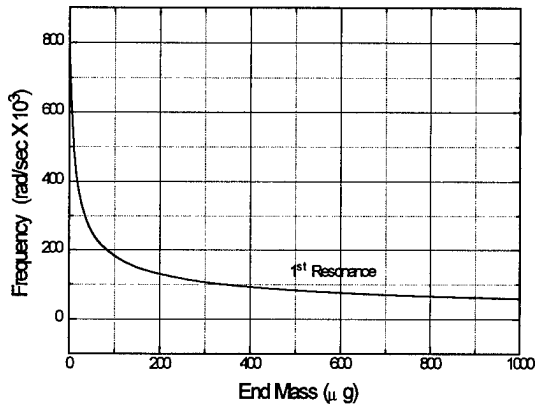


Figure 5.f.)



ZnO $h_2 = 7.2\mu m$
Si $h_1 = 4.8\mu m$
M = 0 $2a = 408\mu m$

Figure 5.) First three resonances of a ZnO-Si bimorph: flexural and shear deformations



Quartz $h_2 = 10mm$ $\bar{x}_1 = 100\ \mu m$
Si $h_1 = 10\mu m$ $\bar{x}_2 = 0$
$2a = 400\mu m$

Figure 6.) Effect of end mass on the fundamental flexural resonance for a Quartz-Si bimorph.

Figure 7.a.)

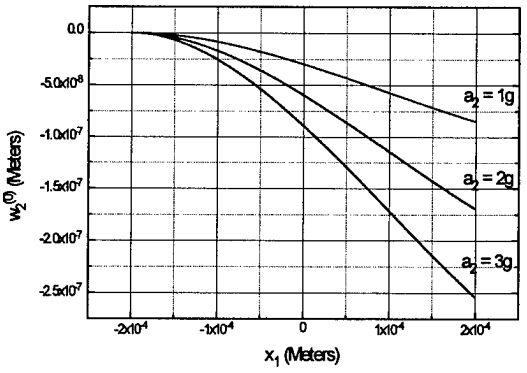


Figure 7.b.)

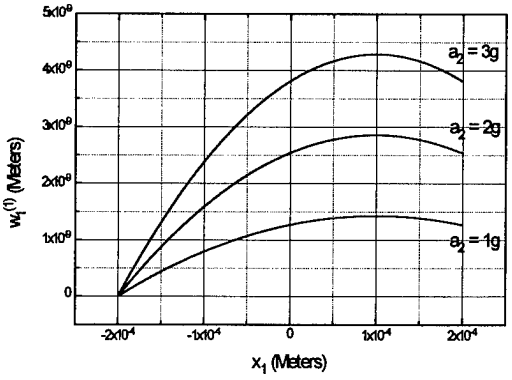


Figure 7.c.)

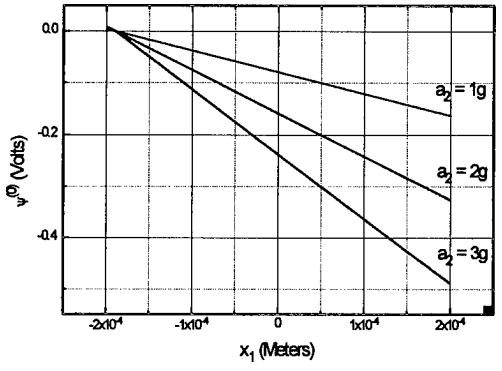


Figure 7.d.)

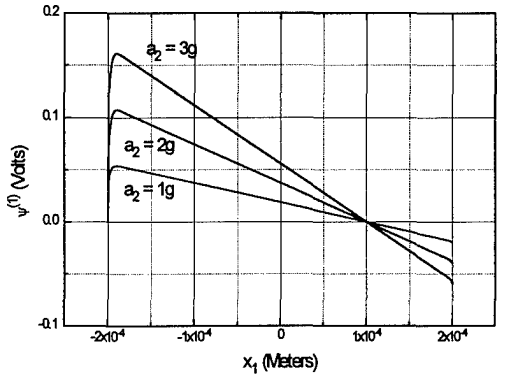


Figure 7.e.)

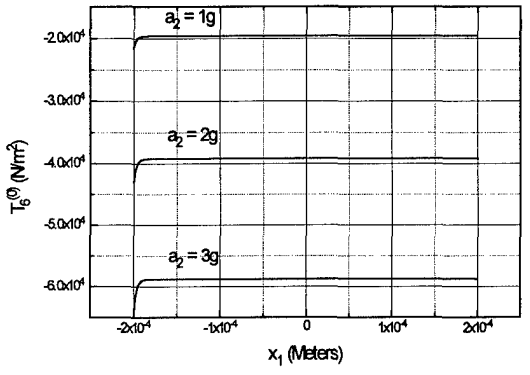


Figure 7.f.)

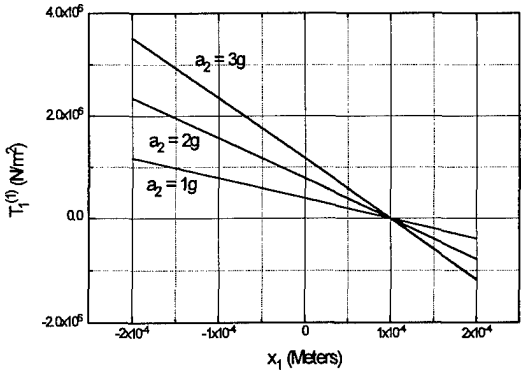


Figure 7.g.)

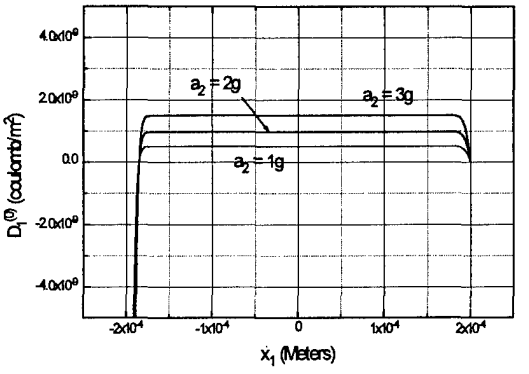
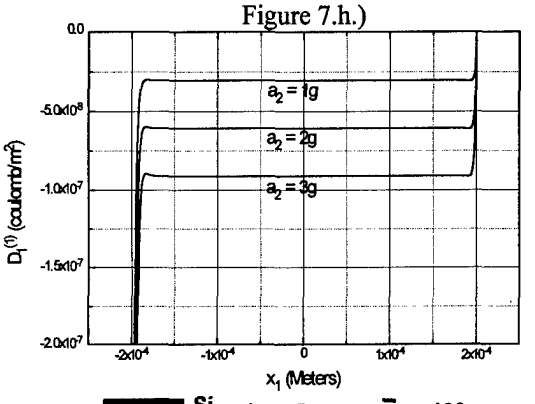
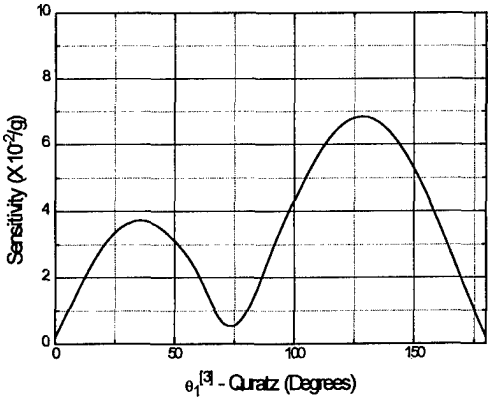


Figure 7.h.)



Si $h_2 = 5\mu m$ $\bar{x}_1 = 100\ \mu m$
ZnO $h_1 = 5\mu m$ $\bar{x}_2 = 0$
$2a = 400\mu m$ $M = 10{,}000\ \mu g$

Figure 7.) Static solutions under acceleration.



$2a = 400.0\mu m$
$M = 500\ \mu g$

Quartz $h_3 = 2.5\mu m$ $\bar{x}_1 = 40\ \mu m$
Si $h_2 = 5.0\mu m$ $\bar{x}_2 = 0$
Quartz $h_1 = 2.5\mu m$

Figure 8.) Acceleration sensitivity of a Quartz-Si-Quartz sandwich as a function of θ_1 in layer 1.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

CALCULATION OF THE OPTIMAL CLIP DIMENSIONING TO MINIMIZE THE INFLUENCE OF FABRICATION IMPERFECTIONS ON THE ACCELERATION SENSITIVITY OF SC–CUT QUARTZ RESONATORS WITH STIFFENED RECTANGULAR SUPPORT SYSTEMS

B.J. Lwo and H.F. Tiersten

Department of Mechanical Engineering, Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute, Troy, New York 12180–3590

Abstract

Recent work on the determination of optimal orientations and aspect ratios of SC–cut quartz resonators stiffened by rectangular quartz plates has shown that the influence of fabrication imperfections is minimized when two clips of reasonable size are used on each of the large sides. In this work existing computer programs are used in the calculation of the influence of the spacing between the two clips and the stiffnesses of the clips on the sensitivity to fabrication imperfections in order to find optimal conditions. Both contoured and flat plate trapped energy resonators are treated. The smaller dimension of the support rectangle is determined by a criterion associated with the energy in the mode shape for each particular harmonic. A summary of recommendations based on all the calculations is presented.

1. Introduction

It has been shown[1,2] that the resultant acceleration sensitivity of contoured quartz resonators vanishes for perfect symmetry of the resonator with respect to the support system. It has further been shown[1] that any loss of symmetry in the combined resonator plus support configuration results in a linear increase in the acceleration sensitivity. Since it is impossible to construct a perfectly symmetric resonator plus support configuration in practice, a stiffened structure has been under investigation. The structure consists of the active biconvex quartz resonator attached to identical top and bottom quartz cover plates by means of small sidewalls around the periphery. The mounting clips are attached to the top and bottom cover plates without touching the active plate.

In recent work[3] it was shown that the influence of fabrication imperfections on the acceleration sensitivity of SC–cut quartz resonators stiffened by rectangular quartz cover plates supported by clips was minimized for an aspect ratio of 1.5 for all resonators considered. In addition it was shown that the optimal orientation is very nearly the same for all the contoured resonators considered, but that it varies significantly for the flat plate trapped energy resonators depending primarily on the size of the electrodes. It was further shown that the use of two clips of reasonable size on the large sides gives better results than one large clip with one clip used on each of the small sides.

In this work existing computer programs[4,5] are used in the calculation of the influence of the spacing between the two clips, the relative sizes of the clips on the small and large sides and the thickness of the clips on the sensitivity to fabrication imperfections. Results are presented for a number of SC–cut contoured and flat plate trapped energy resonators. Finally, a list of recommendations based on all our calculations is presented. However, it should be noted that since all our calculations employ a particular clip configuration, the results could change somewhat for significantly different clip configurations.

2. Preliminary Considerations

For purely elastic nonlinearities the equation for the perturbation in eigenfrequency[6] may be written in the form given in Eq.(2.1) of Ref.4, which is the form

$$\Delta_\mu = H_\mu/2\omega_\mu\,, \quad \omega = \omega_\mu - \Delta_\mu\,, \tag{2.1}$$

where ω_μ and ω are the unperturbed and perturbed eigenfrequencies, respectively. It has further been shown that H_μ may be written in the form given in Eq.(2.9) of Ref.4, which is the form

$$H_\mu = -\int_V \hat{c}_{L\gamma M\alpha} g^\mu_{\gamma,L} g^\mu_{\alpha,M}\, dV\,, \tag{2.2}$$

where $\hat{c}_{L\gamma M\alpha}$ is defined in Eq.(2.4) of Ref.4 and all other quantities are defined in Sec.2 of Ref.4. Since g^μ_α denotes the normalized mode shape in the resonator and from Eq.(2.4) of Ref.4 $\hat{c}_{L\gamma M\alpha}$ depends on the biasing state, H_μ can be evaluated when the mode shape in the resonator and biasing state are known.

A schematic diagram of a contoured crystal resonator along with the associated coordinate system is shown in Fig.1. It has been shown that the eigensolution for the dominant displacement of the harmonic modes in a contoured quartz resonator can be written in the form[7]

0-7803-1945-1/94 $3.00 © 1994 IEEE

$$u_{1n} = \sin\frac{n\pi X_2}{2h}\, u_{noo}\, e^{i\omega_{noo}t},$$

$$n = 1,3,5,\ldots, \qquad (2.3)$$

where

$$u_{noo} = e^{-\alpha_n \frac{\hat{X}_1^2}{2}} e^{-\beta_n \frac{\hat{X}_3^2}{2}}, \qquad (2.4)$$

in which

$$\alpha_n^2 = \frac{\alpha^2\pi^2\hat{c}(1)}{8Rh_o^3 M_n'}, \quad \beta_n^2 = \frac{n^2\pi^2\hat{c}(1)}{8Rh_o^3 P_n'} \qquad (2.5)$$

and M_n' and P_n' are involved lengthy expressions that are defined in Sec.II of Ref.7 along with $\hat{X}_1$ and $\hat{X}_3$, which are orthogonal directions in the plane of the plate for which the scalar differential equation for the nth anharmonic family of modes does not contain mixed derivatives and 2R denotes the radii of the spherical contours. The eigenfrequencies corresponding to the eigensolutions for the harmonic modes are given by

$$\omega_{noo}^2 = \frac{n^2\pi^2\hat{c}(1)}{4h_o^2\rho}\left[1 + \frac{1}{n\pi}\left[\frac{2h_o}{R}\right]^{1/2} \cdot \left[\sqrt{\frac{M_n'}{\hat{c}(1)}} + \sqrt{\frac{P_n'}{\hat{c}(1)}}\right]\right], \qquad (2.6)$$

where $\hat{c}^{(1)}$ is defined in Eq.(78) of Ref.7.

In the in–plane case only the dominant displacement field in (2.3) is needed. However, in the normal case the equivalent trapped energy mode shape, which is much more complicated, is needed. Since this is discussed in Sec.VI of Ref.4 along with Sec.5 of Ref.8 and the analyses and computer programs from Refs.4 and 5 are simply used in this work, we do not present the equivalent trapped energy mode shape here. In addition to the contoured resonators, which were treated in all the other references, as in Ref.3 in this work we also treat the flat plate trapped energy resonator with rectangular electrodes. However, the mode shape is the same as the equivalent trapped energy mode shape discussed in detail in Sec.V of Ref.8, with the exception that Eqs.(63), (64) and (66), (67) of Ref.8 are replaced by the real trapped energy mode equations

$$\hat{\xi}\tan\hat{\xi}\ell = \xi^{SR}, \quad \hat{\nu}\tan\hat{\nu}\ell = \nu^{TM}, \qquad (2.7)$$

$$\xi^{SR} = [(k_n/M_n')\,\Delta_n - \hat{\xi}^2]^{1/2},$$

$$\nu^{TM} = [(k_n/P_n')\,\Delta_n - \hat{\nu}^2]^{1/2}, \qquad (2.8)$$

in which 2ℓ is the dimension of the square electrode and where

$$k_n = (\bar{c}^{(1)}\rho)^{1/2}\frac{n\pi}{h},$$

$$\Delta_n = \frac{n\pi}{2h}\left[\frac{\bar{c}^{(1)}}{\rho}\right]^{1/2}\left[\frac{4k_1^2}{n^2\pi^2} + \hat{R}\right], \qquad (2.9)$$

and $\bar{c}^{(1)}$ is the eigenvalue for the pure thickness mode of the unelectroded plate with nonzero displacement $u_1^{(n)}$ and k_1^2 and $\hat{R}$ are defined in Eqs.$(78)_{2,3}$ of Ref.7. Equations (2.7)–(2.9) respectively, correspond to Eqs,(2.12), (2.10), (2.23) and (2.24) of Ref.9. Since the normal and in–plane biasing states are known from Sec.V of Ref.4 and Sec.VI of Ref.5, respectively, and the mode shapes are known from the discussion earlier in this section, the $\hat{c}_{L\gamma M\alpha}$ and g_γ^μ are known and the perturbation integral for H_μ in (2.2) can be evaluated.

3. Small Offset of Mode Center

A plan view and cross–section of the stiffened structure is shown in Fig.2 along with the coordinate systems. The origin of the unprimed coordinate system is located at the center of the supports and the origin of the primed coordinate system is located at the center of the mode shape, which we permit to be displaced with respect to each other by an arbitrary small distance $\underset{\sim}{d}$ in an arbitrary direction θ. Hence, $\underset{\sim}{d}$ has arbitrary components d_1 and d_3. From Fig.2 it is clear that we may write

$$X_A = d_A + X_A', \quad A = 1,3, \neq 2, \qquad (3.1)$$

which may then be substituted in the expression for the biasing state, as in Eqs.(8.2)–(8.4) of Ref.5. Since the perturbation integral in (2.2) vanishes for $d_A = 0$, substitution of the above–mentioned expression and expansion to first order in d_A yields, with the aid of (2.1),

$$\Delta\omega/\omega)^L = \Gamma^L = C_A^L d_A, \qquad (3.2)$$

and we note that the C_A^L are complicated expressions which are not terribly revealing and have been calculated.

Since we are interested in the maximum value of the magnitude of the resultant acceleration sensitivity $\underset{\sim}{\Gamma}$ for a given $|\underset{\sim}{d}|$, we write

$$\Gamma^2 = \Gamma^L\Gamma^L = C_A^L d_A C_B^L d_B = C_A^L C_B^L\, Q_A Q_B d^2 \qquad (3.3)$$

where

$$Q_1 = \cos\theta, \quad Q_3 = \sin\theta, \qquad (3.4)$$

and θ is the angle between the X_1–axis and $\underset{\sim}{d}$. Since the maximum value of the resultant sensitivity per unit error d is the quantity of interest, we define γ by

$$\gamma^2 = \Gamma^2/d^2 , \tag{3.5}$$

which may be maximized by

$$\partial\gamma^2/\partial\theta = 0 , \quad \partial^2\gamma^2/\partial\theta^2 < 0 , \tag{3.6}$$

to obtain

$$\gamma^2)_{max} = \tfrac{1}{2}[C_A^L C_A^L + \Delta] , \tag{3.7}$$

where

$$\Delta = \sqrt{(C_1^L C_1^L - C_3^L C_3^L) + 4(C_1^L C_3^L)^2} . \tag{3.8}$$

Since the direction of the mispositioning of any clip is fixed, the C_f compose vectorially, i.e.,

$$C_f^2 = C_f^L C_f^L , \tag{3.9}$$

where analogous to (3.2), we have

$$(\Delta\omega/\omega)_{f_i}^L = C_{f_i}^L f_i , \quad (\Delta\omega/\omega)_{f_i} = C_{f_i} f_i . \tag{3.10}$$

4. Calculated Results

From the discussion in Sec.2 we now know g_γ^μ and from Refs.4 and 5, we know $\hat{c}_{L\gamma M\alpha}$ for normal and in–plane acceleration, respectively. Hence, we can now evaluate H_μ in Eq.(2.2). Since the perturbation integral H_μ vanishes for a perfectly symmetric mode–shape plus support configuration, i.e., when both d_A and f_i vanish, in this work we calculate the $\gamma)_{max}$ in (3.7) and C_{f_i} defined in $(3.10)_2$. As noted earlier the f_i denote the mispositioning of a clip and are scalars, while the d_A denote the components of the planar displacement of the center of the mode shape from the center of the rectangle. The symbol $\gamma)_{max}$ is sometimes represented by C, which denotes the mispositioning coefficient for the ith clip shown in Fig.2. Since for a perfectly symmetric system the resultant acceleration sensitivity vanishes, the C and the C_{f_i} are the important quantities to study because when they are minimized, the influence of an error in fabrication is minimized. Since C and C_{f_i} will be plotted in this work and the significance of their value is not familiar, the relation between the values of C and the acceleration sensitivity for a range of mispositioning dimensions is shown by the straight lines in Fig.3.

Calculations of the different C and C_{f_i} have been performed using the known values of the second order[10] and the third order[11] elastic constants of quartz. Since the earlier work in Refs.4 and 5 has shown that low values of C are consistently achieved for ratios of the thickness of the cover plates to the active plate of 2.0, which ratio was adhered to in Ref.3, this ratio is adhered to in all calculations presented in this work. Furthermore, since it was shown in Ref.3 that an aspect ratio of the support rectangle of 1.5 always yields optimal results, the aspect ratio of 1.5 is adhered to in all calculations presented in this work. In addition, only orientations in the vicinity of those that yielded optimal results in Ref.3 are investigated here. Moreover, the energy ratio[12] of 10^{-6}, which was used in most of Ref.4, is consistently used here to determine the distance between sidewalls in the short direction. It should be noted that for all the biconvex resonators considered in this work, the radius of each contour is 40 cm.

In Fig.4 we have plotted $\gamma)_{max} \equiv C$ as a function of the center–to–center distance between the two clips on the large side for a fundamental 1.5 MHz and a third overtone 5 MHz contoured SC–cut quartz resonator for the two orientations of the support rectangle shown. The clip dimensions used in these calculations are $t_1 = t_2 = .12a$ and the clip thickness is .1 mm. The figure clearly shows that the support orientation β of –50° is quite a bit better than –45° for the fundamental, while the –45° orientation is just slightly better than the –50° orientation for the third overtone resonator. The figure also shows that for all cases considered the optimal spacing 2f between the two clips on the large side is very close to .9a. Figures 5 and 6 show the mispositioning coefficients C_{f_1} and C_{f_2}, respectively, for the same resonators as Fig.4. Since the C_{f_1} and C_{f_2} are relatively flat and about one–third of the values of C in Fig.4, they are not particularly significant. Since this general behavior of the C_f curves relative to the C curves is maintained in all cases treated, we do not consider it purposeful to show the others and, hence, do not exhibit them.

Figure 7 shows C vs f/a curves for a number of fundamental SC–cut flat plate trapped energy resonators, each for two support orientations near the optimal for that case. It can be seen from the figure that the support orientation can depend significantly on the electrode size. The figure also shows that for all cases considered the optimal spacing 2f between the two clips is very close to a.

Figure 8 shows C as a function of the size of a clip t_2 on the large side for a fixed size t_1 on the small side for a fundamental 1.5 MHz contoured SC–cut resonator. For this case the distance between the clips 2f = a and the thickness of all clips is .1 mm. The support orientation $\beta = -50°$. The figure clearly shows that very nearly optimal conditions are obtained for $t_1 = t_2$ between .1a and .12a. Figure 9 shows C vs t_2 for a fixed t_1 for a fundamental 10 MHz SC–cut flat plate resonator with an

electrode area of 4 mm^2 and a support orientation $\beta = -65°$. The figure indicates that near optimal conditions can be realized for t_1 between .1a and .15a with $t_2 = .15a$.

Figure 10 shows C as a function of the spring thickness for the particular spring model considered for a number of SC–cut contoured resonators with the different clip sizes and support orientation β shown in the figure. Although the curves are relatively flat they support the optimal t_1 and t_2 for the clips noted in the discussion of Fig.8. Figure 11 shows C vs spring thickness for a number of SC–cut flat plate trapped energy resonators with the different clip dimensions and support orientations shown in the figure. Although the curves are not quite as flat as those in Fig.10, they are still relatively flat. The figure indicates that the larger the electrode size, the higher the mispositioning coefficient. The rectangular dimensions of the support configuration for all the resonators considered are given in Table I.

In closing we present our recommendations based on all our calculations. For all resonators the a/b ratio should be 1.5. For contoured resonators we recommend that 2f = .9a, that $t_1 = t_2$ should be between .1a and .12a and that β should be between –45° and –50°. For flat plate trapped energy resonators we recommend that 2f be between .9a and a, that t_1 be between .1a and .15a and that t_2 be about .15a. For the flat plate we note that the orientation β seems to depend on the details of the resonator.

Acknowledgements

We wish to thank R.C. Smythe of Piezo Technology, Inc. for providing the design of the clip used in the calculations.

This work was supported in part by the Army Research Office under Grant No.DAAL03–91–G–0120.

References

1. Y.S. Zhou and H.F. Tiersten, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators with the Mode Shape Displaced with Respect to Rectangular Supports," J. Appl. Phys., 69, 2862 (1991).

2. H.F. Tiersten and Y.S. Zhou, "On the In–Plane Acceleration Sensitivity of Contoured Quartz Resonators Supported Along Rectangular Edges," J. Appl. Phys., 70, 4708 (1991).

3. Y.S. Zhou and H.F. Tiersten, "Calculated Orientations and Aspect Ratios of Stiffened Rectangular Support Systems of SC–Cut Quartz Resonators which Minimize the Influence of Fabrication Imperfections on Acceleration Sensitivity," Proceedings of the 1993 IEEE International Frequency Control Symposium, IEEE Cat. No.93CH 3244–1, Institute of Electrical and Electronics Engineers, New York, 442 (1993).

4. Y.S. Zhou and H.F. Tiersten, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators Stiffened by Quartz Cover Plates Supported by Clips," J. Appl. Phys., 72, 1244 (1992).

5. Y.S. Zhou and H.F. Tiersten, "In–Plane Acceleration Sensitivity of Contoured Quartz Resonators Stiffened by Quartz Cover Plates Supported by Clips," J. Appl. Phys., 74, 7067 (1993).

6. H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., 64, 832 (1978).

7. 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).

8. H.F. Tiersten and D.V. Shick, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported Along Rectangular Edges," J. Appl. Phys., 67 60 (1990).

9. H.F. Tiersten, "Analysis of Trapped Energy Resonators Operating in Overtones of Coupled Thickness Shear and Thickness Twist," J. Acoust. Soc. Am., 59, 879 (1976)

10. R. Bechmann, "Elastic and Piezoelectric Constants of Alpha–Quartz, Phys. Rev., 110, 1060 (1958).

11. R.N. Thurston, H.J. McSkimin and P. Andreatch, Jr., "Third Order Elastic Constants of Quartz," J. Appl. Phys., 37, 267 (1966).

12. The energy ratio is defined as the energy in the mode beyond the long edge of the support configuration divided by 1/2 the total energy in the mode.

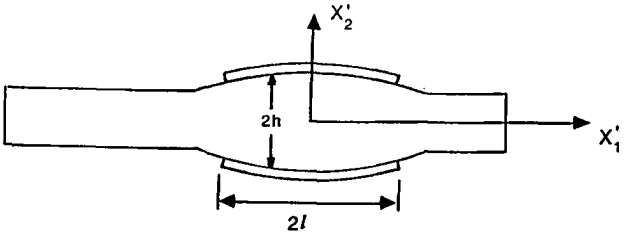


Figure 1 Cross-Section of a Biconvex Resonator

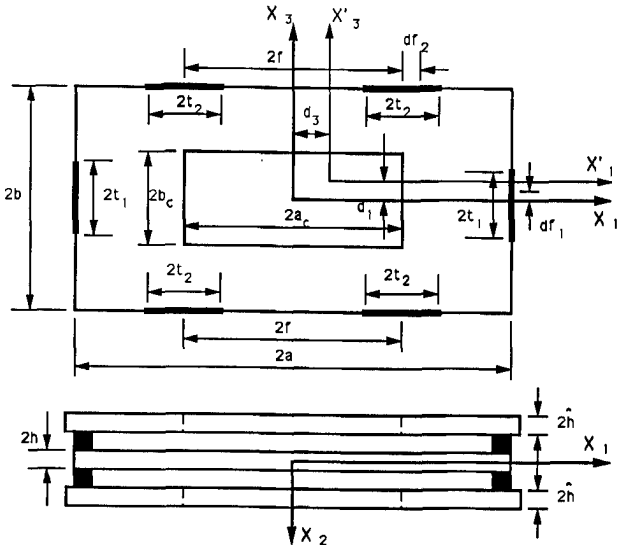


Figure 2 Plan View and Cross-Section of Box Structure Showing Active Plate, Cover Plates, Clips and Sidewalls and Displaced Location of Mode Center and Clips

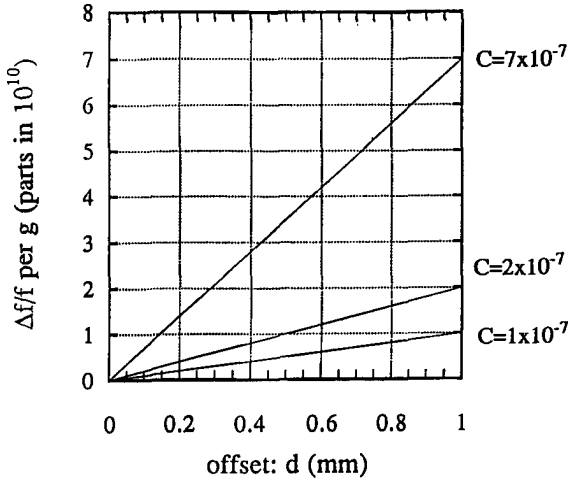


Figure 3 Acceleration Sensitivity Versus Mispositioning Dimension for Different Values of the Mispositioning Coefficient

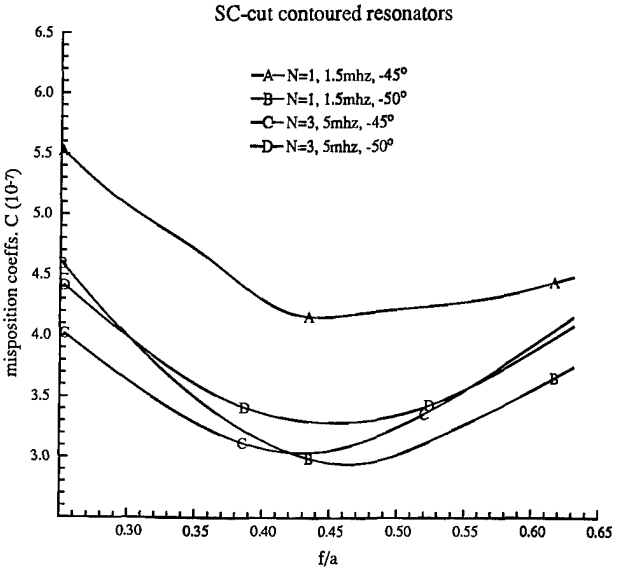


Figure 4 Mispositioning Coefficients for Mode Center Versus Distance Between Centers of Clips for Contoured SC-cut Resonators

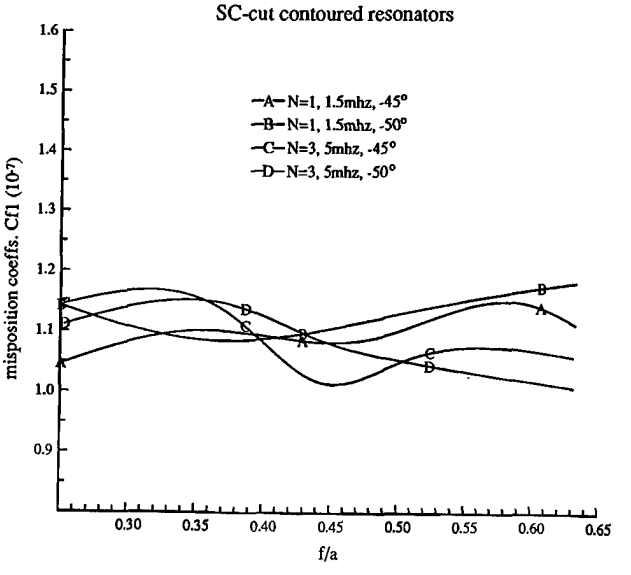


Figure 5 Mispositioning Coefficients for Clips on Large Side Versus Distance Between Centers of Clips for Same Resonators as Figure 4

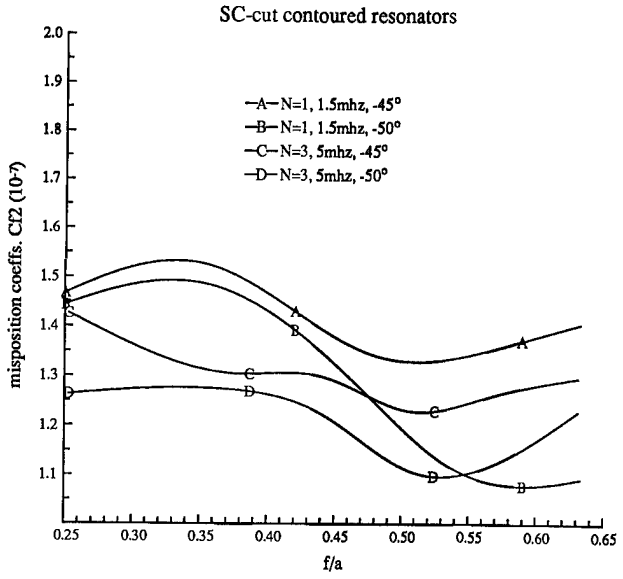


Figure 6 Mispositioning Coefficients for Clips on Large Side Versus Distance Between Centers of Clips for Same Resonators as Figure 4

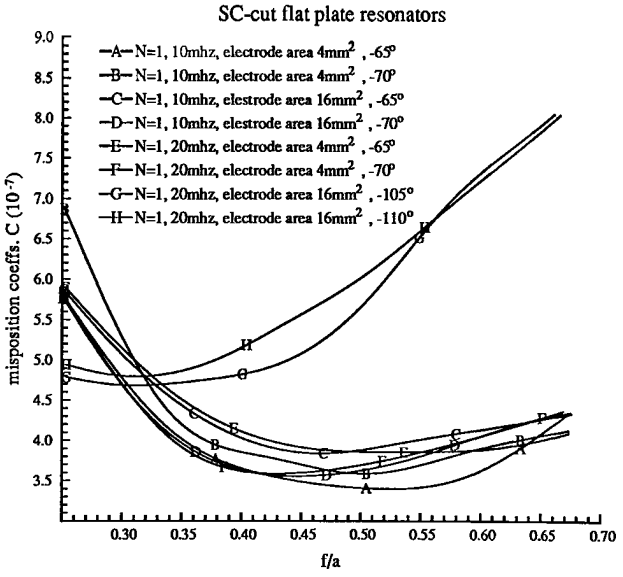


Figure 7 Mispositioning Coefficients for Mode Center Versus Distance Between Centers of Clips for Flat Plate SC-Cut Resonators

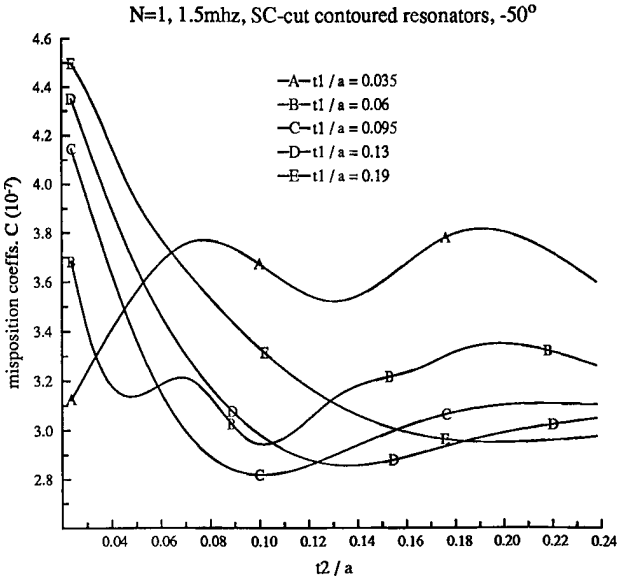


Figure 8 Mispositioning Coefficients for Mode Center Versus Size of Clip on Large Side for a Fixed Clip Size on Small Side for Contoured SC-Cut Resonators

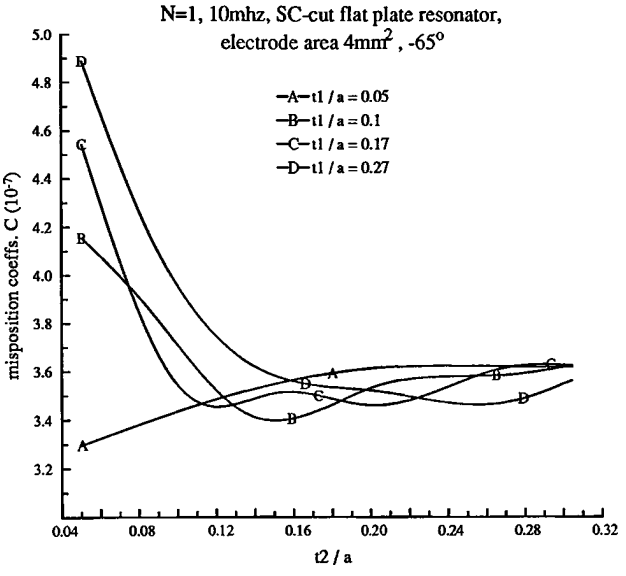


Figure 9 Mispositioning Coefficients for Mode Center Versus Size of Clip on Large Side for a Fixed Clip Size on Small Side for Flat Plate SC-Cut Resonators

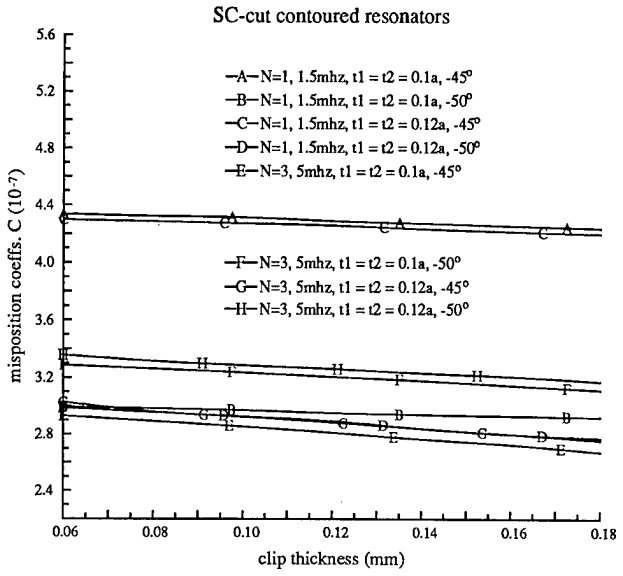


Figure 10 Mispositioning Coefficients for Mode Center Versus Clip Thickness for Fixed Clip Sizes on Both Sides for Contoured SC-Cut Resonators

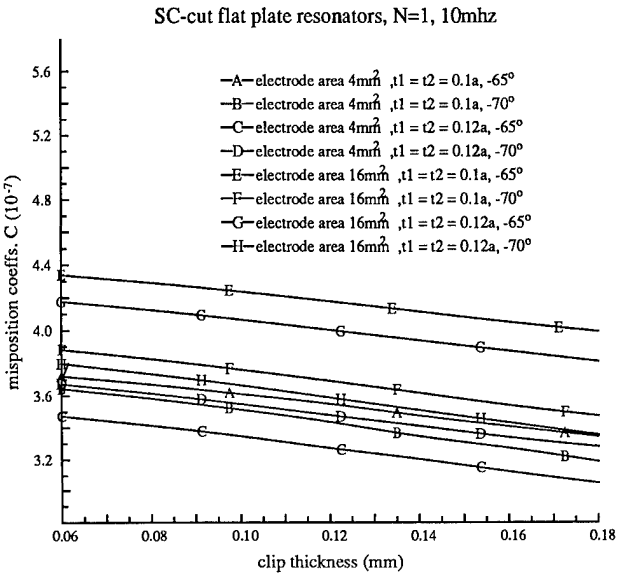


Figure 11 Mispositioning Coefficients for Mode Center Versus Clip Thickness for Fixed Clip Sizes on Both Sides for Flat Plate SC-Cut Resonators

TABLE I

RECTANGULAR DIMENSIONS OF RESONATORS

N	Frequency MHz	Type	a mm	b mm	Electrode mm x mm
1	1.5	Contoured	16.81	11.21	
3	5	Contoured	7.92	5.28	
1	10	Flat plate	5.99	3.96	2 x 2
1	10	Flat plate	6.45	4.31	4 x 4
1	20	Flat plate	3.22	2.15	2 x 2

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

AN ANALYSIS OF TRANSVERSELY VARYING THICKNESS MODES IN TRAPPED ENERGY RESONATORS WITH SHALLOW CONTOURS

H.F. Tiersten, B.J. Lwo and B. Dulmet*

Rensselaer Polytechnic Institute, Troy, NY 12180–3590

*Ecole Nationale Supérieure de Mécanique et des Microtechniques; La Bouloie, Route de Gray, 25030, Besançon Cedex, France

Abstract

The equation for transversely varying thickness modes in doubly–rotated quartz resonators is applied in the analysis of contoured resonators with rectangular electrodes. The influence of both the contouring and the continuity conditions at the edges of the electrodes are included in the analysis. The steady–state forced vibrations of contoured trapped energy resonators is treated and 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 trapped energy resonators with shallow contours.

1. Introduction

The equation for transversely varying thickness modes[1] in doubly–rotated quartz resonators has been applied in the analysis of both flat plate trapped energy resonators[2] and contoured resonators[3]. In the case of the flat plate resonator the electrodes cause the trapping and the continuity conditions at the edges of the electrodes play an important role in the analysis. In the existing treatments of the contoured resonator[4,3] the contouring causes the trapping and the continuity conditions at the edges of the electrodes are ignored in the analysis. This leaves a small region of shallow contour which is not described by either treatment.

In this work the equation for transversely varying thickness modes is applied in the analysis of contoured resonators with rectangular electrodes. The influence of both the contouring and the continuity conditions at the edges of the electrodes are included in the analysis. The corner regions are treated as in earlier work[2]. This treatment requires the use of parabolic cylinder functions[5,6] in both the electroded and unelectroded regions. The satisfaction of the continuity conditions at the edges of the electrodes results in transcendental equations, the roots of which determine the eigenfunctions and associated resonant frequencies of the harmonic and anharmonic modes. An analysis of the steady–state forced vibrations of contoured trapped energy resonators is performed and 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 trapped energy resonators with both shallow and ordinary contours.

The treatment presented here indicates why and when the analysis for the contoured resonator which ignores the conditions at the edges of the electrodes works as well as it does. This treatment should also be used for contoured resonators with ordinary contours operating primarily in the fundamental mode when the electrodes are in a certain range of relatively small size. It is shown that this treatment may not be used for too shallow a contour because of loss of accuracy and, hence, may not be used for the flat plate.

2. Basic Equations

A plan view and cross–section of the trapped energy resonator with shallow contours and rectangular (square) electrodes is shown in Fig.1 along with the associated coordinate system. It has been shown[7] 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

$$M_n' \frac{\partial^2 \tilde{u}^n}{\partial x_1'^2} + P_n' \frac{\partial^2 \tilde{u}^n}{\partial x_3'^2} - \frac{n^2\pi^2}{4h^2} \bar{c}^{(1)} \tilde{u}^n - \rho \ddot{\tilde{u}}^n = F_n e^{i\omega t}, \qquad (2.1)$$

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.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 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 eigennvalue $\bar{c}^{(1)}$ when $u^{(n)} = u^{(n)}(t)$ and 2h and ρ denote the variable thickness and mass density of the resonator, respectively.

0-7803-1945-1/94 $3.00 © 1994 IEEE

In the electroded region we have

$$F_n = \frac{\rho\omega^2(-1)^{\frac{n-1}{2}} e_{26}4V}{c^{(1)}{}_n{}^2\pi^2}, \quad \tilde{c}^{(1)} = \hat{c}^{(1)},$$
$$\tilde{u}^n = \bar{u}^n, \quad \tilde{\omega}_n = \bar{\omega}_n, \tag{2.3}$$

and in the unelectroded region we have

$$F_n = 0, \quad \tilde{c}^{(1)} = \bar{c}^{(1)}, \quad \tilde{u}^n = u^n,$$
$$\tilde{\omega}_n = \omega_n, \tag{2.4}$$

where V is the driving voltage, ω is the driving frequency, e_{26} is the piezoelectric constant referred to the thickness eigendisplacement direction $u_1^{(n)}$ and $\bar{\omega}_n$ and ω_n are the respective pure thickness eigenfrequencies for the 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

$$\Delta = \frac{8k_1^2}{n^2\pi^2} + 2\hat{R}, \quad k_1^2 = \frac{e_{26}^2}{\bar{c}^{(1)}\epsilon_{22}},$$
$$\hat{R} = \frac{2\rho' h'}{\rho h}, \tag{2.7}$$

and $2h'$ and ρ' denote the thickness and mass density of each electrode. It has been shown[4,3] that to lowest order the variable thickness of a plano–convex resonator may be written in the form

$$2h = 2h_o\left[1 - \frac{(x_1'^2 + x_3'^2}{4Rh_o}\right]. \tag{2.8}$$

Taking $\tilde{u}^n$ to have $e^{i\omega t}$ dependence, omitting the $e^{i\omega t}$ and substituting from (2.8) into (2.1) while employing (2.5), we obtain the generic form

$$M_n\tilde{u}^n{}_{,11} + P_n\tilde{u}^n{}_{,33} - \rho\tilde{\omega}_n^2\left[1 + \frac{x_1^2 + x_3^2}{2Rh_o}\right]\tilde{u}^n$$
$$+ \rho\omega^2\tilde{u}^n = F_n, \tag{2.9}$$

in which we have taken the liberty of omitting the primes and still use $\tilde{u}^n$ without the time dependence.

3. Free Vibrations

In this section we obtain the eigensolutions for the associated homogeneous problem, i.e., with V = 0, which makes $F_n = 0$ everywhere. To this end we rewrite the generic equation (2.9) in the homogeneous form

$$M_n\tilde{u}^n{}_{,11} - \rho\tilde{\omega}_n^2\frac{x_1^2}{2Rh_o}\tilde{u}^n + P_n\tilde{u}^n{}_{,33} -$$
$$\rho\tilde{\omega}_n^2\frac{x_3^2}{2Rh_o}\tilde{u}^{(n)} + \rho(\omega^2 - \tilde{\omega}_n^2)\tilde{u}^{(n)} = 0, \tag{3.1}$$

and following Vangheluwe[5] we introduce the variables $\tilde{x}$ and $\tilde{z}$, defined by

$$\tilde{x} = \tilde{\alpha}x_1, \quad \tilde{z} = \tilde{\beta}x_3, \tag{3.2}$$

where

$$\tilde{\alpha} = \sqrt{n\pi}\left[\frac{\tilde{c}^{(1)}}{2M_nRh_o^3}\right]^{1/4},$$
$$\tilde{\beta} = \sqrt{n\pi}\left[\frac{\tilde{c}^{(1)}}{2P_nRh_o^3}\right]^{1/4}. \tag{3.3}$$

We now take the solution of (3.1) in the form of a product thus

$$\tilde{u} = \tilde{F}(\tilde{x})\tilde{G}(\tilde{z}), \tag{3.4}$$

and note that

$$\tilde{u}_{,11} = \tilde{F}''\tilde{\alpha}^2\tilde{G}, \quad \tilde{u}_{,33} = \tilde{F}\tilde{G}''\tilde{\beta}^2, \tag{3.5}$$

and that we have omitted writing the n in many places. The solution (3.4) satisfies (3.1) provided

$$M_n\tilde{\alpha}^2\tilde{a} + P_n\tilde{\beta}^2\tilde{b} + \rho(\omega^2 - \tilde{\omega}_n^2) = 0, \tag{3.6}$$

where $\tilde{a}$ and $\tilde{b}$ are separation constants and

$$\tilde{F}'' - \left[\tilde{a} + \frac{\tilde{x}^2}{4}\right]\tilde{F} = 0,$$
$$\tilde{G}'' - \left[\tilde{b} + \frac{\tilde{z}^2}{4}\right]\tilde{G} = 0, \tag{3.7}$$

which are equations for parabolic cylinder functions[8]. Standard solutions y_1 (symmetric) and y_2 (antisymmetric) exist[9]. Both solutions diverge as $\tilde{x}(\tilde{z}) \longrightarrow \pm\infty$ unless the parameter $\tilde{a}$ (or $\tilde{b}$) take certain values for which the series terminate and the solutions become Hermite functions. Asymmetric standard solutions U and V exist[10], which are composed of y_1 and y_2, and approach zero as $\tilde{x}(\tilde{z}) \longrightarrow \pm\infty$.

Hence, since only the symmetric solutions can be driven by a uniform voltage, we must use y_1 in the electroded region and U (or V) in the unelectroded region for the direction becoming infinite. The generic equations (3.1), (3.6) and (3.7) hold for each region defined in the analysis for different values of $\tilde{\omega}_n$, $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{a}$ and $\tilde{b}$.

Since for analytical convenience in the determination of the eigensolutions we have replaced the circular electrodes by the circumscribed square, we need consider only one quadrant of the resonator, the plan view of which is shown in Fig.2. It should be noted that the edges of the rectangular (square) electrodes are oriented along the directions of the primed coordinate system discussed in Sec.2. Since P_n and M_n are the same in the electroded and unelectroded regions and $\Delta << 1$, the continuity conditions at the edges of the electrodes are given by

$$\bar{u} = u^S\,,\quad \bar{u}_{,1} = u^S_{,1}\,,\ \text{at } x_1 = \ell_1\,,$$
$$\bar{u} = u^T\,,\quad \bar{u}_{,3} = u^T_{,3}\ \text{at } x_3 = \ell_3\,. \tag{3.8}$$

By virtue of the fact that the same continuity conditions exist along the continued dotted lines shown in Fig.2, we also have

$$u^T = u^C\,,\ u^T_{,1} = u^C_{,1}\ \text{at } x_1 = \ell_1\,,$$
$$u^S = u^C\,,\ u^S_{,3} = u^C_{,3}\ \text{at } x_3 = \ell_3\,. \tag{3.9}$$

From the generic equation (3.2), for the electroded and unelectroded regions, respectively, we write

$$\bar{x} = \bar{\alpha}x_1\,,\ \bar{z} = \bar{\beta}x_3\,,\ x = \alpha x_1\,,\ z = \beta x_3\,, \tag{3.10}$$

which with the generic equations (3.7) permits us to write

$$\bar{F}'' - \left[\bar{a} + \frac{\bar{x}^2}{4}\right]\bar{F} = 0\,,\ \bar{G}'' - \left[\bar{b} + \frac{\bar{z}^2}{4}\right]\bar{G} = 0\,,$$
$$F'' - \left[a + \frac{x^2}{4}\right]F = 0\,,\ G'' - \left[b + \frac{z^2}{4}\right]G = 0\,, \tag{3.11}$$

where from the earlier discussion we must have

$$\bar{F} = y_1(\bar{x})\,,\ \bar{G} = y_1(\bar{z})\,,\ F = U(x)\,,\ G = U(z). \tag{3.12}$$

In order to satisfy the edge conditions in (3.8) and (3.9) pointwise, we neglect Δ compared with 1 in the coefficient of the x_3^2 term in the S region and the x_1^2 term in the T region in Eq.(3.1) for the S and T regions and take the eigensolution in the form

$$\bar{u} = A\bar{F}\bar{G}\,,\quad u^S = BF\bar{G}\,,\quad u^T = C\bar{F}G\,,\quad u^C = DFG\,. \tag{3.13}$$

Substituting from $(3.13)_{1-3}$ into Eq.(3.1) for the $-$, S and T regions, respectively, and neglecting Δ in the S and T regions as noted above, we obtain the equivalent of the generic equation (3.6) for each of the three regions, which take the form

$$M_n\bar{\alpha}^2\bar{a} + P_n\bar{\beta}^2\bar{b} + \rho(\omega^2 - \bar{\omega}_n^2) = 0\ \text{ for } -\,,$$
$$M_n\alpha^2 a + P_n\bar{\beta}^2\bar{b} + \rho(\omega^2 - \omega_n^2) = 0\ \text{ for S}\,,$$
$$M_n\bar{\alpha}^2\bar{a} + P_n\beta^2 b + \rho(\omega^2 - \omega_n) = 0\ \text{ for T}\,, \tag{3.14}$$

where

$$\bar{\alpha}^2 = n\pi\left[\frac{\hat{c}(1)}{2M_nRh^3}\right]^{1/2},\quad \bar{\beta}^2 = n\pi\left[\frac{\hat{c}(1)}{2P_nRh_o^3}\right]^{1/2},$$
$$\alpha^2 = n\pi\left[\frac{\bar{c}(1)}{2M_nRh_o^3}\right]^{1/2},\quad \beta^2 = n\pi\left[\frac{\bar{c}(1)}{2P_nRh_o^3}\right]^{1/2}. \tag{3.15}$$

Employing $(2.6)_1$ in (3.15) and the fact that $\Delta << 1$, we obtain

$$\bar{\alpha}^2 = \alpha^2(1 - \Delta/2)\,,\quad \bar{\beta}^2 = \beta^2(1 - \Delta/2)\,. \tag{3.16}$$

Subtracting $(3.14)_1$ from $(3.14)_2$ and $(3.14)_1$ from $(3.14)_3$, respectively, and substituting from (2.5), we obtain

$$M_n(\alpha^2 a - \bar{\alpha}^2\bar{a}) - \frac{n^2\pi^2}{4h_o^2}(\bar{c}^{(1)} - \hat{c}^{(1)}) = 0\,,$$
$$P_n(\beta^2 b - \bar{\beta}^2\bar{b}) - \frac{n^2\pi^2}{4h_o^2}(\bar{c}^{(1)} - \hat{c}^{(1)}) = 0\,. \tag{3.17}$$

Substituting from $(2.6)_1$ and (3.16) into (3.17) and employing the fact that $\Delta << 1$, we obtain

$$a = \bar{a} + \frac{K_n\Delta}{M_n\alpha^2}\,,\quad b = \bar{b} + \frac{K_n\Delta}{P_n\beta^2}\,, \tag{3.18}$$

where

$$K_n = n^2\pi^2\bar{c}^{(1)}/4h_o^2\,. \tag{3.19}$$

At this point we note that (3.13) does not satisfy Eq.(3.1) in the relatively unimportant C region, as in earlier work on the flat plate trapped energy resonator[2]. However, since the solution in the C region satisfies (3.9), it satisfies the equations to first order in the derivatives, but not to second order in the derivatives. Thus, in this eigensolution, the equations associated with the relatively unimportant corner type regions are satisfied to first order

in the derivatives while the equations associated with the other regions are either satisfied, or almost satisfied, to second order in the derivatives. This approximation is nevertheless very accurate because of the slow variation of derivatives in the plane of the plate and the slow variation of the thickness of the plate compared with the thickness differentiated term $-\rho\omega^2\bar{u}_n$ in (3.1).

Substituting from $(3.13)_{1-3}$ into (3.8), we obtain the two sets of equations

$$Ay_1(\bar{\alpha}\ell_1) = BU(\alpha\ell_1)\,,\quad \bar{\alpha}Ay_1'(\bar{\alpha}\ell_1) = \alpha BU'(\alpha\ell_1)\,,$$

$$Ay_1(\bar{\beta}\ell_3) = CU(\beta\ell_3)\,,\quad \bar{\beta}Ay_1'(\bar{\beta}\ell_3) = \beta CU'(\beta\ell_3)\,. \tag{3.20}$$

For a non–trivial solution to each set we obtain the two transcendental equations

$$\bar{\alpha}U(\alpha\ell_1)y_1'(\bar{\alpha}\ell_1) - \alpha y_1(\bar{\alpha}\ell_1)U'(\alpha\ell_1) = 0\,,$$

$$\bar{\beta}U(\beta\ell_3)y_1'(\bar{\beta}\ell_3) - \beta y_1(\bar{\beta}\ell_3)U'(\beta\ell_3) = 0\,, \tag{3.21}$$

the roots of which yield $\bar{a}_{nm}$ and $\bar{b}_{np}$, respectively, in which we have reintroduced the n to denote the harmonic family and the m and p denote the associated anharmonics for the x_1' and x_3' directions, respectively. When $\bar{a}_{nm}$ and $\bar{b}_{np}$ are known, the eigenfrequencies ω_{nmp} for that mode may be obtained from $(3.14)_1$. Clearly, from $(3.20)_{1,3}$ we have the amplitude ratios

$$B = Ay_1(\bar{\alpha}\ell_1)/U(\alpha\ell_1)\,,\quad C = Ay_1(\bar{\beta}\ell_3)/U(\beta\ell_3)\,, \tag{3.22}$$

while from $(3.9)_1$ and (3.22), we have

$$D = Ay_1(\bar{\beta}\ell_3)y_1(\bar{\alpha}\ell_1)/U(\beta\ell_3)U(\alpha\ell_1)\,. \tag{3.23}$$

4. Forced Vibrations

Now that we have the free vibrational modes of contoured trapped energy resonators with rectangular electrodes, we can obtain the forced vibrational solution by expanding in a series of the free vibrational solutions. To this end from (2.9) we write the equation for forced vibrations in the electroded region in the form

$$M_n\bar{u}^n{}_{,11} + P_n\bar{u}^n{}_{,33} - \rho\bar{\omega}_n^2\left[1 + \frac{x_1^2 + x_3^2}{2Rh_o}\right]\bar{u}^n + \rho\omega^2\bar{u}^n = F_n e^{i\omega t}\,, \tag{4.1}$$

and in each of the unelectroded regions S, T and C in the form

$$M_n u^n{}_{,11} + P_n u^n{}_{,33} - \rho\omega_n^2\left[1 + \frac{x_1^2 + x_3^2}{2Rh_o}\right]u^n + \rho\omega^2 u^n = 0\,, \tag{4.2}$$

where F_n, $\bar{\omega}_n$ and ω_n are given in $(2.3)_1$ and (2.5).

We now write the steady–state solution of (4.1) and (4.2) as the sum of the eigensolutions obtained in Sec.3, thus

$$\bar{u}^n = e^{i\omega t}\sum_{mp}{}' K^{nmp}\bar{u}^{nmp}\,,\quad u_S^n = e^{i\omega t}\sum_{mp}{}' K^{nmp}u_S^{nmp}$$

$$u_T^n = e^{i\omega t}\sum_{mp}{}' K^{nmp}u_T^{nmp}\,,\quad u_C^n = e^{i\omega t}\sum_{mp}{}' K^{nmp}u_C^{nmp}$$

, (4.3)

in which the K^{nmp} denote the amplitudes of the driven modes. Substituting from (4.3) into (4.1) and (4.2) and employing the homogeneous forms of (3.1) for each region (–, S, T, C), we obtain

$$\sum_{mp}{}' K^{nmp}\rho(\omega^2 - \omega_{nmp}^2)\,\bar{u}^{nmp} = F_n\,,$$

$$\sum_{mp}{}' K^{nmp}\rho(\omega^2 - \omega_{nmp}^2)\,u_S^{nmp} = 0\,,$$

$$\sum_{mp}{}' K^{nmp}\rho(\omega^2 - \omega_{nmp}^2)\,u_T^{nmp} = 0\,,$$

$$\sum_{mp}{}' K^{nmp}\rho(\omega^2 - \omega_{nmp}^2)\,u_C^{nmp} = 0\,, \tag{4.4}$$

for each of the respective regions. From (4.4) we form

$$4\sum_{mp}{}' K^{nmp}\rho(\omega^2 - \omega_{nmp}^2)\left[\int_0^{\ell_1}\int_0^{\ell_3}\bar{u}^{nmp}\bar{u}^{nst}\,dx_1dx_3 + \int_0^{\ell_1}\int_{\ell_3}^{\infty}u_T^{nmp}u_T^{nst}\,dx_1dx_3 + \int_{\ell_1}^{\infty}\int_0^{\ell_3}u_S^{nmp}u_S^{nst}\,dx_1dx_3 + \int_{\ell_1}^{\infty}\int_{\ell_3}^{\infty}u_C^{nmp}u_C^{nst}\,dx_1dx_3\right] = 4\int_0^{\ell_1}\int_0^{\ell_3}F_n\bar{u}^{nst}\,dx_1dx_3\,, \tag{4.5}$$

and since no energy is lost from the eigensolutions obtained in Sec.3, they satisfy an orthogonality condition[11], which enables us to obtain from (4.5)

$$K^{nst} = F_n \mathcal{J}_{nst}/\rho(\omega^2 - \omega_{nst}^2)\, L_{nst}\,, \tag{4.6}$$

where

$$L_{nmp} = 4\Bigg[\int_0^{\ell_1}\int_0^{\ell_3} (\bar{u}^{nmp})^2\, dx_1 dx_3 + \int_0^{\ell_1}\int_{\ell_3}^{\ell_{1\infty}} (u_T^{nmp})^2 dx_1 dx_3 + \int_{\ell_1}^{\infty}\int_0^{\ell_3} (u_S^{nmp})^2\, dx_1 dx_3 + \int_{\ell_1}^{\infty}\int_{\ell_3}^{\infty} (u_C^{nmp})^2\, dx_1 dx_3\Bigg]\,, \tag{4.7}$$

$$\mathcal{J}_{nmp} = 4\int_0^{\ell_1}\int_0^{\ell_3} \bar{u}^{nmp}\, dx_1 dx_3\,. \tag{4.8}$$

The substitution of $(2.3)_1$ into (4.6) enables us to write

$$K^{nmp} = \frac{(-1)^{\frac{n-1}{2}}\, e_{26} V\, \mathcal{J}_{nmp}}{c^{(1)} n^2\pi^2 (1 - \hat{\omega}_{nmp}^2/\omega^2)\, L_{nmp}}\,, \tag{4.9}$$

in which, as usual, we have replaced ω_{nmp} by $\hat{\omega}_{nmp}$ where

$$\hat{\omega}_{nmp} = \omega_{nmp} + i\omega_{nmp}/2Q_{nmp}\,, \tag{4.10}$$

and Q_{nmp} is the unloaded quality factor in the nmpth mode.

On account of the resonance denominator in (4.9), in the vicinity of a resonance one term in the sums in (4.3) 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 region may be written in the form

$$\bar{u}_1 = e^{i\omega t} K^{nmp} \sin\frac{n\pi x_2}{2h}\, \bar{u}^{nmp} - \frac{e_{26} V x_2}{c^{(1)} 2h}\, e^{i\omega t}\,,$$

$$\bar{\varphi} = \frac{V x_2}{2h} e^{i\omega t} + \frac{e_{26}}{\epsilon_{22}}\left[\sin\frac{n\pi x_2}{2h} - (-1)^{\frac{n-1}{2}}\frac{x_2}{h}\right] K^{nmp} \bar{u}^{nmp} e^{i\omega t}. \tag{4.11}$$

The admittance Y of the nmpth mode of the contoured trapped energy resonator is obtained by first substituting from (4.11) into[12]

$$\bar{D}_2 = e_{26}\bar{u}_{1,2} - \epsilon_{22}\bar{\varphi}_{,2}\,, \tag{4.12}$$

and then substituting from (4.12) into

$$I = -4\int_0^{\ell}\int_0^{\ell} \dot{\bar{D}}_2\, dx_1 dx_3\,, \tag{4.13}$$

for the square electrode or

$$I = -\int_{A_e} \dot{\bar{D}}_2\, dx_1 dx_3\,, \tag{4.14}$$

for the circular electrode. Since the integrals are performed numerically, they may be performed equally readily for the circular as the circumscribed square electrode. Performing the integrations in either (4.13) or (4.14), we obtain

$$Y_{nmp} = \frac{I}{V} = \frac{i\omega\epsilon_{22}}{2h_o}(1 + \hat{k}_1^2)\,\hat{A}_e + \frac{4i\omega\epsilon_{22}\hat{k}_1^2\, \mathcal{J}_{nmp}^2}{[(\hat{\omega}_{nmp}^2/\omega^2) - 1]n^2\pi^2 L_{nmp} h_o} \tag{4.15}$$

where

$$\hat{A}_e = A_e(1 + \ell^2/8Rh_o)\,,\quad \hat{k}_1^2 = e_{26}^2/\epsilon_{22}c^{(1)}. \tag{4.16}$$

The quantities C_o and $\hat{C}_{nmp}$ defined by

$$C_o = \hat{A}_e\frac{\epsilon_{22}(1 + \hat{k}_1^2)}{2h_o}\,,$$

$$\hat{C}_{nmp} = \frac{4\epsilon_{22}\hat{k}_1^2\, \mathcal{J}_{nmp}^2}{n^2\pi^2 L_{nmp} h_o}\,, \tag{4.17}$$

are called the static and motional capacitances, respectively. Clearly, if the numerical integral $\mathcal{J}_{nmp}$ is performed for circular electrodes, the motional capacitance $\hat{C}_{nmp}$ is for circular electrodes.

5. Calculated Results

The results of calculations based on the analyses presented in Secs.3 and 4 are presented in this section. Figure 3 shows the variation in mode shape in the x_1–direction for a fundamental, 2.5 MHz, AT–cut contoured trapped energy resonator for different values of the radius of curvature R of the contour. The dotted line in the figure denotes the edge of the electrode. As expected, the figure clearly shows that the mode becomes

more sharply confined to the center with decreasing R. The figure also clearly shows that for the smallest value of R, i.e., R = 10.6 cm, the point of inflection of the mode occurs well within the electrode. This means that for R = 10.6 cm, the contouring causes the trapping and the influence of the edge of the electrode is negligible. The curve for R = 50 cm shows that the point of inflection still occurs inside the electrode, but not that far from the edge. Hence, curve B is for a contoured, trapped energy resonator with the contouring considerably more significant than the edge of the electrode in causing the trapping. Both curves C and D show that the inflection point occurs at the edge of the electrode. This means that for these mode shapes both the contouring and electrode edge are significant in causing the trapping. Figure 4 shows the mode shape for the lowest symmetric anharmonic for the same resonator and the same values of R as Fig.3.

Figure 5 shows the variation in mode shape for a third overtone, 5 MHz, SC–cut, B.V.A. contoured resonator with R = 30 cm for different electrode sizes. The curves presented are for the fundamental mode of this resonator because they are much more informative than those for the third overtone. The figure shows that the mode shape is essentially the same for all electrode sizes treated. However, the curves show that for the smallest electrode size, ℓ = 4 mm, the inflection point is at the edge of the electrode, while for the larger electrode sizes, the inflection point is well within the electrode. This means that for ℓ = 4mm both the contouring and electrode edge are significant in causing the trapping, while for the larger electrode sizes only the contouring is significant in causing the trapping.

Figures 6 and 7 show the mode shapes in the x_1–direction for a fundamental, 2.5 MHz, AT–cut contoured trapped energy resonator for different electrode sizes, for radii of curvature R = 200 cm and R = 1000 cm, respectively. The curves show that both the contouring and the location of the electrode edge strongly influence the trapping, since quite different curves are obtained for different electrode sizes. This difference becomes more pronounced with increasing R. The analysis of contoured trapped energy resonators presented in this work is essential for the description of these resonators. At this point it should also be noted that even for the small values of R used in practical contoured resonators, for sufficiently small, but not too small, electrode size this analysis is essential.

Calculations have been performed for a fundamental 2.5 MHz, AT–cut contoured resonator with the dimensions shown in Table I and for which measurements have been made for some of the modes. The thickness $2h_o$ = 0.68783 mm was found by matching the measured frequency for the 300 mode with the calculated one. It can be seen that the agreement between the calculated and measured resonant frequencies is extremely good as in the earlier work[3,1], except for the 122 mode, for which strong coupling to flexure is believed to exist. More importantly, we also compare these calculated resonant frequencies with those obtained using the Hermite functions employed in Ref.1. The results are almost identical. The motional capacitances of some of the modes have also been tabulated in Table I along with the measured values and those calculated using the Hermite functions of Ref.1. As observed in earlier work[3] the agreement between calculated and measured values is reasonably good. As expected, the calculated values of C_{nmp} for the square electrode in this work and using the Hermite functions of Ref.1 are very close. The relation between the C_{nmp} calculated in this work for the square and circular electrodes is quite interesting. Although for n = 1 there is generally a significant difference, for n = 3 and n = 5, there will be very little difference or none at all unless the electrodes are very small. This is a result of the fact that for n > 1, the entire mode is essentially under the electrode. Although the C_{100} is always slightly smaller for the circular than for the square electrode, the other C_{1mp} can be significantly different for the circular and the square electrode, and in fact can be larger for the circular than for the square electrode as shown by C_{122} for R = 50 cm. This is a consequence of the fact that anharmonic mode shapes such as shown in Fig.4 can result in an increase for the circle as compared with the circumscribed square.

Calculations have been performed for a third overtone 5 MHz, SC–cut contoured BVA resonator with the dimensions shown in Table II and for which frequency measurements have been made for many of the modes. Again the thickness was obtained by matching the measured frequency for the 300 mode with the calculated one. Since the calculated frequencies[13] are almost identical with those in Table IV of Ref.1, the agreement with measurement is as good as in Ref.1. The table also shows the calculated[14] C_{nmp} from Table V of Ref.1 and for the square and circular electrodes from the analysis presented in this work. As in the other case the comparison is quite interesting. Table III shows the results of calculations that have been performed for a hypothetical AT–cut contoured resonator for a radius of contour R = 200 cm for the dimensions shown. The calculations were performed for two different electrode sizes. The comparison of the C_{nmp} for the circular and square electrodes is quite interesting.

Table IV contains calculations performed for an SC–cut of the dimensions shown for different values of R in an attempt to see if the analysis presented in this work can be used for the flat plate. The columns labeled flat plate are taken from Table III of Ref.1 and are considered to be correct. The table shows that the frequencies become closer as R increases but that the motional capacitances become worse. This means that the description presented here cannot be used for the flat plate. This is not surprising because the $\tilde{\alpha}$ and $\tilde{\beta}$ in (3.3) approach zero as $R \longrightarrow \infty$ for the flat plate and the entire series description breaks down. In closing it should also be noted that although the calculations for R = 1000 cm are accurate, for R = 5000 cm they have become quite inaccurate.

Acknowledgements

We wish to thank R.C. Smythe of Piezo Technology, Inc. for a valuable comment concerning small electrode

sizes on realistic resonators and D.S. Stevens of AT&T Bell Laboratories for mentioning the flat plate.

This work was supported in part by the Army Research Office under Grant No.DAAL03–91–G–0120.

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, "Analysis of Trapped Energy Resonators Operating in Overtones of Coupled Thickness Shear and Thickness Twist," J. Acoust. Soc. Am., 59, 879 (1976).

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. D.C. Vangheluwe, "The Frequency and Motional Capacitance of Partial Contoured Crystal Resonators," Proceedings of the 34th Annual Symposium on Frequency Control, U.S. Army Electronics Command, Fort Monmouth, New Jersey, 412 (1980).

6. B. Dulmet and H.F. Tiersten, "On the Effect of Edge Boundaries on Nearly Trapped Vibrations of Contoured Rectangular Quartz Resonators," 1986 Ultrasonics Symposium Proceedings, IEEE Cat. No.86CH2375–4, Institute of Electrical and Electronics Engineers, New York, 385 (1986).

7. Ref.1, Eq.(99).

8. Handbook of Mathematical Functions, edited by M. Abramowitz and I. Stegun, (Dover, New York, 1965), Chap.19, "Parabolic Cylinder Functions.' by J.C.P. Miller.

9. Ref.8, Eqs.19.2.1 and 19..2.3.

10. Ref.8, Eqs.19.3.1 and 19.3.2.

11. For more detail see H.F. Tiersten, B.J. Lwo and B. Dulmet. "On Transversely Varying Thickness Modes in Trapped Energy Resonators with Shallow Contours," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, New York.

12. Equation (135) of Ref.1.

13. The calculations for N = 1 in the column labeled Hermite were recently performed by D.S. Stevens.

14. The calculations of C_{nmp} appearing in Table V of Ref.1 were for an assumed electrode size 2ℓ = 9.4 mm. This was inadvertently done because neither the static or motional capacitances or electrode size were provided in the publication[15] on the BVA resonator.

15. R. Bourquin, D. Nassour and D. Hauden, "Amplitude Frequency Effect of SC–Cut Quartz Trapped Energy Resonators," Proceedings of the 36th Annual Symposium on Frequency Control, U.S. Army Electronics Command, Fort Monmouth, New Jersey, 200 (1982).

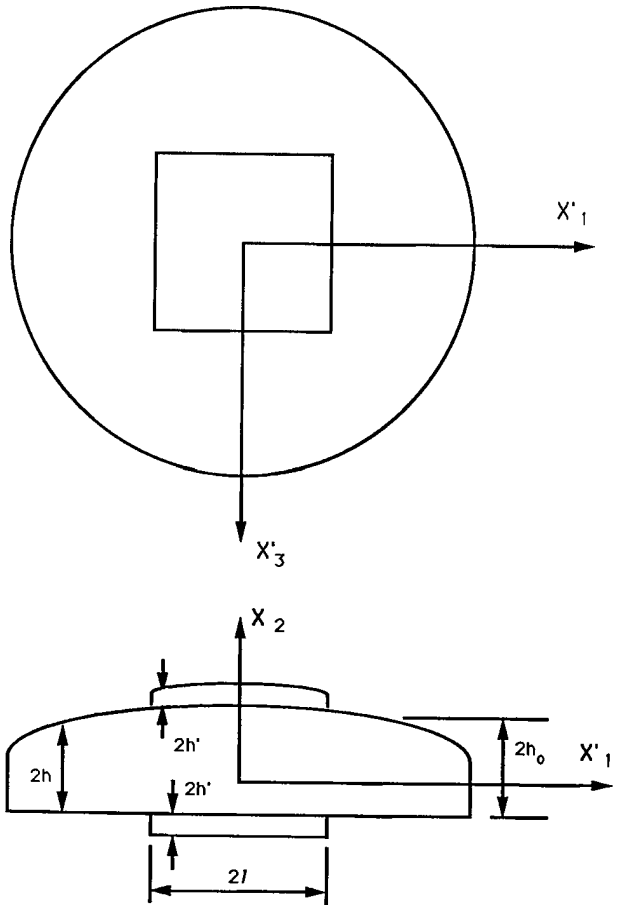


Figure 1 Plan View and Cross-Section of the Contoured Trapped Energy Resonator

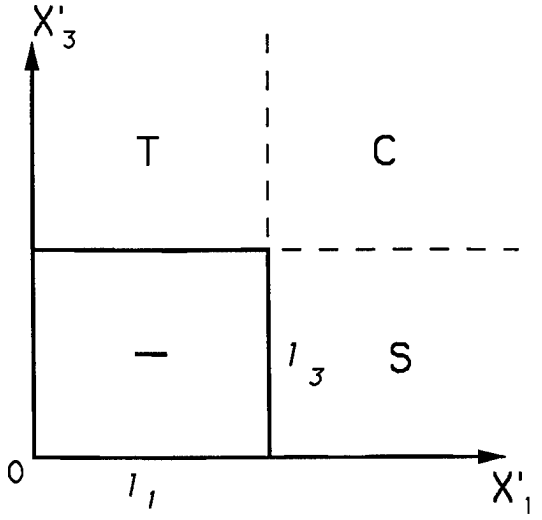


Figure 2 Plan View of the Contoured Trapped Energy Resonator Showing One Quadrant

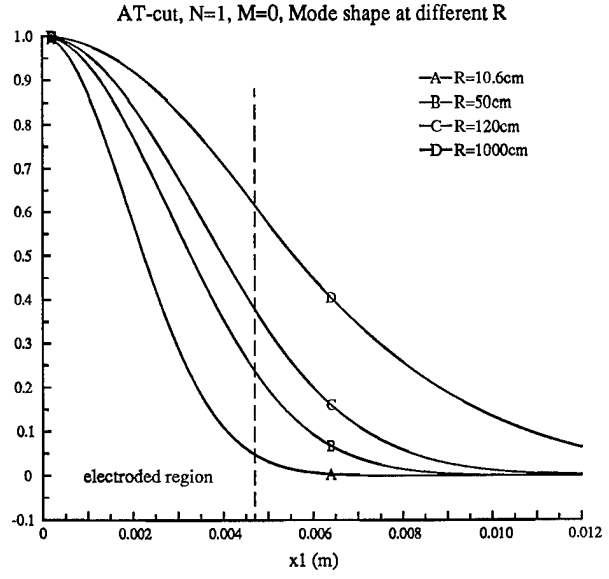


Figure 3 In-Plane Dependence of Mode Shape for a Fundamental, 2.5 MHz, AT-Cut Contoured Trapped Energy Resonator for Different Radii of Curvature of the Contour

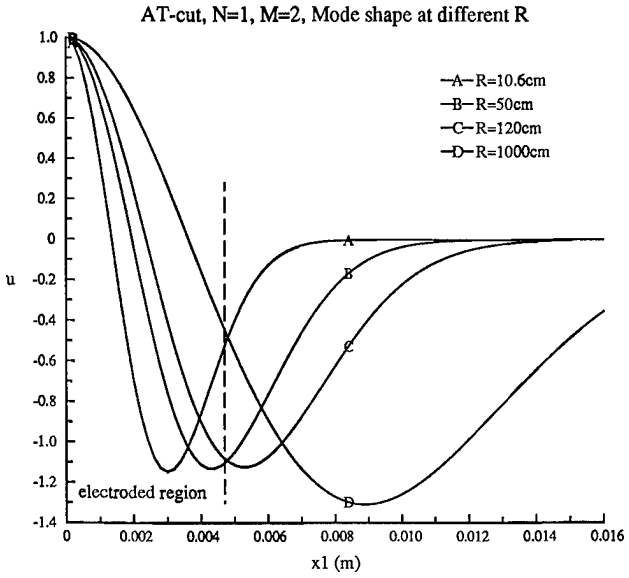


Figure 4 In-Plane Dependence of Mode Shape for the Lowest Symmetric Anharmonic for the Same Resonator and Same Radii of Curvature of the Contour as Figure 3

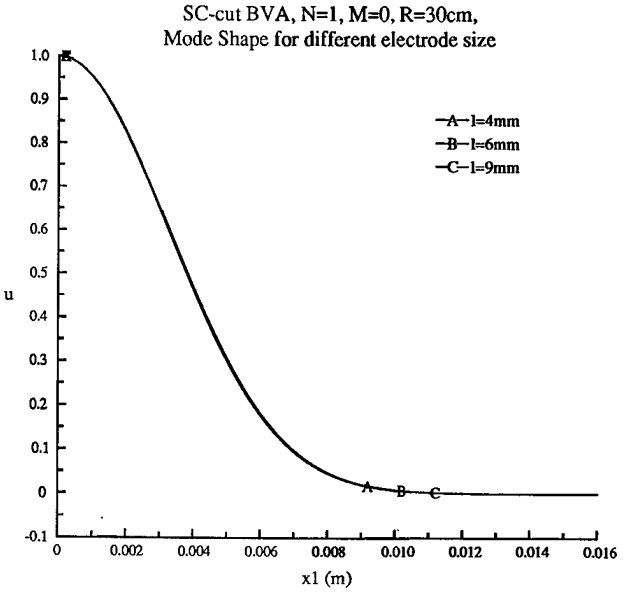


Figure 5 In-Plane Dependence of Mode Shape for the Fundamental of a 5 MHz, Third Overtone, SC-Cut Contoured BVA Resonator with R = 30 cm for Different Electrode Sizes

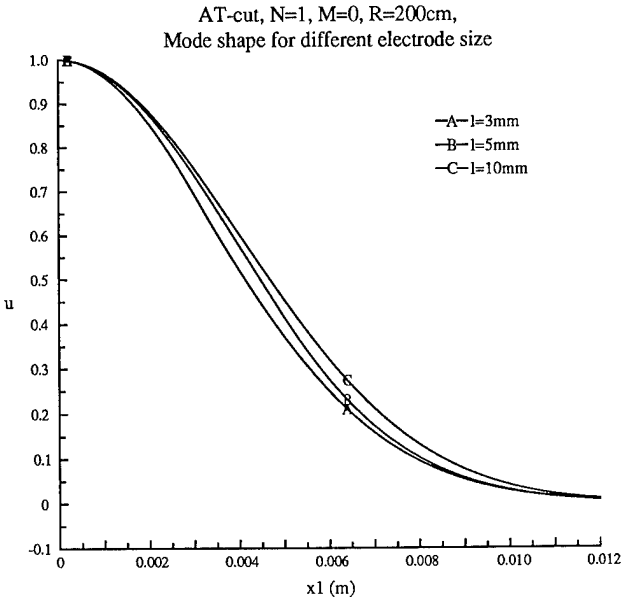


Figure 6 In-Plane Dependence of Mode Shape for a Fundamental, 2.5 MHz, AT-Cut Contoured Trapped Energy Resonator with R = 200 cm for Different Electrode Sizes

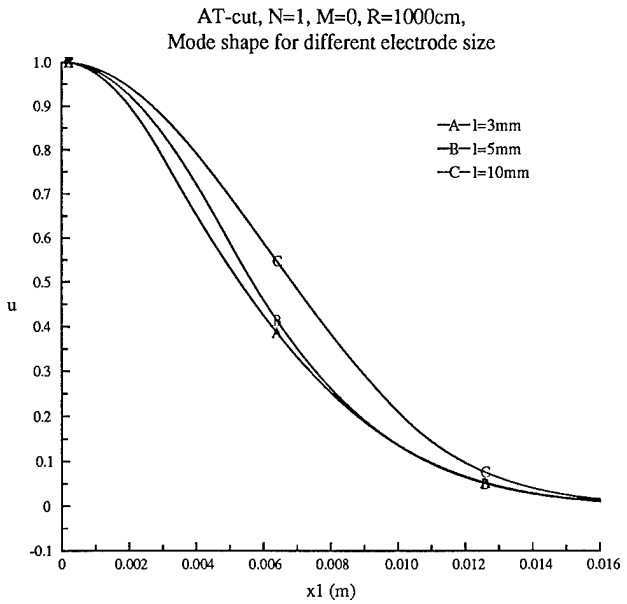


Figure 7 In-Plane Dependence of Mode Shape for a Fundamental, 2.5 MHz, AT-Cut Contoured Trapped Energy Resonator with R = 1000 cm for Different Electrode Sizes

TABLE I

CALCULATED AND MEASURED RESONANT FREQUENCIES AND MOTIONAL CAPACITANCES OF 2.5 MHz, FUNDAMENTAL, CONTOURED AT-CUT RESONATOR

AT-CUT RESONATOR, $2h_0$ = 0.68783 mm, $\hat{R}$ = 1.864 x 10^{-3}, 2ℓ = 9.4 mm

R = 10.6 cm N M P	Freq.(Khz) Hermite	Present	Measured	C(fF) Hermite	Square	Circular	Measured
1 0 0	2508.1	2508.1	2505.5	14.20	14.31	13.88	13.21
1 0 2	2683.7	2683.9	2683.4	5.81	5.90	4.60	6.25
1 2 0	2728.4	2729.3	2727.7	4.64	4.67	3.57	2.16
1 2 2	2890.7	2891.3	2843.2	1.90	1.93	0.147	....
3 0 0	7325.8	= 7325.8	= 7325.8	0.50	0.50	0.50	0.52
3 0 2	7510.7	7510.6	7514.1				
3 2 0	7519.9	7519.8	7520.1				
3 2 2	7700.1	7700.0	7693.4				
5 0 0	12152.8	12153.1	12154.1				
5 0 2	12339.4	12338.1	12343.0				
5 2 0	12366.6	12366.9	12367.7				
5 2 2	12550.1	12548.7	12532.0				

R = 50 cm N M P	Freq. (Khz)	C-square (fF)	C-circular (fF
1 0 0	2452.2	23.47	20.32
1 0 2	2539.2	3.41	0.97
1 2 0	2562.6	1.04	0.159
1 2 2	2645.9	0.152	0.600

TABLE II

CALCULATED AND MEASURED RESONANT FREQUENCIES AND CALCULATED MOTIONAL CAPACITANCES OF A 5 MHz, THIRD OVERTONE, CONTOURED SC-CUT BVA RESONATOR

SC-CUT BVA RESONATOR, $2h_0$ = 1.0905 mm, $\hat{R}$ = 0, 2ℓ = 9.4 mm, R = 30 cm

Frequency (KHz)						
N	M	P	Hermite	Present		Measured
1	0	0	1698	1698		
1	0	2	1775	1776		
1	0	4	1850	1851		
1	2	0	1798	1799		
1	2	2	1872	1873		
1	2	4	1942	1944		
1	4	0	1894	1895		
1	4	2	1963	1965		
3	0	0	5000	= 5000	=	5000
3	0	2	5085	5085		5096
3	0	4	5168	5168		5194
3	2	0	5078	5078		5089
3	2	2	5161	5161		5185
3	2	4	5243	5243		5278
3	4	0	5154	5154		5178
3	4	2	5236	5237		5268
5	0	0	8312	8312		8309
5	0	2	8407	8407		8411
5	0	4	8500	8500		8512
5	2	0	8402	8402		8404
5	2	2	8496	8496		8502
5	2	4	8588	8589		8599
5	4	0	8491	8492		8495
5	4	2	8584	8584		8589

Motional Capacitance (fF)					
				Present	
N	M	P	Hermite	Square	Circular
1	0	0	4.87	4.977	4.084
1	0	2	0.157	0.276	0.0092
1	2	0	0.006	0.0075	0.0259
3	0	0	0.265	0.266	0.261
3	0	2	0.109	0.111	0.094
3	2	0	0.115	0.116	0.099
5	0	0	0.066	0.066	0.066
5	0	2	0.032	0.032	0.031
5	2	0	0.032	0.032	0.031

TABLE III

CALCULATED RESONANT FREQUENCIES AND MOTIONAL CAPACITANCES OF A 2.5 MHz, FUNDAMENTAL CONTOURED TRAPPED ENERGY AT-CUT RESONATOR WITH R = 200 cm FOR TWO DIFFERENT ELECTRODE SIZES

AT-CUT RESONATOR, $2h_o$ = 0.68783 mm, R = 200 cm, $\hat{R}$ = 1.869 x 10^{-3}

ℓ = 0.3 cm					
N	M	P	Freq. (Khz)	C-square (fF)	(C-circular (fF)
1	0	0	2433.1	10.040	7.001
1	0	2	2483.1	4.758	4.981
1	0	4	2525.4	0.684	1.613
1	2	0	2493.6	2.937	2.669
1	2	2	2542.4	1.392	1.577
1	2	4	2583.8	0.200	0.364
1	4	0	2547.0	1.012	1.385
1	4	2	2594.8	0.479	0.611
1	4	4	2635.4	0.069	0.064
3	0	0	7252.6	1.189	0.968
3	0	2	7303.3	0.0325	3.03 x 10^{-4}
3	0	4	7349.2	0.144	0.053
3	2	0	7306.0	0.012	0.0012
3	2	2	7356.4	3.18 x 10^{-4}	0.0275
3	2	4	7401.9	0.00141	0.0227
3	4	0	7353.5	0.0501	0.0169
3	4	2	7403.5	0.00137	0.0150
3	4	4	7448.7	0.00608	0.00363

ℓ = 0.5 cm					
N	M	P	Freq. (Khz)	C-square (fF)	(C-circular (fF)
1	0	0	2428.9	30.125	23.784
1	0	2	2477.7	0.299	0.231
1	0	4	2521.1	3.354	0.844
1	2	0	2490.0	0.084	0.668
1	2	2	2537.7	8.30 x 10^{-4}	0.336
1	2	4	2580.0	0.0093	0.594
1	4	0	2543.1	0.886	0.087
1	4	2	2589.8	0.0088	0.589
1	4	4	2631.3	0.0986	0.359
3	0	0	7251.6	2.037	1.946
3	0	2	7295.6	0.680	0.475
3	0	4	7342.9	0.0853	0.0141
3	2	0	7297.9	0.621	0.431
3	2	2	7341.6	0.207	0.0013
3	2	4	7388.7	0.026	0.0339
3	4	0	7347.9	0.0415	0.0044
3	4	2	7391.3	0.0139	0.0259
3	4	4	7438.1	0.0017	2.15 x 10^{-4}

TABLE IV

COMPARISON OF CALCULATED RESONANT FREQUENCIES AND MOTIONAL CAPACITANCES OF A 10 MHz, FUNDAMENTAL CONTOURED TRAPPED ENERGY SC-CUT RESONATOR FOR INCREASING LARGE VALUES OF R WITH CALCULATIONS FOR THE FLAT PLATE

SC-CUT RESONATOR, $2h_0$ = .17979 mm, 2ℓ = 3.0 mm, M = P = 0

Frequency (Mhz)

N	$N\hat{R}$	Flat Plate	Present		
			R = 50 cm	R = 500 cm	R = 1000 cm
1	0.01	9.934	9.982	9.937	9.934
1	0.02	9.843	9.882	9.842	9.840
1	0.03	9.748	9.781	9.744	9.741
3	0.01	29.912	29.975	29.922	29.917
3	0.02	29.814	29.874	29.882	29.817
3	0.03	29.715	29.773	29.722	29.715

Motional Capacitance (fF)

N	$N\hat{R}$	Flat Plate	Present		
			R = 50 cm	R = 500 cm	R = 1000 cm
1	0.01	2.51	2.990	3.384	3.395
1	0.02	3.27	2.955	3.350	3.371
1	0.03	3.71	2.928	3.304	3.328
3	0.01	0.45	0.147	0.316	0.342
3	0.02	0.52	0.147	0.308	0.331
3	0.03	0.57	0.147	0.304	0.349

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A PERTURBATION METHOD FOR MODELLING THE THERMAL SENSITIVITY OF SURFACE TRANSVERSE WAVE (STW) PROPAGATION ON A PIEZOELECTRIC SUBSTRATE

S. Ballandras, E. Gavignet, E. Bigler

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

Abstract

A perturbation method has been developed in order to correctly predict the sensitivity of Surface Transverse Waves (STW) to quasi-static temperature effects. This approach is based on the coupling of unperturberd STW characteristics and thermoelastic properties of the substrate. The unperturbed STW parameters are calculated taking piezoelectricity into account for propagation under shallow grooves and thin metal strips. An analytical expression of the first order temperature coefficient is obtained in the case of grooves.

Finite element analysis has been used for the case of metal strips deposited one crystalline substrate submitted to slow temperature variations. The validity of the results presented is carefully analysed and comparison with available experimental data are dicussed

Introduction

The propagation of surface transverse waves (STW) on piezoelectric substrates has been used for the design of high frequency delay lines and resonators [1,2,3]. A theoretical model has been proposed for modelling temperature effects in STW devices in the nonpiezoelectric approximation [1, 4] and temperature compensated crystal orientations have been predicted. However, most devices have been built on existing cuts of quartz (AT-cut) for which the frequency-temperature dependence of STW devices has been found with a parabolic curve. Some discrepancies exist between the model proposed by Auld and Thompson [4] and experimental results, these discrepancies being attributed to quasi-static stress effects. More experimental work by Avramov [5] has confirmed the interest of the AT cut of quartz as a thermally compensated STW cut but no other theoretical approach has been presented yet. These experimental results reinforce the need of a precise model of thermal sensitivity of surface transverse wave propogation in order to optimize the existing devices and to check the existence of other temperature compensated crystal cuts for STW applications.

It is known for Rayleigh wave sensitivity to temperature effects that conventionnal models [6] based on the variation of elastic constants versus temperature introduce some fundamental errors due to the use of non rotationaly invariant temperature coefficients [7].

Then, Tiersten and Sinha have developed a perturbation method to predict more precisely the sensitivity of acoustic wave propogation to temperature effects [8]. this model is based on variational equations of small acoustic fields superposed on a bias [9] and requires the temperature derivatives of the fundamental elastic constants of the propagation substrate [10].

Following this approach, a new model is proposed to improve theoretical predictions of temperature sensitivity of STW propagation. First, derivation of STW propagation under metal strips or grooves has been implemented with piezoelectricity taken into account. This has been performed using a calculation procedure close to Bright and Hunt's one [11] but not limited to cubic crystals. This model is described in the first section of the present paper.

Then, results obtained using this model have been combined to a perturbation method in order to calculate the first order temperature coefficient of STW. In the case of STW propagation on quartz under shallow grooves, an analytical expression has been obtained allowing a systematic investigation of STW temperature coefficient versus cut angle Theta and the groove depth. In the case of STW propagation under metal strips, stresses and strains near the surface are derived by finite element analysis and then coupled with the unperturbed STW parameters using a numerical procedure close to the one implemented for predicting the mechanical and thermal stress sensitivities of Rayleigh waves [12]. Calculations have been focused in the AT cut region in order to compare theorical predictions to available experimental data [4, 5].

As a conclusion, improvements and critics of the presented model are discussed.

0-7803-1945-1/94 $3.00 © 1994 IEEE

I - STW propagation on piezoelectric substrates

In this section, derivation of STW propagation on piezoelectric crystals corrugated by shallow grooves or strips (fig. 1) will be described. In the case of quartz, that significates that only singly rotated Y cuts will be suitable for STW applications (propagation along Z' axis). It is known [13] that STW can propagate under the same conditions as Bleustein-Gulyaev waves (considering a substrate with a metallized surface). Then, piezoelectric STW are described following an approach proposed by Bright and Hunt [11]

$$u_1 = \sum_{n=-\infty}^{+\infty} \sum_{r=1}^{2} A_{nr} e^{-a_{nr} a_2} e^{-j\beta_n a_3} e^{j\omega t}$$
$$\Phi = \sum_{n=-\infty}^{+\infty} \sum_{r=1}^{2} B_{nr} e^{-a_{nr} a_2} e^{-j\beta_n a_3} e^{j\omega t} \quad (1)$$

where u_1 and Φ represent respectively the mechanical displacement and its coupled electrical potential, A_{nr} and B_{nr} are the mechanical and electrical amplitudes, a_{nr} is a coefficient relative to the wave penetration along the substrate depth. βn is the wave number for the nth mode and is defined by the relation $\beta n = \beta o + 2\pi/\Lambda$ where Λ is the periodicity of the corrugation.

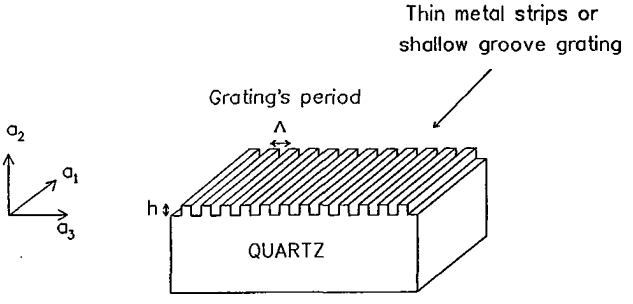


Fig. 1. *Piezoelectric substrate corrugazed by grooves or strips*

The propagation of STW will be governed by the propagation equation associated with Poisson's condition for a non-conducting media which are written for singly rotated Y quartz cuts as follows

$$\rho_o \frac{d^2 u_1}{dt^2} = C_{55} \frac{\partial^2 u_1}{\partial a_3^2} + 2C_{56} \frac{\partial^2 u_1}{\partial a_2 \partial a_3} + C_{66} \frac{\partial^2 u_1}{\partial a_2^2}$$
$$+ e_{35} \frac{\partial^2 \Phi}{\partial a_3^2} + (e_{25} + e_{36}) \frac{\partial^2 \Phi}{\partial a_3^2} + e_{26} \frac{\partial^2 \Phi}{\partial a_2^2} \quad (2)$$

$$e_{35} \frac{\partial^2 u_1}{\partial a_3^2} + (e_{25} + e_{36}) \frac{\partial^2 u_1}{\partial a_2 \partial a_3} + e_{26} \frac{\partial^2 u_1}{\partial a_2^2}$$
$$= \varepsilon_{33} \frac{\partial^2 \Phi}{\partial a_3^2} + 2\varepsilon_{23} \frac{\partial^2 \Phi}{\partial a_2 \partial a_3} + \varepsilon_{22} \frac{\partial^2 \Phi}{\partial a_2^2}$$

where ρ_o is the density of the substrate in the material coordinates (a_1, a_2, a_3) represented in fig. 1 and C_{ij} and e_{ij} are respectively the rotated elastic and piezoelectric constants expressed in contracted indexes. ε_{ij} are the rotated dielectric constants.The expression of u_1 and Φ given in equation (1) are then injected in equations (2), yielding a relation between a_{nr} and β_n under the form of a polynomial of the 4th degree in a_{nr}.

$$a_{nr}^4 (C_{66}\varepsilon_{22} + e_{26}^2)$$
$$+ a_{nr}^4 2j\beta_n (C_{66}\varepsilon_{32} + C_{56}\varepsilon_{22} - e_{26}(e_{36} + e_{25}))$$
$$- a_{nr}^2 \beta_n^2 (\varepsilon_{22} C_{55} + \varepsilon_{33} C_{66} + 4\varepsilon_{23} C_{56}$$
$$- \varepsilon_{22}(\rho_o \omega^2 / \beta_n^2) + (e_{36} + e_{25})^2 + 2e_{26}\varepsilon_{35}) \quad (3)$$
$$+ a_{nr} 2j\beta_n^3 (C_{56}\varepsilon_{33} + C_{55}\varepsilon_{23} - \varepsilon_{23}(\rho_o \omega^2 / \beta_n^2)$$
$$+ \varepsilon_{35}(e_{36} + e_{25}))$$
$$+ \beta_n^4 (C_{55}\varepsilon_{33} + \varepsilon_{35}^2 - \varepsilon_{33}(\rho_o \omega^2 / \beta_n^2)) = 0$$

For a given β_n, equation (3) will provide 4 values of a_{nr} ensuring the compatibility between the propagation equation and Poisson's condition (eq. (2)). only 2 values of a_{nr} will be considered which correspond to waves with a decreasing energy along the substrate depth. These a_{nr} are those of the two partial waves of equation (1).

Mechanical boundary conditions

It is necessary to establish the boundary conditions for both grooves and metal strips in order to calculate the other characteristics of the STW propagation. In the case of metal strips, the mechanical boundary conditions usually applied are those established by Datta and Hunsinger for a rectangular profile of the strips [14]. In this paper, a more general description of these boundary conditions based on previous works of Tuan and Parekh [15] is proposed and applied to STW calculations. This approach is only valid for shallow grooves or thin strips but it allows to consider all corrugation shape which can be represented by a function along the propagation direction a_3 defined as follows

$$a_2 = hf(a_3)$$

where h is the maximum height of the obstacle (cf. fig. 2) which is supposed to be non limited in the a_1 direction.

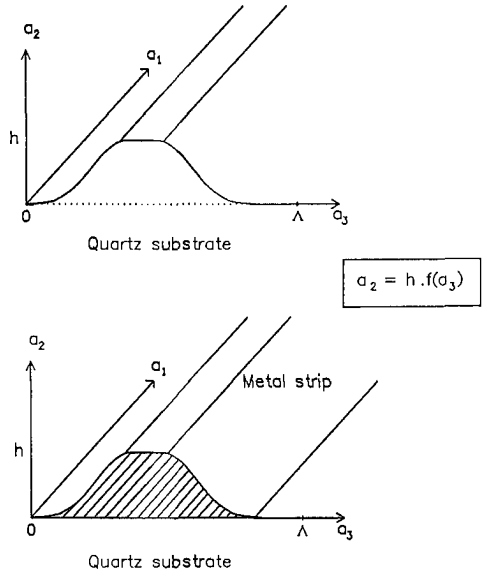


Fig. 2. *Profile of shallow grooves or metal strips*

For both cases of grooves and strips, the mechanical propagation conditions at the surface will be written

$$T_{ij}\, \nu_j = 0 \tag{4}$$

where T_{ij} is the dynamic stress field and ν_j the normal vector to the surface which is approximated using the following formula

$$\nu_1 = 0\,;\ \nu_2 = 1\,;\ \nu_3 = -\,h\frac{\partial f(a_3)}{\partial a_3} \tag{5}$$

In the case of shallow grooves, the dynamic stress field at the propagation surface is developed using Taylor-Mc Laurin series and is injected in equation (4). Taking equation (5) into account and limiting the series to the terms of first order along h, the mechanical boundary will be expressed as follows

$$T_{i2}(0,a_3) = h\left(\frac{\partial f(a_3)}{\partial a_3} T_{i3}(0,a_3) - f(a_3)\frac{\partial T_{i2}(0,a_3)}{\partial a_2}\right) \tag{6}$$

For singly rotated Y quartz cuts, equation (6) has to be verified only for i = 1.

In the case of shallow metal strips, equations (4), (5) and (6) will describe the stress field in the strip and will be noted with a prime. But it is necessary to add continuity conditions on stresses and displacements at the interface between the strip and the substrate. These conditions will be expressed as follows

$$T'_{i2} = \left|_{a_2=0} = T_{i2}\right|_{a_2=0} ;\ u'_1\Big|_{a_2=0} = u_1\Big|_{a_2=0} \tag{7}$$

Considering equation (6) and the propagation equation in the strip, mechanical boundary conditions are obtained under the following form

$$T'_{12}(0,a_3) = h\left[f(a_3)(\rho\,\omega^2 u_1(0,a_3) + \frac{\partial T'_{13}(0,a_3)}{\partial a_3} + \frac{\partial f(a_3)}{\partial a_3} T'_{13}(0,a_3)\right] \tag{8}$$

For a rectangular profile of the strip, $f(a_3)$ will be replaced by the Heaviside function and the usual Datta and Hunsinger boundary conditions will be obtained

Electrical boundary conditions

For shallow grooves, the continuity of the electrical displacement vector D_j and of the potential Φ must be respected and will be written

$$(D_j^s - D_j^v)\,\nu_j = 0 \text{ and } \Phi^s = \Phi^v \text{ in } a_2 = h f(a_3) \tag{9}$$

where the upperscript s stands for "substrate" and v for "vaccum". Φ^v is calculated using Poisson's condition in vaccum (without charges) and assuming that the electrical potentiel vanishes in $a_2 = +\infty$. Following the same approach as in the case of mechanical stresses, equation (9) is developed in Taylor-Mc Laurin series and finally, the electrical boundary condition for shallow grooves is written as follows

$$D_2^s(0,a_3) - D_2^v(0,a_3) = -h\left(f(a_3)\frac{\partial}{\partial a_2}(D_2^s(0,a_3) - D_2^v(0,a_3)) + \frac{\partial f(a_3)}{\partial a_3}(D_2^s(0,a_3) - D_2^v(0,a_3))\right) \tag{10}$$

In the case of metal strips, a simplified electrical boundary condition has been used for the present study. The electrical potential Φ has been assigned to zero in $a_2 = 0$, corresponding to an academic calculation of a completely metallized surface with local mass-loading effects due to the strips. Thus, the electrical boundary condition is reduced to the simple equation

$$\Phi^s = 0 \text{ in } a_2 = 0 \tag{11}$$

For both cases, a Fortran program has been written which calculates all the characteristics of the STW propagation using an approach identical to this implemented by Bright and Hunt for cubic crystals [11]. It is then possible to have access to all the STW parameters for all values of ω. It is particularly interesting to plot the relation between and ω and β_0 which charaterize the dispersive nature of STW propagation. The dispersion curves appear in fig. 3 for both cases of metal strips deposited at the surface of an AT quartz cut ($\Phi = 0°$, $\theta = 37°$). The spatial period Λ is equal to 4,9 μm and the calculations bave been performed for h = 100 nm, 150 nm and 200 nm which corresponds to usual practical height of the corrugation [5, 16].

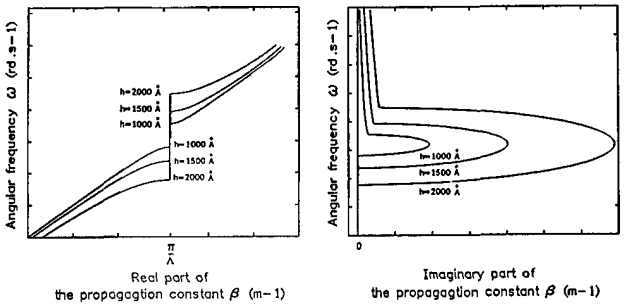


Fig. 3. *Dispersion curves for STW on (θ=37°) cut caered with metal strip of different height*

The classical stop-band phenomenon for the periodic structure composed by metal strips is illustrated in fig. 3 by the curve of the real part of the propagation constant βo. For values lower than π/Λ, βo is real, which to a propagating comportment of the wave. When βo is higher than π/Λ, its imaginary part grows which correspond to an attenuation of the wave in the corrugation network. This general behavior of STW propagation versus Λ allows to realize resonators. The characteristics of STW have been then calculated for a given structure reported in fig. 4. Now the parameters of this unperturbed STW propagation will be used to modelize first order quasi static temperature effects.

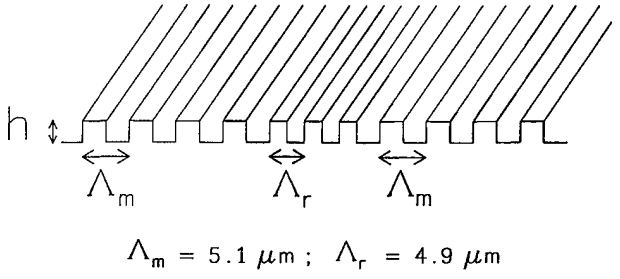


Fig. 4. *Structure of corrugations (grooves on metal strips) retained for calculation of STW parameters*

III. Perturbation method for 1st order temperature effects

In the most general case, following forms of variationnal propagation equations have to be considered to implement Tiersten's perturbation method

$$-\rho_o \omega_o^2 \iiint_v u_i^* u_i^o \, dv =$$

$$\iiint_v u_i \left[\frac{\partial}{\partial_{ak}} \left(C_{ikjm} \frac{\partial uj^o}{\partial_{am}} \right) \right] dv$$

$$-\rho_o \omega^2 \iiint_v u_i^{o*} u_i \, dv = \qquad (12)$$

$$\iiint_v u_i^{o*} \left[\frac{\partial}{\partial_{ak}} \left(\overline{\overline{A_{ikjm}}} \frac{\partial u_j}{\partial_{am}} \right) \right] dv$$

where the index o denotes an unperturbed parameter of STW and $\overline{\overline{A_{ikjm}}}$ is a tensor taking into account the non linear effects due to the perturbation considered [12]. Equations (12) may be simplified using Green-Ostrogradsky relation. The degree of derivation on u_j is then decreased but a surface integral arises. This surface term will be equal to zero because of the mechanical boundary conditions (4) which must be preserved during the perturbation phenomenon. Following then the usual perturbation procedure and applying the convenient approximations described in [8], an expression of the frequency shift due to the perturbation can be obtained

$$\frac{F - F_o}{F_o} = \frac{\iiint_v \frac{\partial u_i^{o*}}{\partial a_k} \overline{\overline{H_{ikjm}}} \frac{\partial u_j^o}{\partial a_m} dv}{2\rho_o \omega_o^2 \iiint_v u_\ell^{o*} u_\ell^o \, dv} \qquad (13)$$

where H_{ikjm} is the general thermoelastic perturbation tensor [12]. In the case of shallow grooves, no stresses arise during thermal dilatation and $\overline{\overline{H_{ikjm}}}$ will be simplified using approximations described in many references [8, 10, 12] and considered homogeneous in the whole substrate. After development of the propagation terms of equation (13) the following expression of frequency shift is obtained

$$\frac{F - F_o}{F_o} = \frac{\sum_{n,m=-\infty}^{+\infty} \sum_{r,s=1}^{2} \Gamma_{nmrs} \cdot I_{nmrs} \cdot A_{nr}^* A_{ms}}{2\rho_o \omega_o^2 \sum_{r,s=1}^{+\infty} A_{nr}^* A_{ms} I_{nmrs}} \qquad (14)$$

$$with \quad \Gamma_{nmrs} = (a_{nr}^* a_{ms} \overline{\overline{H_{1212}}} + (\beta_n a_{ms} - a_{nr}^* \beta_n) \overline{\overline{H_{1213}}} + \beta_n \beta_m \overline{\overline{H_{1313}}})$$

and

$$I_{nmrs} = \int_0^1 \int_0^{\Lambda} \int_{-\infty}^{hf(a_3)} e^{-(a_{nr}^{*} + a_{ms})\, a_2} \; e^{j\frac{2\pi}{\Lambda}(n-m)a_3} \, da_2 \, da_3 \, da_1 \qquad (15)$$

It should be noted that since (F - Fo)/Fo is calculated here only for a STW delay line or for the cavity of a resonator (cf. fig. 4), the propagation coefficient βo will remain real. Finaly, an analytic formula of the first order temperature coefficient θα is obtained for STW propagating under shallow grooves. For propagation under thin metal strips, equation (13) will be expressed differently in order to take account of the influence of the strips. As in the case of grooves, surface integrals will be supressed because the boundary conditions (4) are assumed to be respected by the perturbed propagation. Equation (13) becomes in the present case

$$\frac{F - F_o}{F_o} = \frac{\int_{-\infty}^{+\infty}\int_0^{\Lambda} \left[\int_{-\infty}^{o} \frac{\partial u_j^{o*}}{\partial a_m} H_{ikjm} \frac{\partial u_j^{o}}{\partial a_m} da_2 \right.}{2\omega_o^2 \left[\rho_o \int_{-\infty}^{+\infty}\int_0^{\Lambda}\int_{-\infty}^{o} u_\ell^{o*} u_\ell^{o} \, dv \right.} \qquad (16)$$

$$\left. + \int_0^{hf(a_3)} \frac{\partial u_i^{o*'}}{\partial a_k} H_{ikjm'} \frac{\partial u_j^{o*'}}{\partial a_m} da_2 \right] da_1 \, da_3$$

$$\left. + \rho_o' \int_{-\infty}^{+\infty}\int_0^{\Lambda}\int_0^{hf(a_3)} u_\ell^{o*'} u_\ell^{o'} \, dv \right]$$

where the primed terms are relatives to the metal strip (only one spatial period along a_3 is considered). Now, two difficulties have to be solved. First, it is necessary to know the STW parameters in the strip to calculate equation (16) and $\overline{H_{ikjm}'}$ and $\overline{H_{ikjm}}$ must take account of the bimorph effect between the metal strip and the crystal substrate which present different thermoelastic properties.

Parameters of STW in the strip can be evaluated using a calculation process inspired from [17]. Since continuity of the mechanical displacement has to be verified at the interface $a_2 = 0$, u'_i will present the same polarization than u_i along a_3 and will be expressed as follows

$$u_1' = \sum_{n=-\infty}^{+\infty} G_n(a_2)\, e^{-j\beta_n a_3} e^{j\omega t}, \; u_2' = u_3' = 0 \qquad (17)$$

where $G_n(a_2)$ is a function to be defined using propagation equation in the strip [17]. Injecting (17) in this equation yields the following differential equation

$$\frac{\partial^2 G_n(a_2)}{\partial a_2^2} + \left(\rho' \frac{\omega^2}{C_{66}'} - \beta_n^2\right) G_n(a_2) = 0 \qquad (18)$$

Then it is possible to evaluate a solution of $G_n(a_2)$ under the form

$$G_n(a_2) = a_n e^{\lambda n a_2} + b_n e^{-\lambda n a_2}, \; \lambda n = \left(\beta_n^2 - \rho' \frac{\omega^2}{C_{66}'}\right)^{1/2} \qquad (19)$$

where a_n and b_n are two terms of amplitude which can be calculated using continuity relations (7).

Now the problem of $\overline{H_{ikjm}}$ must be considered. It is unrealistic to solve analytically the thermoelastic problem of a thin metal strip deposited on a crystal substrate submitted to temperature variations. On the contrary, a finite element package will allow the evaluation of stresses, strains and displacement gradients for such a problem. The finite element package MODULEF has been used to calculate these static parameters in a 2D model. One spatial period along a_3 has been meshed (fig. 5) and stresses, strains and displacement gradients have been considered in the vicinity of the surface where the STW propagates. Equation (16) has been adapted for the treatment of discrete values of H_{ikjm} and H_{ikjm}'. To be also coherent with all the studies devoted to thermal influence on acoustic resonators and delay lines, the first derivatives of the elastic constants of the metal strip had to be calculated.

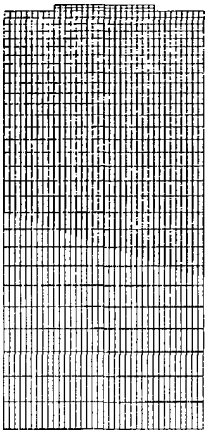


Fig.5. *Mesh of one spatial period of a STW structure*

IV. Numerical results

In both cases of grooves and strips, calculations have been done for quartz substrates, the metal of the strip was Aluminum and the profile of the strip was considered to be rectangular. For STW propagating

under shallow grooves, a systematic investigation of the first order temperature coefficient versus cut angle theta and metal thickness has been done using a non-piezoelectric analysis of STW propagation. results are reported on fig. 6, showing three regions where θ_0 is close to zero. However, it is not possible to verify the relevance of these theoretical calculations since the thermal behaviour of STW devices built with grooves has been reported until now. Nevertheless, it is interesting to compare results obtained respectivelly with non-piezoelectric and piezoelectric analyses of STW propagation. Fig. 7 shows such a comparaison between both calculations in the vicinity of AT cuts ($31 < \theta < 39$, $h = 0.1\mu m$). A good agreement between the results can be remarked. The characteristics used to define the STW device are those given in fig. 4.

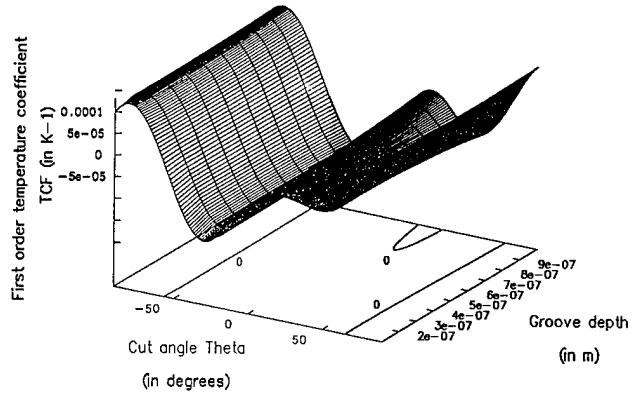


Fig. 6. *Map of the first order temperature coefficient versus cut angle θ and groove depth h*

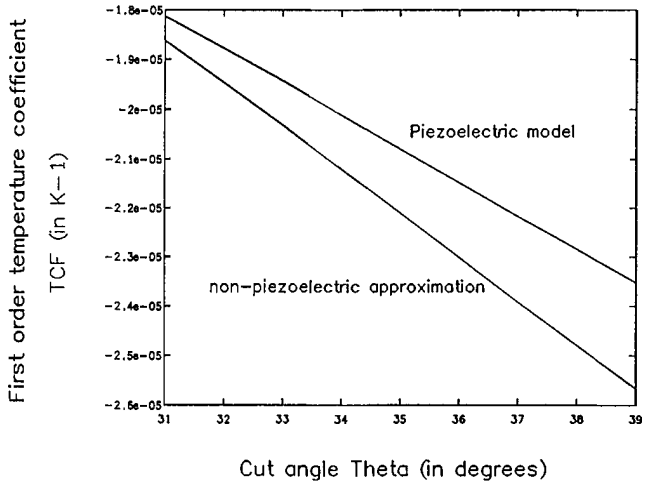


Fig. 7. *Comparison between results obtained using piezoelectric model for the prediction of first order temperature coefficient*

It is also very interesting to compare the results obtained using this model with the first order temperature coefficient of bulk acoustic waves (BAW/SSBW) on singly rotated quartz cut.

Fig. 8 shows the superimposition of STW and BAW/SSBW first order temperature coefficients. To comment this figure, one has to precize that STW excited on an Y cut presents characteristics close to the piezoelectrically coupled thickness shear bulk wave propagating normal to a Y+90° plate. Consequently the well known curve of the first order temperature coefficient of BAW resonators has been shifted by 90° along Theta. This shift allows to properly compare temperature sensitivities of waves of similar propagation charcteristics and of same thermoelastic properties.

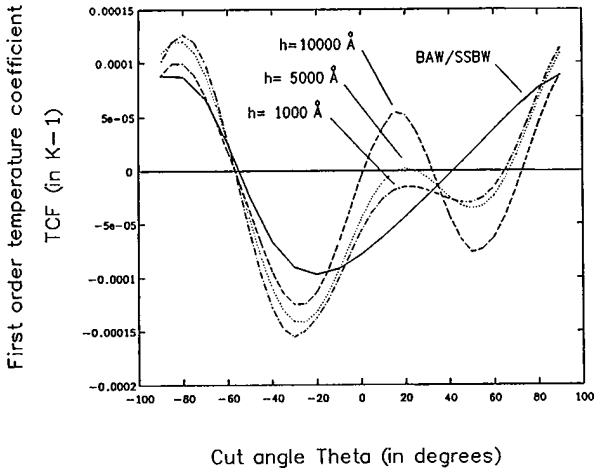


Fig. 8. *Comparison between BAW/SSBW and STW first order temperature coefficient (case of grooves of varying depth)*

In the vicinity of STW's BT cut, a good agreement between the two models can be emphasized, showing that a STW propagating at the surface of a BT cut should present a frequency-temperature behaviour close to the slow thickness shear BAW excited on an AT cut. A cubic dependance of frequency versus temperature is then expected for STW or BT cuts. Nevertheless, in the vicinity of STW's AT cut descrepancies between the two curves clearly appear.

It is then more interesting to compare with metal strip STW devices, for which experimental data on the thermal sensitivity exist [2,5].

The problem of STW propagation under Aluminum strips submitted to quasi-static temperature effects has been then numerically implemented for quartz cuts in the region of AT cuts.

These devices were built on various substrates with cut angles in the vicinity of the AT cut. On the same graph are ploted the theoretical value of the temperature coefficient of thickness shear bulk wave resonators and the predictions for STW propagating under various metal thickness, according to the present model.

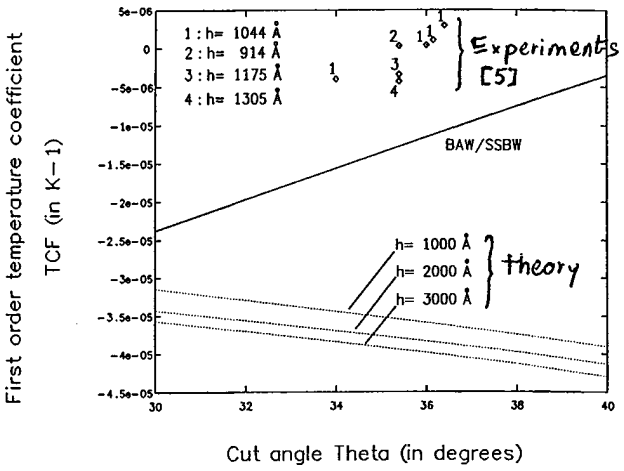


Fig. 9. *Comparison between theoretical and experimental data for the first order temperature coefficient of STW under Aluminum strips*

A rather high value of the temperature coefficient was found theoretically, which is a clear discrepancy with experimental data.

This discrepancy between the model and Avramov's experiments could be explained by the fact that the theoretical calculations presented here are only valid for delay lines and not for resonators. However an analysis of Thompson measurements [2] performed on delay lines shows that a positive value of first order temperature should be found for cuts in the vicinity of the AT cut.

Improvements of the model could be made by using a full 3-D thermoelastic simulation of the Aluminum/Quartz of the simplified 2-D approach used in the present study.

V. Conclusion

A model based on a variationnal perturbation method has been proposed to predict the first order temperature coefficient of STW devices built on quartz substrates. To increase the precision of the calculation, piezoelectricity has been taken into account to solve the problem of STW unperturbed propagation. From the propagation parameters of an unperturbed STW, the model is coupled to a perturbation method and predicts the existence of temperature-compensated orientations on quartz for grooved or metal strip gratings. The influence of the groove depth is caculated and a whole family of temperature compensated STW devices using grooved gratings can be found on (AT-cut) plates.

In the present state of the model, some discrepancies exist between experimental measurements on temperature sensitivities for metal strips devices and values predicted by the model. Further improvements are required to take into account 2nd order temperature effects, and a more intricate 3-D simulation of thermoelastic effects will be useful. However the model can be extended to many kinds of crystalline symmetries and to any combinations of grooved and metal strip grating that could yield attractive temperature compensation effects for high frequency STW applications on other crystals.

References

[1] D.F. Thompson, B.A. Auld, "Surface Transverse Wave propagation under metal strip gratings", Proc. 1986 Ultrasonics Symp., IEEE cat. n° 86CN2375-4, pp. 261-266.

[2] D. F. Thompson, "Temperature compensation of microwave acoustic resonators", Ph. D. Dissertation, Ginzton Laboratory, Stanford University, (june 1986).

[3] 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.

[4] B.A. Auld and D.F. Thompson, "Temperature compensation of Surface Transverse Waves for stable oscillator applications", Proc. 1987 Ultrasonics Symp., IEEE cat. n° 87CH2492-7, pp. 305-312.

[5] I.D. Avramov, "Microwave oscillators stabilized with surface transverse wave resonant devices", Proc. 46th Ann. Freq. Cont. Symp., Hershey-PA, june 1992.

[6] J.J. Campbell and W.R. Jones, "A methode for estimating optimal crystal cuts and propagation direction for excitation of piezoelectric surface waves", IEEE Trans. Sonics Ultrason., vol.SU-15, pp.209-217 (1968).

[7] R. Bechmann, A.D. Ballato, T.J. Lukaszek, "Higher-Order Temperature Coefficients of the Elastic Stiffnesses and Compliances of Alpha-Quartz", Proceedings of the IRE, vol. 50, 1962, pp. 1812-1822.

[8] B.K. Sinha and H.F. Tiersten, "On the Temperature dependance of the velocity of surface waves in quartz", J. Appl. Phys., vol. 51, n° 1, pp. 4659-4665 (janv. 1980).

[9] J.C. Baumhauer and H.F. Tiersten , "Nonlinear electroelastic equations for small fields superposed on a bias", J. Acoust. Soc. Am., vol. 54, n° 4, pp. 1017-1074 (oct. 1973).

[10] B.K. Sinha and H.F. Tiersten, "First temperature derivatives of the fundamental elastic constants of quartz", J. Appl. Phys., Vol. 50, n° 4, pp. 2732-2739 (1979).

[11] 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).

[12] S. Ballandras, E. Bigler, "Surface acoustic wave devices with low sensitivity to mechanical and thermoelastic stresses", J. Appl. Phys. Lett., vol. 72, n° 8, pp. 3272-3281 (october 1992).

[13] D.L. Lee, "S-Band SSBW Delay lines for Oscillator Applications", IEEE Proc. 1980, Ultrasonics Symposium, pp. 245-250.

[14] S. Datta, B.J. Humsinger, "First order reflection coefficient of surface acoustic waves from thin strip overlays", J. Appl., vol. 50, n°9, 1979, pp. 5661-5665.

[15] J.P. Parek, M. S. Tuan, "Reflexion and Bulk wave conversion of Rayleigh wave at single groove", J. Appl. Phys., Vol. 48, n° 3, March 1977, pp. 994-1003.

[16] 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.

[17] A.R. Baghai-Wadji, A.A. Maradudin, 'Rigourous analysis of surface transverse waves in periodic aways with arbitrary electrode profiles", IEEE Proc. of the 1990 Ultrasonics Symposium, pp. 425-428.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Two-Dimensional Equations For Guided Electromagnetic Waves In Anisotropic Dielectric Plates Surrounded By Free Space

P.C.Y. Lee and J.D. Yu

Department of Civil Engineering & Operations Research
Princeton University, Princeton, N.J. 08544

Abstract

Two-dimensional governing equations for guided electromagnetic waves in an anisotropic dielectric plate surrounded by free space are deduced from the three-dimensional Maxwell's equations by expanding the electromagnetic vector potential in a series of trigonometric functions of thickness coordinate in the plate and in exponentially decaying functions of thickness coordinate in the upper and lower halves of free space. By further satisfying the continuity conditions of the electromagnetic field at the interfaces between the plate and free spaces, a single system of two-dimensional governing equations is obtained.

Dispersion relations are obtained from the two-dimensional equations and computed for the doubly rotated cut of lithium niobate $(yxlt)45°/5°$. These dispersion curves are then compared with the corresponding ones obtained from the exact solutions of the three-dimensional Maxwell's equations. It is shown that the agreement between the approximate and exact dispersion curves is very close.

1 Introduction

Solutions of straight-crested waves propagating in an isotropic dielectric plate of infinite extent and surrounded by free space is one of the very few simple closed form solutions of the three-dimensional Maxwell's equations and it can be found in Ref. 1 and 2. Solutions of straight-crested waves propagating in an anisotropic dielectric plate of infinite extent and surrounded by free space was obtained recently by Lee and Yu.[3] Closed form solutions for isotropic dielectric resonators of finite extent are known only for a sphere and a circular ring of small circular cross section, which were studied by Richtmyer.[4] Closed form solutions of approximate two-dimensional equations for a circular dielectric disk were obtained by Lee and Yang.[5] However, exact closed form solutions of the free vibrations in anisotropic dielectric plate of finite extent, rectangular or circular, have not yet been found in published articles.

In the present paper, we follow the procedures of Lee and Yang[6] to derive the two-dimensional equations for guided electromagnetic (EM) waves in anisotropic dielectric plates surrounded by free space. The three-dimensional Maxwell's equations and continuity conditions in terms of the electromagnetic vector potential A, for source-free regions, are introduced in Section 2. In Section 3, the two-dimensional nth-order wave equations are derived in terms of the nth-order vector potential. A truncation procedure is then introduced in section 4 where the accuracy of the two-dimensional equations are examined by comparing these dispersion curves of the straight-crested waves propagating in the x_1 direction from the two-dimensional equations with those from the exact solutions of the three dimensional Maxwell's equations.[3]

2 3-D Equations and Continuity Conditions

We consider a linear and anisotropic dielectric plate of infinite extent referred to a rectangular coordinate system x_i 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 (see Fig. 1). The plate is surrounded by free space.

The governing equations of the EM field in the plate comprise Maxwell's equations

$$\epsilon_{ijk}E_{k,j} = -B_{i,t}\ , \quad B_{i,i} = 0\ , \tag{1}$$

$$\epsilon_{ijk}H_{k,j} = D_{i,t}\ , \quad D_{i,i} = 0\ , \tag{2}$$

0-7803-1945-1/94 $3.00 © 1994 IEEE

and the constitutive equations

$$D_i = \epsilon_0 \epsilon_{ij} E_j \ , \quad B_i = \mu_0 H_i \ . \tag{3}$$

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_0 is the dielectric permittivity of free space, ϵ_{ij} the relative dielectric permittivity tensor, μ_0 the magnetic permeability of free space, and ϵ_{ijk} the unit alternating tensor. We have 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, a vector potential function A_i may be introduced from which E_i and B_i can be obtained by[7]

$$E_i = -A_{i,t} \ , \quad B_i = \epsilon_{ijk} A_{k,j} \ . \tag{4}$$

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 ϵ_{ij} is replaced by δ_{ij}.

The continuity conditions to be satisfied at $x_2 = \pm b$, the interfaces between the plate and 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 $a = 1, 3$.

By substituting (4) into (5), it may be seen that conditions (5) may be replaced by the continuity conditions on the tangential components of A_i, i.e.

$$A_a(x_2 = \pm b) = \bar{A}_a(x_2 = \pm b) \ , \quad a = 1, 3 \ . \tag{7}$$

3 Two-Dimensional Equations

In the region of the plate, i.e. for $|x_2| \leq b$, components of the vector potential A_i are expanded in an infinite series with their thickness-dependence expressed by the trigonometrical functions[6]

$$\begin{aligned} A_a &= \sum_{n=0}^{\infty} A_a^{(n)}(x_1, x_3, t) \cos \frac{n\pi}{2}(1-\psi) \ , \quad a = 1, 3 \\ A_2 &= \sum_{n=0}^{\infty} A_2^{(n)}(x_1, x_3, t) \sin \frac{n+1}{2}\pi(1-\psi) \ , \end{aligned} \tag{8}$$

where $\psi = \frac{x_2}{b}$, $A_i^{(n)}$ are called components of the nth-order vector potential. We note that they are functions of x_1, x_3, and t only.

By substituting (8) into (4), we have

$$\begin{aligned} E_a &= \sum_{n=0}^{\infty} E_a^{(n)} \cos \frac{n\pi}{2}(1-\psi) \ , \quad a = 1, 3 \\ E_2 &= \sum_{n=0}^{\infty} E_2^{(n)} \sin \frac{n+1}{2}\pi(1-\psi) \ , \\ B_a &= \sum_{n=0}^{\infty} B_a^{(n)} \sin \frac{n+1}{2}\pi(1-\psi) \ , \\ B_2 &= \sum_{n=0}^{\infty} B_2^{(n)} \cos \frac{n\pi}{2}(1-\psi) \ , \end{aligned} \tag{9}$$

where

$$\begin{aligned} E_i^{(n)} &= -A_{i,t}^{(n)} \ , \\ B_1^{(n)} &= \frac{n+1}{2b}\pi A_3^{(n+1)} - A_{2,3}^{(n)} \ , \\ B_2^{(n)} &= k(A_{1,3}^{(n)} - A_{3,1}^{(n)}) \ , \\ B_3^{(n)} &= -\frac{n+1}{2b}\pi A_1^{(n+1)} + A_{2,1}^{(n)} \ . \end{aligned} \tag{10}$$

Eqs. (10) are the two-dimensional nth-order field-potential relations. We note that a correction factor k has been introduced in the expression for $B_2^{(n)}$ in $(10)_3$, to permit some minor adjustment of frequency branches as it will be explained in Section 4.

By multiplying $(2)_1$ (for $i = 1, 3$) and $(2)_2$ by $\cos \frac{n\pi}{2}(1-\psi)$ and $(2)_1$ (for $i = 2$) by $\sin \frac{n+1}{2}\pi(1-\psi)$, integrating the resulting expressions with respect to ψ from -1 to $+1$, and using the identities

$$\begin{aligned} \int_{-1}^{1} \sin \frac{m\pi}{2}(1-\psi) \sin \frac{n\pi}{2}(1-\psi) d\psi &= \delta_{mn} - \delta_{m0}\delta_{n0} \ , \\ \int_{-1}^{1} \cos \frac{m\pi}{2}(1-\psi) \cos \frac{n\pi}{2}(1-\psi) d\psi &= \delta_{mn} + \delta_{m0}\delta_{n0} \ , \end{aligned} \tag{11}$$

we have the two-dimensional field equations

$$
\begin{aligned}
&-H_{2,3}^{(n)} - \frac{n\pi}{2b} H_3^{(n-1)} + \frac{1}{b}\mathcal{H}_3^{(n)} = D_{1,t}^{(n)} \ , \\
&H_{1,3}^{(n)} - H_{3,1}^{(n)} = D_{2,t}^{(n)} \ , \\
&H_{2,1}^{(n)} + \frac{n\pi}{2b} H_1^{(n-1)} - \frac{1}{b}\mathcal{H}_1^{(n)} = D_{3,t}^{(n)} \ , \\
&D_{a,a}^{(n)} - \frac{n\pi}{2b} D_2^{(n-1)} + \frac{1}{b}\mathcal{D}_2^{(n)} = 0 \ ,
\end{aligned} \tag{12}
$$

where δ_{mn} is the Kronecker delta, the components of the nth-order magnetic field intensities and electric flux densities are defined by

$$
\begin{aligned}
H_a^{(n)} &= \int_{-1}^{1} H_a \sin\frac{n+1}{2}\pi(1-\psi)d\psi \ , \\
H_2^{(n)} &= \int_{-1}^{1} H_2 \cos\frac{n\pi}{2}(1-\psi)d\psi \ , \\
D_a^{(n)} &= \int_{-1}^{1} D_a \cos\frac{n\pi}{2}(1-\psi)d\psi \ , \\
D_2^{(n)} &= \int_{-1}^{1} D_2 \sin\frac{n+1}{2}\pi(1-\psi)d\psi \ ,
\end{aligned} \tag{13}
$$

and the components of the nth-order magnetic face field intensity and electric face charge density are defined by

$$
\begin{aligned}
\mathcal{H}_a^{(n)} &= H_a(b) - (-1)^n H_a(-b) \ , \\
\mathcal{D}_2^{(n)} &= D_2(b) - (-1)^n D_2(-b) \ .
\end{aligned} \tag{14}
$$

By inserting (9) in (3) and, in turn, into (13), we obtain the two-dimensional constitutive relations of nth-order in terms of the vector potential $A_i^{(n)}$:

$$
\begin{aligned}
H_1^{(n)} &= \frac{1}{\mu_0}\left(\frac{n+1}{2b}\pi A_3^{(n+1)} - A_{2,3}^{(n)}\right) \ , \\
H_2^{(n)} &= \frac{1}{\mu_0} k(1+\delta_{n0})\left(A_{1,3}^{(n)} - A_{3,1}^{(n)}\right) \ , \\
H_3^{(n)} &= \frac{1}{\mu_0}\left(-\frac{n+1}{2b}\pi A_1^{(n+1)} + A_{2,1}^{(n)}\right) \ , \\
D_a^{(n)} &= -\epsilon_0\left[(1+\delta_{n0})\epsilon_{ab}A_{b,t}^{(n)} + \epsilon_{a2}\sum_{m=0}^{\infty}\gamma_{(m+1)n}A_{2,t}^{(m)}\right] \ , \\
D_2^{(n)} &= -\epsilon_0\left[\epsilon_{22}A_{2,t}^{(n)} + \sum_{m=0}^{\infty}\gamma_{(n+1)m}\epsilon_{2b}A_{b,t}^{(m)}\right] \ ,
\end{aligned} \tag{15}
$$

where $a, b = 1, 3$ and

$$
\begin{aligned}
\gamma_{mn} &= \int_{-1}^{1} \sin\frac{m\pi}{2}(1-\psi)\cos\frac{n\pi}{2}(1-\psi)d\psi \\
&= \begin{cases} 0 & m+n \text{ even} \\ \dfrac{4m}{(m^2-n^2)\pi} & m+n \text{ odd} \ . \end{cases}
\end{aligned} \tag{16}
$$

We see that for the plate ($|x_2| \le b$) the three-dimensional governing equations (2), (3), and (4) are now replaced by an infinite set of two-dimensional equations (12), (14), and (15), respectively.

Further substitution of (15) in (12) leads to the two-dimensional nth-order wave equations for $A_i^{(n)}$:

$$
\begin{aligned}
&(1+\delta_{n0})k(A_{1,aa}^{(n)} - A_{a,a1}^{(n)}) + \frac{n\pi}{2b}A_{2,1}^{(n-1)} \\
&- \left(\frac{n\pi}{2b}\right)^2 A_1^{(n)} - \frac{\mu_0}{b}\mathcal{H}_3^{(n)} \\
&= \mu_0\epsilon_0\left[(1+\delta_{n0})\epsilon_{1a}A_{a,tt}^{(n)} + \epsilon_{12}\sum_{m=0}^{\infty}\gamma_{(m+1)n}A_{2,tt}^{(m)}\right] , \\
&A_{2,aa}^{(n)} - \frac{n+1}{2b}\pi A_{a,a}^{(n+1)} \\
&= \mu_0\epsilon_0\left[\epsilon_{22}A_{2,tt}^{(n)} + \sum_{m=0}^{\infty}\gamma_{(n+1)m}\epsilon_{2a}A_{a,tt}^{(m)}\right] , \\
&(1+\delta_{n0})k(A_{3,aa}^{(n)} - A_{a,a3}^{(n)}) + \frac{n\pi}{2b}A_{2,3}^{(n-1)} \\
&- \left(\frac{n\pi}{2b}\right)^2 A_3^{(n)} + \frac{\mu_0}{b}\mathcal{H}_1^{(n)} \\
&= \mu_0\epsilon_0\left[(1+\delta_{n0})\epsilon_{3a}A_{a,tt}^{(n)} + \epsilon_{32}\sum_{m=0}^{\infty}\gamma_{(m+1)n}A_{2,tt}^{(m)}\right] ,
\end{aligned} \tag{17}
$$

$$
(1+\delta_{n0})\epsilon_{ab}A_{b,at}^{(n)} + \sum_{m=0}^{\infty}\gamma_{(m+1)n}\epsilon_{a2}A_{2,at}^{(m)}
- \frac{n\pi}{2b}\left[\epsilon_{22}A_{2,t}^{(n-1)} + \sum_{m=0}^{\infty}\gamma_{nm}\epsilon_{2a}A_{a,t}^{(m)}\right] - \frac{1}{\epsilon_0 b}\mathcal{D}_2^{(n)} = 0,
$$

for $n = 0, 1, 2, \ldots$. We note in (17) that $\mathcal{H}_a^{(n)}$ and $\mathcal{D}_2^{(n)}$, which are defined in (14), may be regarded as the "forcing functions" specified at the faces of the plate.

The two-dimensional governing equations for the evanescent field in the upper and lower halves of free space ($|x_2| \ge b$) have been derived in Ref. 6 which are summarized as follows.

For $x_2 \ge b$:

$$
\bar{A}_i = A_i^+(x_1, x_3, t)e^{-\bar{\eta}b(\psi-1)} \ , \tag{18}
$$

and

$$
\begin{aligned}
\bar{E}_i &= E_i^+ e^{-\bar{\eta}b(\psi-1)} \ , \\
\bar{B}_i &= B_i^+ e^{-\bar{\eta}b(\psi-1)} \ ,
\end{aligned} \tag{19}
$$

where

$$
\begin{aligned}
&E_i^+ = -A_{i,t}^+ \ , \quad B_1^+ = -\bar{\eta}A_3^+ - A_{2,3}^+ \ , \\
&B_2^+ = k(A_{1,3}^+ - A_{3,1}^+) \ , \quad B_3^+ = \bar{\eta}A_1^+ + A_{2,1}^+ \ .
\end{aligned} \tag{20}
$$

Field equations:

$$\begin{aligned} &-H^+_{2,3} + \bar{\eta} H^+_3 - \frac{1}{b}\bar{H}_3(b) = D^+_{1,t} \ , \\ &H^+_{1,3} - H^+_{3,1} = D^+_{2,t} \ , \\ &H^+_{2,1} - \bar{\eta} H^+_1 + \frac{1}{b}\bar{H}_1(b) = D^+_{3,t} \ , \\ &D^+_{a,a} + \bar{\eta} D^+_2 - \frac{1}{b}\bar{D}_2(b) = 0 \ , \end{aligned} \tag{21}$$

where

$$\begin{aligned} H^+_i &= \int_1^\infty \bar{H}_i e^{-\bar{\eta} b(\psi-1)} d\psi \ , \\ D^+_i &= \int_1^\infty \bar{D}_i e^{-\bar{\eta} b(\psi-1)} d\psi \ . \end{aligned} \tag{22}$$

Constitutive equations:

$$\begin{aligned} D^+_i &= -\frac{\epsilon_0}{2b\bar{\eta}} A^+_{i,t} \ , \\ H^+_1 &= \frac{1}{2b\bar{\eta}\mu_0}(-\bar{\eta} A^+_3 - A^+_{2,3}) \ , \\ H^+_2 &= \frac{k}{2b\bar{\eta}\mu_0}(A^+_{1,3} - A^+_{3,1}) \ , \\ H^+_3 &= \frac{1}{2b\bar{\eta}\mu_0}(\bar{\eta} A^+_1 + A^+_{2,1}) \ . \end{aligned} \tag{23}$$

Wave equations of A^+_i:

$$\begin{aligned} &k(A^+_{1,aa} - A^+_{a,a1}) - \bar{\eta} A^+_{2,1} \\ &-\bar{\eta}^2 A^+_1 + 2\bar{\eta}\mu_0 \bar{H}_3(b) = \epsilon_0 \mu_0 A^+_{1,tt} \ , \\ &A^+_{2,aa} + \bar{\eta} A^+_{a,a} = \epsilon_0 \mu_0 A^+_{2,tt} \ , \\ &k(A^+_{3,aa} - A^+_{a,a3}) - \bar{\eta} A^+_{2,3} \\ &-\bar{\eta}^2 A^+_3 - 2\bar{\eta}\mu_0 \bar{H}_1(b) = \epsilon_0 \mu_0 A^+_{3,tt} \ , \\ &A^+_{a,at} + \bar{\eta} A^+_{2,t} + \frac{2\bar{\eta}}{\epsilon_0}\bar{D}_2(b) = 0 \ . \end{aligned} \tag{24}$$

For $x_2 \le -b$:

$$\bar{A}_i = A^-_i(x_1, x_3, t) e^{\bar{\eta} b(\psi+1)} \ , \tag{25}$$

and

$$\begin{aligned} \bar{E}_i &= E^-_i e^{\bar{\eta} b(\psi+1)} \ , \\ \bar{B}_i &= B^-_i e^{\bar{\eta} b(\psi+1)} \ . \end{aligned} \tag{26}$$

Field-potential relations:

$$\begin{aligned} &E^-_i = -A^-_{i,t} \ , \quad B^-_1 = \bar{\eta} A^-_3 - A^-_{2,3} \ , \\ &B^-_2 = k(A^-_{1,3} - A^-_{3,1}) \ , \quad B^-_3 = -\bar{\eta} A^-_1 + A^-_{2,1} \ . \end{aligned} \tag{27}$$

Field equations:

$$\begin{aligned} &-H^-_{2,3} - \bar{\eta} H^-_3 + \frac{1}{b}\bar{H}_3(-b) = D^-_{1,t} \ , \\ &H^-_{1,3} - H^-_{3,1} = D^-_{2,t} \ , \\ &H^-_{2,1} + \bar{\eta}\ H^-_1 - \frac{1}{b}\bar{H}_1(-b) = D^-_{3,t} \ , \\ &D^-_{a,a} - \bar{\eta} D^-_2 + \frac{1}{b}\bar{D}_2(-b) = 0 \ , \end{aligned} \tag{28}$$

where

$$\begin{aligned} H^-_i &= \int_{-\infty}^{-1} \bar{H}_i e^{\bar{\eta} b(\psi+1)} d\psi \ , \\ D^-_i &= \int_{-\infty}^{-1} \bar{D}_i e^{\bar{\eta} b(\psi+1)} d\psi \ . \end{aligned} \tag{29}$$

Constitutive equations:

$$\begin{aligned} D^-_i &= -\frac{\epsilon_0}{2b\bar{\eta}} A^-_{i,t} \ , \\ H^-_1 &= \frac{1}{2b\bar{\eta}\mu_0}(\bar{\eta} A^-_3 - A^-_{2,3}) \ , \\ H^-_2 &= \frac{k}{2b\bar{\eta}\mu_0}(A^-_{1,3} - A^-_{3,1}) \ , \\ H^-_3 &= \frac{1}{2b\bar{\eta}\mu_0}(-\bar{\eta} A^-_1 + A^-_{2,1}) \ . \end{aligned} \tag{30}$$

Wave equations of A^-_i:

$$\begin{aligned} &k(A^-_{1,aa} - A^-_{a,a1}) + \bar{\eta} A^-_{2,1} \\ &-\bar{\eta}^2 A^-_1 - 2\bar{\eta}\mu_0 \bar{H}_3(-b) = \epsilon_0 \mu_0 A^-_{1,tt} \ , \\ &A^-_{2,aa} - \bar{\eta} A^-_{a,a} = \epsilon_0 \mu_0 A^-_{2,tt} \ , \\ &k(A^-_{3,aa} - A^-_{a,a3}) + \bar{\eta} A^-_{2,3} \\ &-\bar{\eta}^2 A^-_3 + 2\bar{\eta}\mu_0 \bar{H}_1(-b) = \epsilon_0 \mu_0 A^-_{3,tt} \ , \\ &A^-_{a,at} - \bar{\eta} A^-_{2,t} - \frac{2\bar{\eta}}{\epsilon_0}\bar{D}_2(-b) = 0 \ . \end{aligned} \tag{31}$$

Substitutions of (8), (18), and (25) into (7), the continuity conditions of the tangential components of vector potential, gives

$$\begin{aligned} A^+_a &= \sum_{m=0}^{\infty} A^{(m)}_a \ , \\ A^-_a &= \sum_{m=0}^{\infty} (-1)^m A^{(m)}_a \ , \end{aligned} \tag{32}$$

and the insertion of (6), the continuity conditions on H_a and D_2, into (14) yields

$$\begin{aligned} \mathcal{H}^{(n)}_a &= \bar{H}_a(b) - (-1)^n \bar{H}_a(-b) \ , \\ \mathcal{D}^{(n)}_2 &= \bar{D}_2(b) - (-1)^n \bar{D}_2(-b) \ . \end{aligned} \tag{33}$$

By substituting $(17)_4$, $(24)_4$, and $(31)_4$ into $(33)_2$, integrating the resulting relation with respect to t, and setting the integration constant to zero for time-varying field, we obtain

$$
\begin{aligned}
&\frac{1}{2b}\left[A_2^+ - (-1)^n A_2^-\right] \\
&= -\left[(1+\delta_{n0})\,\epsilon_{ab} A_{b,a}^{(n)} + \sum_{m=0}^{\infty} \gamma_{(m+1)n}\epsilon_{a2} A_{2,a}^{(m)}\right] \\
&\quad + \frac{n\pi}{2b}\left(\epsilon_{22} A_2^{(n-1)} + \sum_{m=0}^{\infty} \gamma_{nm}\epsilon_{2a} A_a^{(m)}\right) \\
&\quad - \frac{1}{\bar{\eta} b}\sum_{m=0}^{\infty} \hat{\delta}_{mn} A_{a,a}^{(m)} ,
\end{aligned}
\tag{34}
$$

where

$$
\hat{\delta}_{mn} = \frac{1}{2}[1+(-1)^{m+n}] = \begin{cases} 1, & m+n = \text{even} \\ 0, & m+n = \text{odd.} \end{cases} \tag{35}
$$

By further substituting the rest of (17), (24), and (31) into $(33)_1$, using the relations (32) and (34), and listing the resulting equations together with $(17)_2$, we have

$$
\begin{aligned}
&(1+\delta_{n0})k(A_{3,aa}^{(n)} - A_{a,a3}^{(n)}) - \left(\frac{n\pi}{2b}\right)^2 A_3^{(n)} \\
&+\frac{1}{\bar{\eta} b}\sum_{m=0}^{\infty} \hat{\delta}_{mn}\left[k A_{3,aa}^{(m)} + (1-k) A_{a,a3}^{(m)} - \bar{\eta}^2 A_3^{(m)}\right] \\
&-\frac{n\pi}{2b}\left[(\epsilon_{22}-1) A_{2,3}^{(n-1)} + \sum_{m=0}^{\infty} \gamma_{nm}\epsilon_{2a} A_{a,3}^{(m)}\right] \\
&+(1+\delta_{n0})\epsilon_{ab} A_{b,a3}^{(n)} + \sum_{m=0}^{\infty} \gamma_{(m+1)n}\epsilon_{a2} A_{2,a3}^{(m)} \\
&= \mu_0\epsilon_0\left[(1+\delta_{n0})\epsilon_{3a} A_{a,tt}^{(n)} + \epsilon_{32}\sum_{m=0}^{\infty} \gamma_{(m+1)n} A_{2,tt}^{(m)} \right. \\
&\qquad \left. + \frac{1}{\bar{\eta} b}\sum_{m=0}^{\infty} \hat{\delta}_{mn} A_{3,tt}^{(m)}\right] ,
\end{aligned}
$$

$$
A_{2,aa}^{(n-1)} - \frac{n\pi}{2b} A_{a,a}^{(n)} = \mu_0\epsilon_0\left[\epsilon_{22} A_{2,tt}^{(n-1)} + \sum_{m=0}^{\infty} \gamma_{nm}\epsilon_{2a} A_{a,tt}^{(m)}\right] , \tag{36}
$$

$$
\begin{aligned}
&(1+\delta_{n0})k(A_{1,aa}^{(n)} - A_{a,a1}^{(n)}) - \left(\frac{n\pi}{2b}\right)^2 A_1^{(n)} \\
&+\frac{1}{\bar{\eta} b}\sum_{m=0}^{\infty} \hat{\delta}_{mn}\left[k A_{1,aa}^{(m)} + (1-k) A_{a,a1}^{(m)} - \bar{\eta}^2 A_1^{(m)}\right] \\
&-\frac{n\pi}{2b}\left[(\epsilon_{22}-1) A_{2,1}^{(n-1)} + \sum_{m=0}^{\infty} \gamma_{nm}\epsilon_{2a} A_{a,1}^{(m)}\right] \\
&+(1+\delta_{n0})\epsilon_{ab} A_{b,a1}^{(n)} + \sum_{m=0}^{\infty} \gamma_{(m+1)n}\epsilon_{a2} A_{2,a1}^{(m)} \\
&= \mu_0\epsilon_0\left[(1+\delta_{n0})\epsilon_{1a} A_{a,tt}^{(n)} + \epsilon_{12}\sum_{m=0}^{\infty} \gamma_{(m+1)n} A_{2,tt}^{(m)} \right. \\
&\qquad \left. + \frac{1}{\bar{\eta} b}\sum_{m=0}^{\infty} \hat{\delta}_{mn} A_{1,tt}^{(m)}\right] ,
\end{aligned}
$$

for $n = 0, 1, 2, \ldots$. These are the two-dimensional nth-order equations of $A_i^{(n)}$ for trapped or guided waves in the anisotropic dielectric plate surrounded by free space. In (36), the interaction of the field in the plate with the evanescent field in free space has been taken into account by satisfying the continuity conditions at the faces of the plate.

4 Dispersion Relations

In the case of straight-crested waves propagating in the x_1 direction, i.e. $A_i^{(p)} = A_i^{(p)}(x_1, t)$, (36) and (10) reduce to

$$
\begin{aligned}
&-\left(\frac{n\pi}{2b}\right)^2 A_1^{(n)} + \frac{1}{\bar{\eta} b}\sum_{m=0}^{\infty} \hat{\delta}_{mn}\left[A_{1,11}^{(m)} - \bar{\eta}^2 A_1^{(m)}\right] \\
&-\frac{n\pi}{2b}\left[(\epsilon_{22}-1) A_{2,1}^{(n-1)} + \sum_{m=0}^{\infty} \gamma_{nm}\epsilon_{2a} A_{a,1}^{(m)}\right] \\
&+(1+\delta_{n0})\epsilon_{1a} A_{a,11}^{(n)} + \epsilon_{12}\sum_{m=0}^{\infty} \gamma_{(m+1)n} A_{2,11}^{(m)} \\
&= \mu_0\epsilon_0\left[(1+\delta_{n0})\epsilon_{1a} A_{a,tt}^{(n)} + \epsilon_{12}\sum_{m=0}^{\infty} \gamma_{(m+1)n} A_{2,tt}^{(m)} \right. \\
&\qquad \left. + \frac{1}{\bar{\eta} b}\sum_{m=0}^{\infty} \hat{\delta}_{mn} A_{1,tt}^{(m)}\right] , \\
&A_{2,11}^{(n-1)} - \tfrac{n\pi}{2b} A_{1,1}^{(n)} \\
&= \mu_0\epsilon_0\left(\epsilon_{22} A_{2,tt}^{(n-1)} + \sum_{m=0}^{\infty} \gamma_{nm}\epsilon_{2a} A_{a,tt}^{(m)}\right) , \\
&(1+\delta_{n0})k A_{3,11}^{(n)} - \left(\frac{n\pi}{2b}\right)^2 A_3^{(n)} \\
&+\frac{1}{\bar{\eta} b}\sum_{m=0}^{\infty} \hat{\delta}_{mn}\left(k A_{3,11}^{(m)} - \bar{\eta}^2 A_3^{(m)}\right) \\
&= \mu_0\epsilon_0\left[(1+\delta_{n0})\epsilon_{3a} A_{a,tt}^{(n)} + \epsilon_{32}\sum_{m=0}^{\infty} \gamma_{(m+1)n} A_{2,tt}^{(m)} \right. \\
&\qquad \left. + \frac{1}{\bar{\eta} b}\sum_{m=0}^{\infty} \hat{\delta}_{mn} A_{3,tt}^{(m)}\right] ,
\end{aligned}
\tag{37}
$$

and

$$
\begin{aligned}
E_i^{(n)} &= -A_{i,t}^{(n)} , \\
B_1^{(n)} &= \frac{n+1}{2b}\pi A_3^{(n+1)} , \\
B_2^{(n)} &= -k A_{3,1}^{(n)} , \\
B_3^{(n)} &= -\frac{n+1}{2b}\pi A_1^{(n+1)} + A_{2,1}^{(n)} .
\end{aligned}
\tag{38}
$$

We see that $A_3^{(p)}$ is coupled to $A_1^{(p)}$ and $A_2^{(p)}$ through the anisotropy, but $A_1^{(p)}$ and $A_2^{(p)}$ are always coupled even if the material is isotropic.

In anisotropic dielectrics, TE and TM modes are, in general, coupled together and so are the dispersion curves. In order to extract the nth-order governing equations from the infinite system, we employ the following truncation procedures:

1. Consider the governing equations for vector potentials $A_1^{(n-1)}$, $A_1^{(n)}$, $A_2^{(n-1)}$, $A_3^{(n-1)}$, and $A_3^{(n)}$.
2. Disregard all other higher and lower order components in these equations.

Thus for the nth-order equations, we have

$$
\begin{aligned}
&(1+\delta_{n1})\,\epsilon_{1a}A_{a,11}^{(n-1)} + \frac{1}{\bar{\eta}b}A_{1,11}^{(n-1)} \\
&-\left[\left(\frac{n-1}{2b}\pi\right)^2 + \frac{\bar{\eta}}{b}\right]A_1^{(n-1)} \\
&+\epsilon_{12}\gamma_{n(n-1)}A_{2,11}^{(n-1)} - \frac{(n-1)\pi}{2b}\gamma_{(n-1)n}\epsilon_{2a}A_{a,1}^{(n)} \\
&= \mu_0\epsilon_0\left[(1+\delta_{n1})\,\epsilon_{1a}A_{a,tt}^{(n-1)}\right. \\
&\qquad \left. + \frac{1}{\bar{\eta}b}A_{1,tt}^{(n-1)} + \epsilon_{12}\gamma_{n(n-1)}A_{2,tt}^{(n-1)}\right], \\
&A_{2,11}^{(n-1)} - \frac{n\pi}{2b}A_{1,1}^{(n)} \\
&= \mu_0\epsilon_0\left[\epsilon_{22}A_{2,tt}^{(n-1)} + \gamma_{n(n-1)}\epsilon_{2a}A_{a,tt}^{(n-1)}\right], \\
&\left(1+\delta_{n1}+\frac{1}{\bar{\eta}b}\right)kA_{3,11}^{(n-1)} - \left[\left(\frac{n-1}{2b}\pi\right)^2 + \frac{\bar{\eta}}{b}\right]A_3^{(n-1)} \\
&= \mu_0\epsilon_0\left[(1+\delta_{n1})\,\epsilon_{3a}A_{a,tt}^{(n-1)}\right. \\
&\qquad \left. + \epsilon_{32}\gamma_{n(n-1)}A_{2,tt}^{(n-1)} + \frac{1}{\bar{\eta}b}A_{3,tt}^{(n-1)}\right], \\
&\left(\epsilon_{11}+\frac{1}{\bar{\eta}b}\right)A_{1,11}^{(n)} - \left[\left(\frac{n\pi}{2b}\right)^2 + \frac{\bar{\eta}}{b}\right]A_1^{(n)} + \epsilon_{13}A_{3,11}^{(n)} \\
&-\frac{n\pi}{2b}\left[\gamma_{n(n-1)}\epsilon_{2a}A_{a,1}^{(n-1)} + (\epsilon_{22}-1)A_{2,1}^{(n-1)}\right] \\
&= \mu_0\epsilon_0\left[\left(\epsilon_{11}+\frac{1}{\bar{\eta}b}\right)A_{1,tt}^{(n)} + \epsilon_{13}A_{3,tt}^{(n)}\right], \\
&\left(1+\frac{1}{\bar{\eta}b}\right)kA_{3,11}^{(n)} - \left[\left(\frac{n\pi}{2b}\right)^2 + \frac{\bar{\eta}}{b}\right]A_3^{(n)} \\
&= \mu_0\epsilon_0\left(\epsilon_{3a}A_{a,tt}^{(n)} + \frac{1}{\bar{\eta}b}A_{3,tt}^{(n)}\right),
\end{aligned}
\tag{39}
$$

where $n = 1, 2, 3, \ldots$. By letting

$$
A_i^{(n)} = C_i^{(n)}e^{i(\xi x_1 - \omega t)} = C_i^{(n)}e^{i\xi(x_1 - vt)}, \tag{40}
$$

where ξ is the wave number, ω the angular frequency, v the phase velocity, and $C_i^{(n)}$ a constant, we have the dispersion relation

$$
\det\begin{bmatrix}
\epsilon_{11}(1+\delta_{n1})\bar{\lambda}^2 + \frac{(n-1)^2}{X^2} + \frac{4\bar{\lambda}}{\pi X} & \epsilon_{12}\gamma_{n(n-1)}\bar{\lambda}^2 \\
\epsilon_{21}\gamma_{n(n-1)}V^2 & \epsilon_{22}V^2 - 1 \\
(1+\delta_{n1})\,\epsilon_{31}V^2 & \epsilon_{32}\gamma_{n(n-1)}V^2 \\
-\epsilon_{21}\gamma_{n(n-1)}\frac{n}{X} & -(\epsilon_{22}-1)\frac{n}{X} \\
0 & 0
\end{bmatrix}
$$

$$
\begin{bmatrix}
(1+\delta_{n1})\,\epsilon_{13}\bar{\lambda}^2 \\
\epsilon_{23}\gamma_{n(n-1)}V^2 \\
(1+\delta_{n1})(\epsilon_{33}V^2 - k) + \frac{2(2V^2-k-1)}{\pi\bar{\lambda}X} - \frac{(n-1)^2}{X^2} \\
-\epsilon_{23}\gamma_{n(n-1)}\frac{n}{X} \\
0
\end{bmatrix}
$$

$$
\begin{bmatrix}
\epsilon_{21}\gamma_{(n-1)n}\frac{n-1}{X} & \epsilon_{23}\gamma_{(n-1)n}\frac{n-1}{X} \\
-\frac{n}{X} & 0 \\
0 & 0 \\
\epsilon_{11}\bar{\lambda}^2 + \frac{n^2}{X^2} + \frac{4\bar{\lambda}}{\pi X} & \epsilon_{13}\bar{\lambda}^2 \\
\epsilon_{31}V^2 & \epsilon_{33}V^2 - k + \frac{2(2V^2-k-1)}{\pi\bar{\lambda}X} - \frac{n^2}{X^2}
\end{bmatrix} = 0, \tag{41}
$$

where

$$
X = \frac{2\xi b}{\pi}, \quad V = \frac{v}{c_0}, \quad \bar{\lambda} = \frac{\bar{\eta}}{\xi}, \quad c_0 = \left(\sqrt{\mu_0\epsilon_0}\right)^{-1}. \tag{42}
$$

For a given X, (41) has three roots in V. The two smaller roots give the nth pair of dispersion curves.

For numerical computations, we consider a doubly rotated cut of lithium niobate $(yxlt)45°/5°$ for which ϵ_{ij} has the value

$$
[\epsilon_{ij}] = \begin{bmatrix} 43.878 & -0.647 & -0.650 \\ -0.647 & 36.537 & -7.426 \\ -0.650 & -7.426 & 36.481 \end{bmatrix}. \tag{43}
$$

In order to get better fitting with the exact dispersion curves, the correction factors for different pairs of dispersion curves are determined by letting

$$
V_\infty = V|_{X\to\infty} = \begin{cases} 0.970V_{b1} \text{ for } n=1 \\ 0.975V_{b1} \text{ for } n=2 \\ 1.0V_{b1} \text{ for } n=3 \end{cases}, \tag{44}
$$

where V_{b1} is the first velocity bound for the anisotropic dielectric plate introduced in a previous work[3], where procedures for the determination of the velocity bounds were given in terms of the relative dielectric impermeability tensor β_{ij} which is the inverse of ϵ_{ij}. For the convenience of reference, we list the procedures in terms of ϵ_{ij} in the Appendix.

From (41) and (44), we have

$$k = \left\{ \left[(1+\delta_{n1})\,\epsilon_{11}\epsilon_{22}\epsilon_{33} + 2\epsilon_{12}\epsilon_{23}\epsilon_{31}\gamma_{n(n-1)}^2 - (1+\delta_{n1})\,\epsilon_{22}\epsilon_{13}^2 - \epsilon_{11}\epsilon_{23}^2\gamma_{n(n-1)}^2 - \epsilon_{33}\epsilon_{12}^2\gamma_{n(n-1)}^2 \right] V_\infty^4 + (1+\delta_{n1})\left(\epsilon_{13}^2 - \epsilon_{11}\epsilon_{33}\right) V_\infty^2 \right\} / \left[(1+\delta_{n1})\,\epsilon_{11}\left(\epsilon_{22}V_\infty^2 - 1\right) - \epsilon_{12}^2\gamma_{n(n-1)}^2 V_\infty^2 \right] . \tag{45}$$

We note that, in (43), ϵ_{12} and ϵ_{13} are much smaller than ϵ_{11}, ϵ_{22}, ϵ_{33}, and ϵ_{23}. Hence, the coupling of $A_1^{(n-1)}$ and $A_3^{(n)}$ with $A_2^{(n-1)}$ is very weak and can be neglected. Hence, by dropping equations $(39)_1$ and $(39)_5$ and the terms $A_1^{(n-1)}$ and $A_3^{(n)}$ in the remaining equations of (39). We have

$$\begin{aligned}
&A_{2,11}^{(n-1)} - \frac{n\pi}{2b}A_{1,1}^{(n)} \\
&= \mu_0\epsilon_0\left[\epsilon_{22}A_{2,tt}^{(n-1)} + \epsilon_{23}\gamma_{n(n-1)}A_{3,tt}^{(n-1)}\right] , \\
&\left(1+\delta_{n1}+\frac{1}{\bar\eta b}\right)kA_{3,11}^{(n-1)} \\
&\quad - \left[\left(\frac{n-1}{2b}\pi\right)^2 + \frac{\bar\eta}{b}\right]A_3^{(n-1)} \\
&= \mu_0\epsilon_0\left\{\epsilon_{32}\gamma_{n(n-1)}A_{2,tt}^{(n-1)} + \left[\epsilon_{33}\left(1+\delta_{n1}\right) + \frac{1}{\bar\eta b}\right]A_{3,tt}^{(n-1)}\right\} , \\
&\left(\epsilon_{11}+\frac{1}{\bar\eta b}\right)A_{1,11}^{(n)} - \left[\left(\frac{n\pi}{2b}\right)^2 + \frac{\bar\eta}{b}\right]A_1^{(n)} \\
&-\frac{n\pi}{2b}\left[(\epsilon_{22}-1)A_{2,1}^{(n-1)} + \epsilon_{23}\gamma_{n(n-1)}A_{3,1}^{(n-1)}\right] \\
&\quad = \mu_0\epsilon_0\left(\epsilon_{11}+\frac{1}{\bar\eta b}\right)A_{1,tt}^{(n)} ,
\end{aligned} \tag{46}$$

the dispersion relation from (46) is

$$\det\begin{bmatrix} \epsilon_{23}V^2\gamma_{n(n-1)} \\ (1+\delta_{n1})(\epsilon_{33}V^2 - k) + \frac{2(2V^2-k-1)}{\pi\bar\lambda X} - \frac{(n-1)^2}{X^2} \\ -\frac{n}{X}\epsilon_{23}\gamma_{n(n-1)} \end{bmatrix} \begin{bmatrix} -\frac{n}{X} & \epsilon_{22}V^2 - 1 \\ 0 & \epsilon_{32}V^2\gamma_{n(n-1)} \\ \epsilon_{11}\bar\lambda^2 + \frac{n^2}{X^2} + \frac{4\bar\lambda}{\pi X} & -(\epsilon_{22}-1)\frac{n}{X} \end{bmatrix} = 0 , \tag{47}$$

and the correction factors are

$$k = \left\{ \left[(1+\delta_{n1})\,\epsilon_{22}\epsilon_{33} - \epsilon_{23}^2\gamma_{n(n-1)}^2 \right] V_\infty^4 - (1+\delta_{n1})\,\epsilon_{33}V_\infty^2 \right\} / \left[(1+\delta_{n1})\left(\epsilon_{22}V_\infty^2 - 1\right) \right] . \tag{48}$$

For a given X, (47) has only two positive roots in V, which gives us the nth pair of dispersion curves.

The first three pairs of dispersion curves computed from (47) are plotted as solid curves in Fig.2 to Fig.4 for $X \leq 2$. The correction factors from (48) for $n = 1, 2, 3$ are respectively 0.877, 0.880, and 0.931. For the convenience of comparison, the dispersion curves obtained from the exact solutions of the three-dimensional Maxwell's equations are plotted in Fig.5. Those dashed lines in Fig.2 to Fig.5 correspond to $V = V_{b1}$, $V = V_{b2}$, and $V = 1$ as explained in Ref. 3. We can see that the dispersion curves from the truncated two-dimensional equations give good approximations.

5 Conclusions

A system of two-dimensional equations is derived for the guided EM waves in an anisotropic dielectric plate surrounded by free space. The interactions between the vibrating field in the plate and the evanescent field in free space are accommodated by satisfying the continuity conditions at the faces of the plate.

Dispersion relations for straight-crested waves are obtained from the two dimensional equations and computed for the doubly rotated cut of lithium niobate $(yxlt)45°/5°$. These dispersion curves are then compared with the corresponding ones obtained from the exact solutions of the three-dimensional Maxwell's equations. It is shown that these approximate dispersion curves are very close to the exact ones.

Acknowledgements

This work was supported by a grant (Contract No. DAAH 04-93-6-0081) from the U. S. Army Research Office.

Appendix

For the determination of V_{b1}, we start from the Maxwell's equations and constitutive equations. Substituting $(1)_1$ into $(2)_1$ and using (3), we have the wave equations in terms of E_i

$$E_{i,kk} - E_{k,ki} = \mu_0\epsilon_0\epsilon_{ij}E_{j,tt} . \tag{49}$$

For straight-crested wave propagating in x_1 direction, i.e. $E_i = E_i(x_1, x_2, t)$, (49) can be written as

$$\begin{aligned}
&E_{1,22} - E_{2,21} = \mu_0\epsilon_0\epsilon_{1j}E_{j,tt} , \\
&-E_{1,12} + E_{2,11} = \mu_0\epsilon_0\epsilon_{2j}E_{j,tt} , \\
&E_{3,11} + E_{3,22} = \mu_0\epsilon_0\epsilon_{3j}E_{j,tt} .
\end{aligned} \tag{50}$$

Let the electric field E_j be the form

$$E_j = C_j e^{i(\xi x_1 + \eta x_2 - \omega t)} = C_j e^{i\xi(x_1 + \lambda x_2 - vt)} , \quad (51)$$

where ξ and η are the wave numbers in the x_1 and x_2 directions, ω the angular frequency, v the phase velocity, and $\lambda = \eta/\xi$ the ratio of the wave numbers. Substituting (51) into (50), we have

$$\begin{bmatrix} \epsilon_{11}V^2 - \lambda^2 & \epsilon_{12}V^2 + \lambda & \epsilon_{13}V^2 \\ \lambda + \epsilon_{12}V^2 & \epsilon_{22}V^2 - 1 & \epsilon_{23}V^2 \\ \epsilon_{13}V^2 & \epsilon_{23}V^2 & \epsilon_{33}V^2 - \lambda^2 - 1 \end{bmatrix} \begin{Bmatrix} C_1 \\ C_2 \\ C_3 \end{Bmatrix} = 0, \quad (52)$$

where X and V, as defined in (42), are the dimensionless wave number and the dimensionless phase velocity, and Ω is the dimensionless angular frequency defined by

$$\Omega = \frac{w}{(\pi/2b)c_0} . \quad (53)$$

Vanishing of the determinant of the coefficient matrix gives

$$\begin{aligned} &\epsilon_{22}\lambda^4 + 2\epsilon_{12}\lambda^3 + [\epsilon_{11} + \epsilon_{22} \\ &-(\epsilon_{11}\epsilon_{22} - \epsilon_{12}^2 + \epsilon_{22}\epsilon_{33} - \epsilon_{23}^2)V^2]\,\lambda^2 \\ &+2\left[(\epsilon_{13}\epsilon_{23} - \epsilon_{12}\epsilon_{33})V^2 + \epsilon_{12}\right]\lambda \\ &+(\epsilon_{11}\epsilon_{22}\epsilon_{33} + 2\epsilon_{13}\epsilon_{23}\epsilon_{12} - \epsilon_{12}^2\epsilon_{33} - \epsilon_{13}^2\epsilon_{22} - \epsilon_{23}^2\epsilon_{11})V^4 \\ &-(\epsilon_{11}\epsilon_{22} + \epsilon_{11}\epsilon_{33} - \epsilon_{12}^2 - \epsilon_{13}^2)V^2 + \epsilon_{11} = 0 . \end{aligned} \quad (54)$$

Eq. (54) is a quartic equation in λ. The characteristic of its roots depends on ϵ_{ij} as well as on V. For a crystal dielectric plate of a given cut orientation, ϵ_{ij} have fixed values. Hence, we regard (54) as a function of V.

Let V_b denote the dimensionless phase velocity at which λ has repeated roots. Then these repeated roots of (54) must also satisfy the equation that is obtined by taking the derivative of (54) with respect to λ and can be written as

$$\begin{aligned} &\left[(\epsilon_{11}\epsilon_{22} + \epsilon_{22}\epsilon_{33} - \epsilon_{12}^2 - \epsilon_{23}^2)\,\lambda - \epsilon_{23}\epsilon_{31} + \epsilon_{12}\epsilon_{33}\right] V^2 \\ &= 2\epsilon_{22}\lambda^3 + 3\epsilon_{12}\lambda^2 + (\epsilon_{11} + \epsilon_{22})\,\lambda + \epsilon_{12} . \end{aligned} \quad (55)$$

Substitution of (55) into (54) results in a sixth-order equation of λ from which roots are computed numerically. By inserting the repeated roots of λ back into (55) to compute V^2 and retaining only the real values of V_b for propagating waves, we obtain, in general, two values that are designated by V_{b1} and V_{b2} such that

$$1 > V_{b2} > V_{b1} > 0 . \quad (56)$$

For the doubly rotated cut of lithium niobate $(yxlt)45°/5°$ with the relative dielectric permittivity tensor given in (43), $V_{b1} = 0.1509$ and $V_{b2} = 0.1855$.

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.D. Yu, "Guided Electromagnetic Waves in Anisotropic Dielectric Plates," *J. Appl. Phys.* 74(8), pp.4823-4839, 1993.

[4] R. D. Richtmyer, "Dielectric resonators," *J. Appl. Phys.*, vol. 10, pp. 391-398, June 1939.

[5] P. C. Y. Lee and J. S. Yang, "Vibrations of circular disk dielectric resonators," *J. Appl. Phys.*, **73** (11), pp. 7083-7092, 1993.

[6] 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.

[7] M. Born and E. Wolf, *Principles of Optics*, 3rd. ed., Pergarmon Press, New York, 1965, pp. 71-73.

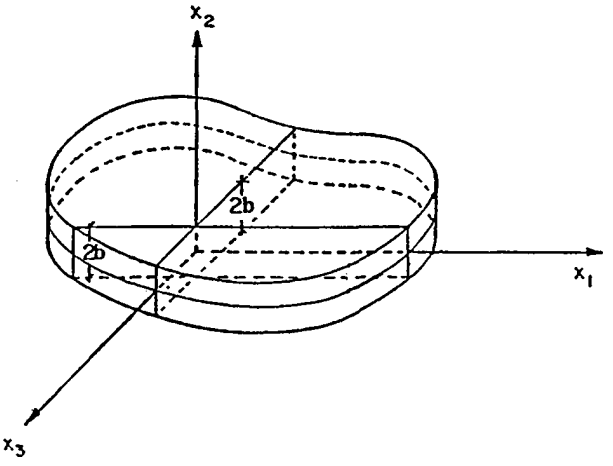


Figure 1: A dielectric plate surrounded by free space.

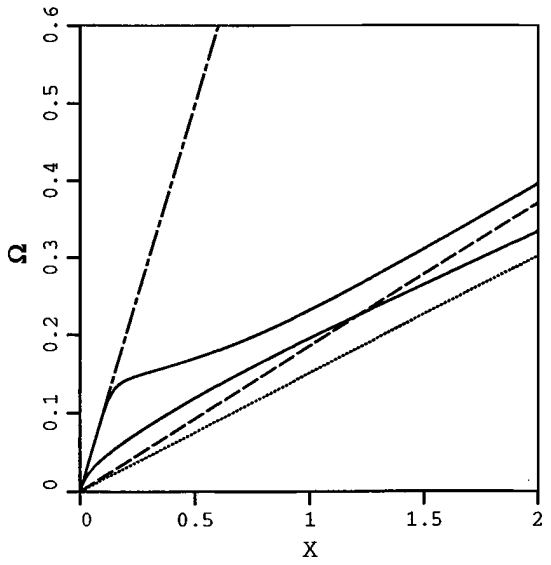


Figure 2: The first pair of dispersion curves for guided EM waves in a dielectric plate defined by (43), according to the 2-D equations.

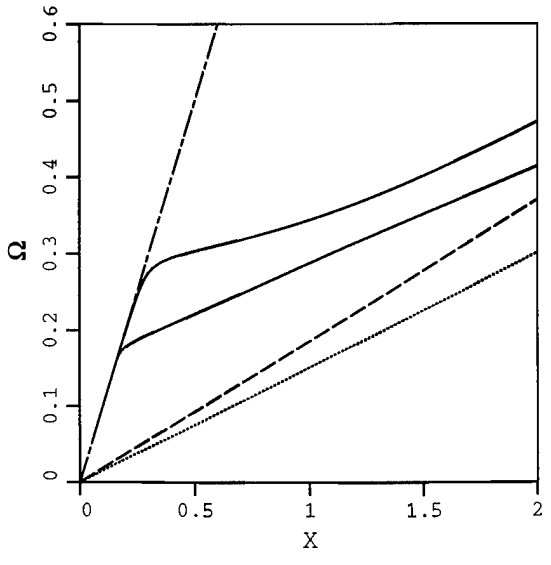


Figure 3: The second pair of dispersion curves for guided EM waves in a dielectric plate defined by (43), according to the 2-D equations.

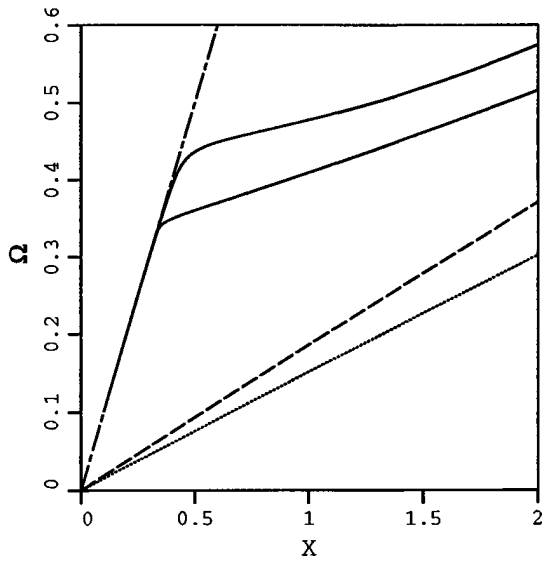


Figure 4: The third pair of dispersion curves for guided EM waves in a dielectric plate defined by (43), according to the 2-D equations.

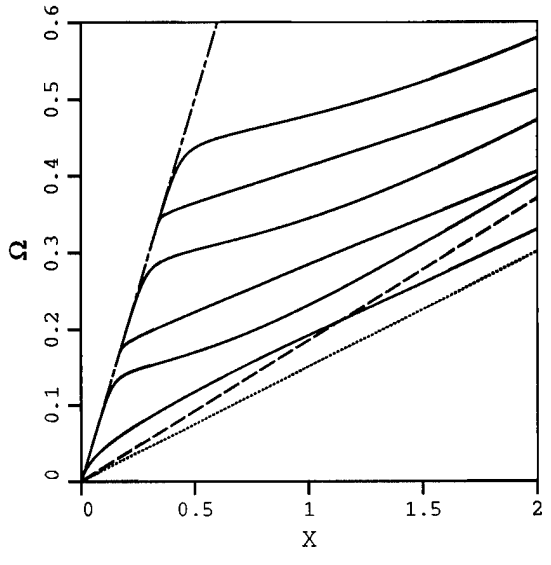


Figure 5: The first three pairs of dispersion curves for guided EM waves in a dielectric plate defined by (43), according to the 3-D equations.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THE EFFECT OF THERMAL DISSIPATION IN THE VIBRATIONS OF A PIEZOELECTRIC BODY

J. S. Yang and R. C. Batra

Department of Mechanical and Aerospace Engineering and Engineering Mechanics
University of Missouri-Rolla, Rolla, Missouri 65401-0249, USA

Abstract

We study the effect of heat conduction on the shift in the frequencies of a freely vibrating linear piezoelectric body by using two perturbation methods. It is shown that the first order effect on frequencies is to shift them by a small imaginary number signifying thereby that the effect of energy dissipation due to heat conduction is to reduce the amplitude of vibration.

I. Introduction

Piezoelectric materials are often used as resonators whose frequencies need to be precisely controlled. Because of the coupling between the thermoelastic and pyroelectric effects, it is important to quantify the effect of heat dissipation on the shift in the frequencies of a piezoelectric body. We first note that the non-dimensional parameter associated with heat conduction is very small, and then use the perturbation method to define zeroth and first order problems. A solution of these problems determines the relative shift in the frequencies caused by the heat conduction.

A thermopiezoelectricity theory was first proposed by Mindlin [1] who subsequently also derived equations governing the deformations of a thermopiezoelectric plate [2]. Several general results for thermopiezoelectric bodies have been proved by W. Nowacki [3-8], J. Nowacki [9], Majorkowska-Knap [10,11], Nowinski [12], and Iesan [13]. Chandrasekharaiah [14,15] has generalized Mindlin's theory of thermopiezoelectricity to account for finite speed of propagation of thermal disturbances. Several investigators [16-22] have studied the propagation of waves in plates, cylinders and general 3-dimensional bodies made of thermopiezoelectric materials. Tauchert [23] has recently applied thermopiezoelectricity theory to composite plates.

Here we employ Mindlin's theory to analyze the effect of energy dissipated due to heat conduction on the frequencies of a freely vibrating piezoelectric body. It is expected that the thermal dissipation will damp out the vibrations. One of our objectives is to show that this is indeed true and we ascertain the damping caused by heat conduction. Our analysis of the problem is based on the observation that for high frequency vibrations, the effect of heat conduction is small. Thus the perturbation method in which the solution is expanded is terms of a small parameter associated with heat conduction can be used to study the problem. The zeroth order perturbation problem corresponds to adiabatic deformations of the body in which the heat generation shifts the frequencies by real numbers. However, in the first order problem, thermal dissipation alters the frequencies by imaginary numbers which reveals the damping effect.

II. Formulation of the Problem

Let the spatial region occupied by the piezoelectric body be V, the boundary of V be S, the unit outward normal to S be n_i, and S be partitioned as

$$\begin{aligned} S_u \cup S_T &= S_\phi \cup S_D = S_\theta \cup S_q = S \\ S_u \cap S_T &= S_\phi \cap S_D = S_\theta \cap S_q = \emptyset \,. \end{aligned} \tag{1}$$

In rectangular Cartesian coordinates, the governing equations for the free vibrations of a linear thermopiezoelectric body can be written as [1]

$$T_{ij,i} = \rho \ddot{u}_j$$
$$D_{i,i} = 0$$
$$\dot{\sigma} = -\frac{q_{i,i}}{\Theta}$$

0-7803-1945-1/94 $3.00 © 1994 IEEE

$$
\begin{aligned}
&T_{ij} = c_{ijkl}S_{kl} - e_{kij}E_k - \lambda_{ij}\theta \ , \quad S_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i}) \\
&D_i = e_{ijk}S_{jk} + \epsilon_{ij}E_j + p_i\theta \ , \quad E_i = -\phi_{,i} \\
&\sigma = \lambda_{ij}S_{ij} + p_iE_i + \alpha\theta \ , \\
&q_i = -\kappa_{ij}\theta_{,j} \ ,
\end{aligned} \tag{2}
$$

with boundary conditions

$$
\begin{aligned}
&u_i = 0 \quad \text{on} \quad S_u \\
&T_{ji}n_j = (c_{ijkl}S_{kl} - e_{kij}E_k - \lambda_{ij}\theta)n_j = 0 \quad \text{on} \quad S_T \\
&\phi = 0 \quad \text{on} \quad S_\phi \\
&D_in_i = (e_{ijk}S_{jk} + \epsilon_{ij}E_j + p_i\theta)n_i = 0 \quad \text{on} \quad S_D \\
&\theta = 0 \quad \text{on} \quad S_\theta \\
&q_in_i = -\kappa_{ij}\theta_{,j}n_i = 0 \quad \text{on} \quad S_q \ .
\end{aligned} \tag{3}
$$

In equations (2) and (3), ρ is the mass density, u_i mechanical displacement, S_{ij} the infinitesimal strain tensor, T_{ij} the stress tensor, E_i electric field, D_i electric displacement, ϵ_{ijk} the permutation tensor, θ the change in the temperature of a material particle, Θ the reference uniform temperature of the body, q_i the heat flux, σ the entropy, κ_{ij} the heat conduction tensor, c_{ijkl} isothermal elastic tensor, e_{kij} piezoelectric moduli, λ_{ij} thermal stress moduli, ϵ_{ij} electric permittivity, p_i pyroelectric moduli, and α is related to the specific heat. Furthermore, a repeated index implies summation over the range of the index and a superimposed dot indicates the time derivative. Equation $(2)_1$ expresses the balance of linear momentum, and $(2)_3$ the balance of internal energy. Equation $(2)_2$ is Gauss's equation. Equation set (2) is within the quasi-static theory of piezoelectricity. For resonators with frequency usually in the order of 10^6 Hz the quasi-static theory of piezoelectricity is very accurate approximation [24]. The boundary conditions (3) signify that the part S_u of the boundary of the body is clamped, S_T is traction free, S_q is thermally insulated, S_θ is kept at the initial uniform temperature of the body, the normal component of the electric displacement vanishes on S_D and the electric potential is zero on S_ϕ. Thus there is no mechanical work done by external forces. When S_θ is not empty, the body can exchange heat with the surroundings through its points on S_θ.

Substitution from equations $(2)_4$ - $(2)_7$ into the balance laws $(2)_1$ - $(2)_3$ gives

$$
\begin{aligned}
&c_{ijkl}u_{k,li} + e_{kij}\phi_{,ki} - \lambda_{ij}\theta_{,i} = \rho\ddot{u}_j \\
&e_{kij}u_{i,jk} - \epsilon_{ij}\phi_{,ij} + p_i\theta_{,i} = 0 \\
&\lambda_{ij}\dot{u}_{i,j} - p_i\dot{\phi}_{,i} + \alpha\dot{\theta} = \frac{\kappa_{ij}}{\Theta}\theta_{,ij}
\end{aligned} \tag{4}
$$

We look for its solutions of the form

$$
\begin{aligned}
u(\mathbf{x},t) &= \tilde{\mathbf{u}}(\mathbf{x})e^{i\omega t} \\
\phi(\mathbf{x},t) &= \tilde{\phi}(\mathbf{x})e^{i\omega t} \\
\theta(\mathbf{x},t) &= \tilde{\theta}(\mathbf{x})e^{i\omega t}
\end{aligned} \tag{5}
$$

which when substituted in (4) give

$$
\begin{aligned}
&-c_{ijkl}u_{k,li} - e_{kij}\phi_{,ki} + \lambda_{ij}\theta_{,i} = \rho\omega^2 u_j \\
&-e_{kij}u_{i,jk} + \epsilon_{ij}\phi_{,ij} - p_i\theta_{,i} = 0 \\
&-\frac{\lambda_{ij}}{\alpha}u_{i,j} + \frac{p_i}{\alpha}\phi_{,i} - \theta = i\frac{\kappa}{\alpha\Theta\Omega L^2}\frac{\kappa_{ij}}{\kappa}\frac{\Omega}{\omega}L^2\theta_{,ij} \ .
\end{aligned} \tag{6}
$$

In equation $(6)_3$, we have introduced a reference frequency Ω, a characteristic length L, set

$$
\kappa = \max_{ij}|\kappa_{ij}| \ , \tag{7}
$$

and have dropped the superimposed tildes.

III. Perturbation Method

In equation $(6)_3$, $\kappa/\alpha\Theta\Omega L^2$ is a dimensionless number. For an AT cut quartz at 298.2 °K, we have [25]

$$
\begin{aligned}
&\kappa \approx 10^6 \quad erg/cm \cdot sec \ ^\circ K \\
&\alpha\Theta \approx 10^7 \quad dyn/cm^2 \ ^\circ K \ .
\end{aligned} \tag{8}
$$

Then for $L = 1cm$ and $\Omega = 10^6/sec$, which is characteristic of a resonator, $\kappa/\alpha\Theta\Omega L^2 \approx 10^{-7}$ is a very small number, especially for high frequency vibrations with large Ω. Denoting

$$
\epsilon = \frac{\kappa}{\alpha\Theta\Omega L^2}
$$

we can write equation $(6)_3$ as

$$
-\frac{\lambda_{ij}}{\alpha}u_{i,j} + \frac{p_i}{\alpha}\phi_{,i} - \theta = i\epsilon\frac{\kappa_{ij}}{\kappa}\frac{\Omega}{\omega}L^2\theta_{,ij} \ . \tag{9}
$$

Since ϵ is very small, a perturbation method with respect to ϵ may be used to solve equations $(6)_1$, $(6)_2$ and (9). We assume that

$$
\begin{aligned}
\omega &= \omega^o + \epsilon\omega^* + O(\epsilon^2) \\
u_i &= u_i^o + \epsilon u_i^* + O(\epsilon^2) \\
\phi &= \phi^o + \epsilon\phi^* + O(\epsilon^2) \\
\theta &= \theta^o + \epsilon\theta^* + O(\epsilon^2)
\end{aligned} \tag{10}
$$

Substitution from equation (10) into $(6)_1$, $(6)_2$, (9) and (3), equating terms of equal powers of ϵ, we obtain the following perturbation problems of successive orders.

Zeroth order problem:

$$
\begin{aligned}
& -c_{ijkl}u^o_{k,li} - e_{kij}\phi^o_{,ki} + \lambda_{ij}\theta^o_{,i} = \rho(\omega^o)^2 u^o_j \\
& -e_{kij}u^o_{i,jk} + \epsilon_{ij}\phi^o_{,ij} - p_i\theta^o_{,i} = 0 \\
& -\frac{\lambda_{ij}}{\alpha}u^o_{i,j} + \frac{p_i}{\alpha}\phi^o_{,i} - \theta^o = 0 \\
& u^o_i = 0 \quad \text{on} \quad S_u \\
& (c_{ijkl}u^o_{k,l} + e_{kij}\phi^o_{,k} - \lambda_{ij}\theta^o)n_j = 0 \quad \text{on} \quad S_T \\
& \phi^o = 0 \quad \text{on} \quad S_\phi \\
& (e_{ijk}u^o_{j,k} - \epsilon_{ij}\phi^o_{,j} + p_i\theta^o)n_i = 0 \quad \text{on} \quad S_D
\end{aligned} \tag{11}
$$

First order problem:

$$
\begin{aligned}
& -c_{ijkl}u^*_{k,li} - e_{kij}\phi^*_{,ki} + \lambda_{ij}\theta^*_{,i} \\
& \quad = \rho(\omega^o)^2 u^*_j + 2\rho\omega^o\omega^* u^o_j \\
& -e_{kij}u^*_{i,jk} + \epsilon_{ij}\phi^*_{,ij} - p_i\theta^*_{,i} = 0 \\
& -\frac{\lambda_{ij}}{\alpha}u^*_{i,j} + \frac{p_i}{\alpha}\phi^*_{,i} - \theta^* = i\frac{\kappa_{ij}}{\kappa}\frac{\Omega}{\omega^o}L^2\theta^o_{,ij} \\
& u^*_i = 0 \quad \text{on} \quad S_u \\
& (c_{ijkl}u^*_{k,l} + e_{kij}\phi^*_{,k} - \lambda_{ij}\theta^*)n_j = 0 \quad \text{on} \quad S_T \\
& \phi^* = 0 \quad \text{on} \quad S_\phi \\
& (e_{ijk}u^*_{j,k} - \epsilon_{ij}\phi^*_{,j} + p_i\theta^*)n_i = 0 \quad \text{on} \quad S_D
\end{aligned} \tag{12}
$$

We note that in the perturbation problems defined by equations (11) and (12) thermal boundary conditions have been dropped. This is because in (9) the small parameter ϵ appears in the term with the highest order derivative of θ. Hence we have a singular perturbation problem in the sense that a boundary layer type solution needs to be constructed near the boundary surface S to satisfy the thermal boundary conditions. In fact, in (11), the thermal boundary condition $q^o_i n_i = 0$ on S_q is satisfied since $q^o_i = 0$. So only a boundary layer near S_θ needs to be constructed. Since boundary layers exist only in a narrow region, and frequencies are global properties, we neglect the effect of such boundary layers on the shift in frequencies caused by the thermal dissipation. The thermal boundary conditions will be considered in another method discussed in Section IV.

III.1 Zeroth Order Problem

The zeroth order problem defined by the equation set (11) represents an adiabatic process characterized by $q^o_i = 0$. This agrees with the usual hypothesis that for high frequency vibrations or fast processes thermal conduction is small and may be neglected. Solving equation $(11)_3$ for θ^o and substituting the result in $(11)_1$ and $(11)_2$, we obtain

$$
\begin{aligned}
& -\bar{c}_{ijkl}u^o_{k,li} - \bar{e}_{kij}\phi^o_{,ki} = \rho(\omega^o)^2 u^o_j \\
& -\bar{e}_{kij}u^o_{i,jk} + \bar{\epsilon}_{ij}\phi^o_{,ij} = 0 \\
& u^o_i = 0 \quad \text{on} \quad S_u \\
& (\bar{c}_{ijkl}u^o_{k,l} + \bar{e}_{kij}\phi^o_{,k})n_j = 0 \quad \text{on} \quad S_T \\
& \phi^o = 0 \quad \text{on} \quad S_\phi \\
& (\bar{e}_{ijk}u^o_{j,k} - \bar{\epsilon}_{ij}\phi^o_{,j})n_i = 0 \quad \text{on} \quad S_D
\end{aligned} \tag{13}
$$

where

$$
\begin{aligned}
\bar{c}_{ijkl} &= c_{ijkl} + \frac{1}{\alpha}\lambda_{ij}\lambda_{kl} \\
\bar{e}_{kij} &= e_{kij} - \frac{1}{\alpha}p_k\lambda_{ij} \\
\bar{\epsilon}_{ij} &= \epsilon_{ij} - \frac{1}{\alpha}p_i p_j
\end{aligned} \tag{14}
$$

are the adiabatic material constants. We note that (13) has the mathematical structure of a piezoelectric problem without thermal effects, except that all of the material constants are changed from isothermal to adiabatic constants. The change will alter the frequencies by real numbers representing a pure shift in the frequencies.

A problem like (13) has been solved for different geometries and under various boundary conditions [26]. In the following, we assume that the zeroth order problem has been solved, thus ω^o, u^o_i, ϕ^o, θ^o are known real quantities.

For later convenience, we introduce the four-vectors **U** and **V**, operators **A** and **B**, and inner product $\langle\ ;\ \rangle$ defined as

$$
\begin{aligned}
& \mathbf{U} = \{u_j\ , \phi\}, \quad \mathbf{V} = \{v_j\ , \psi\} \\
& \mathbf{AU} = \{-\bar{c}_{ijkl}u_{k,li} - \bar{e}_{kij}\phi_{,ki}\ ,\ -\bar{e}_{kij}u_{i,jk} + \bar{\epsilon}_{ij}\phi_{,ij}\} \\
& \mathbf{BU} = \{\rho u_j\ , 0\}
\end{aligned} \tag{15}
$$

$$\langle \mathbf{U} ; \mathbf{V} \rangle = \int_V (u_j v_j + \phi\psi) dV \tag{16}$$

Then (13) can be written as

$$\mathbf{A}\mathbf{U}^o = (\omega^o)^2 \mathbf{B}\mathbf{U}^o \tag{17}$$

with boundary conditions

$$\begin{aligned} & u_i^o = 0 \quad \text{on} \quad S_u \\ & T_{ji}(\mathbf{U}^o) n_j = 0 \quad \text{on} \quad S_T \\ & \phi^o = 0 \quad \text{on} \quad S_\phi \\ & D_i(\mathbf{U}^o) n_i = 0 \quad \text{on} \quad S_D \end{aligned} \tag{18}$$

where

$$\begin{aligned} T_{ji}(\mathbf{U}^o) &= \bar{c}_{jikl} u^o_{k,l} + \bar{e}_{kji} \phi^o_{,k} \\ D_i(\mathbf{U}^o) &= \bar{e}_{ikl} u^o_{k,l} - \bar{\epsilon}_{ik} \phi^o_{,k} \end{aligned} \tag{19}$$

For any four-vectors U and V we have

$$\begin{aligned} \langle \mathbf{A}\mathbf{U} ; \mathbf{V} \rangle &= \langle \mathbf{U} ; \mathbf{A}\mathbf{V} \rangle \\ &- \int_S [T_{ji}(\mathbf{U}) n_j v_i + D_i(\mathbf{U}) n_i \psi] dS \\ &+ \int_S [T_{kl}(\mathbf{V}) n_l u_k + D_k(\mathbf{V}) n_k \phi] dS \\ \langle \mathbf{B}\mathbf{U} ; \mathbf{V} \rangle &= \langle \mathbf{U} ; \mathbf{B}\mathbf{V} \rangle \end{aligned} \tag{20}$$

Relations (20) show that operators A and B are self-adjoint.

III.2 First Order Problem

For the first order problem defined by equation set (12), we solve equation $(12)_3$ for θ^* and substitute the result in the remaining equations to obtain

$$\begin{aligned} & -\bar{c}_{ijkl} u^*_{k,li} - \bar{e}_{kij} \phi^*_{,ki} = \rho(\omega^o)^2 u_j^* \\ & \quad + 2\rho\omega^o\omega^* u_j^o + i\lambda_{ij} \frac{\kappa_{kl}}{\kappa} \frac{\Omega}{\omega^o} L^2 \theta^o_{,kli} \\ & -\bar{e}_{kij} u^*_{i,jk} + \bar{\epsilon}_{ij} \phi^*_{,ij} = -ip_i \frac{\kappa_{kl}}{\kappa} \frac{\Omega}{\omega^o} L^2 \theta^o_{,kli} \\ & u_i^* = 0 \quad \text{on} \quad S_u \\ & (\bar{c}_{ijkl} u^*_{k,l} + \bar{e}_{kij} \phi^*_{,k}) n_j \\ & \quad = -i\lambda_{ij} n_j \frac{\kappa_{kl}}{\kappa} \frac{\Omega}{\omega^o} L^2 \theta^o_{,kl} \quad \text{on} \quad S_T \\ & \phi^* = 0 \quad \text{on} \quad S_\phi \\ & (\bar{e}_{ijk} u^*_{j,k} - \bar{\epsilon}_{ij} \phi^*_{,j}) n_i = ip_i n_i \frac{\kappa_{kl}}{\kappa} \frac{\Omega}{\omega^o} L^2 \theta^o_{,kl} \quad \text{on} \quad S_D \end{aligned} \tag{21}$$

In terms of the abstract vectors and operators, (21) can be written as

$$\begin{aligned} & \mathbf{A}\mathbf{U}^* = (\omega^o)^2 \mathbf{B}\mathbf{U}^* + \omega^* \mathbf{F} + \mathbf{G} \\ & u_i^* = 0 \quad \text{on} \quad S_u \\ & T_{ji}(\mathbf{U}^*) n_j = f_i \quad \text{on} \quad S_T \\ & \phi^* = 0 \quad \text{on} \quad S_\phi \\ & D_i(\mathbf{U}^*) n_i = g \quad \text{on} \quad S_D \end{aligned} \tag{22}$$

where

$$\begin{aligned} \mathbf{F} &= \{2\omega^o \rho u_j^o \ , 0\} \\ \mathbf{G} &= i \frac{\kappa_{kl}}{\kappa} \frac{\Omega}{\omega^o} L^2 \{\lambda_{ij} \theta^o_{,kli} \ , -p_i \theta^o_{,kli}\} \\ f_i &= -i\lambda_{ij} n_j \frac{\kappa_{kl}}{\kappa} \frac{\Omega}{\omega^o} L^2 \theta^o_{,kl} \\ g &= ip_i n_i \frac{\kappa_{kl}}{\kappa} \frac{\Omega}{\omega^o} L^2 \theta^o_{,kl} \end{aligned} \tag{23}$$

Our main interest is in evaluating the perturbation ω^* in the frequency. To do so, we take the inner product of both sides of equation $(22)_1$ with $\mathbf{U}^o$, and obtain

$$\begin{aligned} \langle \mathbf{U}^o ; \mathbf{A}\mathbf{U}^* \rangle &= (\omega^o)^2 \langle \mathbf{U}^o ; \mathbf{B}\mathbf{U}^* \rangle \\ &+ \omega^* \langle \mathbf{U}^o ; \mathbf{F} \rangle + \langle \mathbf{U}^o ; \mathbf{G} \rangle \end{aligned} \tag{24}$$

which, by using equation (20) becomes

$$\begin{aligned} & \langle \mathbf{A}\mathbf{U}^o ; \mathbf{U}^* \rangle - \int_S [T_{ji}(\mathbf{U}^*) n_j u_i^o + D_i(\mathbf{U}^*) n_i \phi^o] dS \\ & \quad + \int_S [T_{kl}(\mathbf{U}^o) n_l u_k^* + D_k(\mathbf{U}^o) n_k \phi^*] dS \\ & = (\omega^o)^2 \langle \mathbf{B}\mathbf{U}^o ; \mathbf{U}^* \rangle + \omega^* \langle \mathbf{U}^o ; \mathbf{F} \rangle + \langle \mathbf{U}^o ; \mathbf{G} \rangle \end{aligned} \tag{25}$$

Recalling the boundary conditions (18) of the zeroth order problem and $(22)_2$ - $(22)_5$ of the first order problem, we write equation (25) as

$$\begin{aligned} & -\int_{S_T} f_i u_i^o dS - \int_{S_D} g\phi^o dS \\ & \quad = \omega^* \langle \mathbf{U}^o ; \mathbf{F} \rangle + \langle \mathbf{U}^o ; \mathbf{G} \rangle \end{aligned} \tag{26}$$

Thus

$$\omega^* = \frac{-\langle \mathbf{U}^o ; \mathbf{G} \rangle - \int_{S_T} f_i u_i^o dS - \int_{S_D} g\phi^o dS}{\langle \mathbf{U}^o ; \mathbf{F} \rangle} \tag{27}$$

Since each term on the right-hand side of equation (27) involves solutions of the zeroth order problem, ω^* can be computed.

III.3 Frequency Shift

From (27), we obtain the shift in the frequency as

$$\frac{\Delta\omega}{\omega^o} = \frac{\epsilon\omega^*}{\omega^o} = i\Big\{ \int_V [-u_j^o \lambda_{ij} \kappa_{kl} \theta^o_{,kli} + \phi_o p_i \kappa_{kl} \theta^o_{,kli}] dV + \int_{S_T} u_i^o \lambda_{ij} n_j \kappa_{kl} \theta^o_{,kl} dS - \int_{S_D} \phi_o p_i n_i \kappa_{kl} \theta^o_{,kl} dS \Big\} \Big/ \Big\{ \alpha\Theta(\omega^o)^3 \int_V 2\rho u_j^o u_j^o dV \Big\} \tag{28}$$

in which θ^o is determined from $(12)_3$, viz,

$$\theta^o = -\frac{\lambda_{ij}}{\alpha} u^o_{i,j} + \frac{p_i}{\alpha} \phi^o_{,i} \tag{29}$$

Since all the zeroth order quantities are real, $\Delta\omega$ in (28) is a pure imaginary number, thus the piezoelectric body undergoes a nonperiodic motion. Physically, we expect that $i\Delta\omega < 0$ because of the damping effect of thermal dissipation, but this is not evident from (28). However, equation (28) does imply that $\Delta\omega$ decreases as ω^o increases.

IV. Another Approach

The above perturbation method neglects the boundary layer needed to satisfy the thermal boundary conditions. We now remedy this situation by using another approximate method to analyze the problem. In it the temperature field is determined in a way similar to that used in finding the temperature field in adiabatic thermoelasticity [27]. Mayer [17] used this method to study the thermoelastic attenuation of surface acoustic waves. This method can be characterised as a perturbation method with respect to some of the thermal coupling constants, since these couplings are usually weak.

First, we introduce a dimensionless number ϵ and write the original problem defined by the equations (2) and (3) as

$$\begin{aligned}
&- c_{ijkl} u_{k,li} - e_{kij} \phi_{,ki} + \epsilon \lambda_{ij} \theta_{,i} = \rho\omega^2 u_j \\
&- e_{kij} u_{i,jk} + \epsilon_{ij} \phi_{,ij} - \epsilon p_i \theta_{,i} = 0 \\
&i\omega \lambda_{ij} u_{i,j} - i\omega p_i \phi_{,i} + i\omega\alpha\theta = \frac{\kappa_{ij}}{\Theta} \theta_{,ij} \\
&u_i = 0 \quad \text{on} \quad S_u \\
&(c_{ijkl} S_{kl} - e_{kij} E_k - \epsilon \lambda_{ij} \theta) n_j = 0 \quad \text{on} \quad S_T \\
&\phi = 0 \quad \text{on} \quad S_\phi \\
&(e_{ijk} S_{jk} + \epsilon_{ij} E_j + \epsilon p_i \theta) n_i = 0 \quad \text{on} \quad S_D \\
&\theta = 0 \quad \text{on} \quad S_\theta \\
&- \kappa_{ij} \theta_{,j} n_i = 0 \quad \text{on} \quad S_q
\end{aligned} \tag{30}$$

We note that $\epsilon = 1$ corresponds to the original problem. Then we make the perturbation expansion

$$\begin{aligned}
\omega &= \omega^o + \epsilon\omega^* + O(\epsilon^2) \\
u_i &= u_i^o + \epsilon u_i^* + O(\epsilon^2) \\
\phi &= \phi^o + \epsilon\phi^* + O(\epsilon^2) \\
\theta &= \theta^o + \epsilon\theta^* + O(\epsilon^2)
\end{aligned} \tag{31}$$

which results in the following perturbation problems.

Zeroth order problem:

$$\begin{aligned}
&- c_{ijkl} u^o_{k,li} - e_{kij} \phi^o_{,ki} = \rho(\omega^o)^2 u_j^o \\
&- e_{kij} u^o_{i,jk} + \epsilon_{ij} \phi^o_{,ij} = 0 \\
&u_i^o = 0 \quad \text{on} \quad S_u \\
&(c_{ijkl} u^o_{k,l} + e_{kij} \phi^o_{,k}) n_j = 0 \quad \text{on} \quad S_T \\
&\phi^o = 0 \quad \text{on} \quad S_\phi \\
&(e_{ijk} u^o_{j,k} - \epsilon_{ij} \phi^o_{,j}) n_i = 0 \quad \text{on} \quad S_D
\end{aligned} \tag{32}$$

$$\begin{aligned}
&i\omega^o \lambda_{ij} u^o_{i,j} - i\omega^o p_i \phi^o_{,i} + i\omega^o \alpha\theta^o = \frac{\kappa_{ij}}{\Theta} \theta^o_{,ij} \\
&\theta^o = 0 \quad \text{on} \quad S_\theta \\
&- \kappa_{ij} \theta^o_{,j} n_i = 0 \quad \text{on} \quad S_q
\end{aligned} \tag{33}$$

First order problem:

$$\begin{aligned}
&- c_{ijkl} u^*_{k,li} - e_{kij} \phi^*_{,ki} + \lambda_{ij} \theta^o_{,i} \\
&\quad = \rho(\omega^o)^2 u_j^* + 2\rho\omega^o\omega^* u_j^o \\
&- e_{kij} u^*_{i,jk} + \epsilon_{ij} \phi^*_{,ij} - p_i \theta^o_{,i} = 0 \\
&u_i^* = 0 \quad \text{on} \quad S_u \\
&(c_{ijkl} u^*_{k,l} + e_{kij} \phi^*_{,k} - \lambda_{ij} \theta^o) n_j = 0 \quad \text{on} \quad S_T \\
&\phi^* = 0 \quad \text{on} \quad S_\phi \\
&(e_{ijk} u^*_{j,k} - \epsilon_{ij} \phi^*_{,j} + p_i \theta^o) n_i = 0 \quad \text{on} \quad S_D
\end{aligned} \tag{34}$$

We note that the zeroth order problem is a one-way coupled system: ω^o, u_j^o, and ϕ^o can be solved from equations (32) without knowing the temperature field. Then, θ^o can be solved from (33), which is complex and satisfies the thermal boundary conditions.

For the first order problem, only the mechanical problem is defined by equations (34); the thermal

part is not needed to determine ω^*. We now write (34) as

$$\begin{aligned}
&\mathbf{AU}^* = (\omega^o)^2\mathbf{BU}^* + \omega^*\mathbf{F} + \mathbf{G} \\
&u_i^* = 0 \quad \text{on} \quad S_u \\
&T_{ji}(\mathbf{U}^*)n_j = f_i \quad \text{on} \quad S_T \\
&\phi^* = 0 \quad \text{on} \quad S_\phi \\
&D_i(\mathbf{U}^*)n_i = g \quad \text{on} \quad S_D
\end{aligned} \tag{35}$$

where

$$\begin{aligned}
&\mathbf{F} = \{2\omega^o\rho u_j^o\ , 0\} \\
&\mathbf{G} = \{-\lambda_{ij}\theta^o_{,i}\ , p_i\theta^o_{,i}\} \\
&f_i = \lambda_{ij}\theta^o n_j \\
&g = -p_i\theta^o n_i\ .
\end{aligned} \tag{36}$$

We note that equation set (35) has the same structure as equation set (22), except that now isothermal material constants are used in the abstract operators **A**, **B**, $T_{ji}(\mathbf{U})$, and $D_i(\mathbf{U})$. Following the procedure of the previous section, we obtain

$$\omega^* = \frac{-\langle \mathbf{U}^o\ ; \mathbf{G}\rangle - \int_{S_T} f_i u_i^o dS - \int_{S_D} g\phi^o dS}{\langle \mathbf{U}^o\ ; \mathbf{F}\rangle}\ . \tag{37}$$

Recalling relations (36) and setting $\epsilon = 1$, we conclude that

$$\begin{aligned}
\frac{\Delta\omega}{\omega^o} = \frac{\epsilon\omega^*}{\omega^o} = \Big\{ &\int_V [u_j^o\lambda_{ij}\theta^o_{,i} - \phi_o p_i\theta^o_{,i}]dV \\
&- \int_{S_T} u_i^o\lambda_{ij}\theta^o n_j dS + \int_{S_D} \phi_o p_i\theta^o n_i dS\Big\} \\
\Big/(\omega^o)^2 &\int_V 2\rho u_j^o u_j^o dV
\end{aligned} \tag{38}$$

Noting that θ^o is complex, the frequency shift is generally a complex number, even though it is not transparent.

V. AnExample

As an example, we consider an infinite AT cut quartz plate occupying the region $0 < x_1 < b$. The plate is fixed and thermally insulated at $x_1 = 0, b$. Since our main concern is in assessing the effect of the thermal dissipation, we assume the plate to be made of an anisotropic thermoelastic material without piezoelectric effects. We consider the thickness stretch vibrations with

$$u_1 = \tilde{u}_1(x_1)e^{i\omega t}\ , \quad u_2 = u_3 = 0\ , \quad \theta = \tilde{\theta}(x_1)e^{i\omega t} \tag{39}$$

The governing equations (6) and boundary conditions (3) in this case simplify to

$$\begin{aligned}
&-c_{1111}u_{1,11} + \lambda_{11}\theta_{,1} = \rho\omega^2 u_1\ , \quad 0 < x_1 < b \\
&i\omega\lambda_{11}u_{1,1} + i\omega\alpha\theta = \frac{\kappa_{11}}{\Theta}\theta_{,11}\ , \quad 0 < x_1 < b \\
&u_1 = 0\ , \quad x_1 = 0, b \\
&\theta_{,1} = 0\ , \quad x_1 = 0, b
\end{aligned} \tag{40}$$

The results of the two perturbation methods are given below.

V.1 Perturbation with Respect to the Thermal Conductivity

The zeroth order problem is

$$\begin{aligned}
&-c_{1111}u^o_{1,11} + \lambda_{11}\theta^o_{,1} = \rho(\omega^o)^2 u_1^o\ , \quad 0 < x_1 < b \\
&i\omega^o\lambda_{11}u^o_{1,1} + i\omega^o\alpha\theta^o = 0\ , \quad 0 < x_1 < b \\
&u_1^o = 0\ , \quad x_1 = 0, b
\end{aligned} \tag{41}$$

with solution

$$\begin{aligned}
&\omega^o = \frac{n\pi}{b}\left(\frac{\bar{c}_{1111}}{\rho}\right)^{\frac{1}{2}}\ , \quad n = 1, 2, 3, \cdots \\
&u_1^o = \sin(n\pi\frac{x_1}{b}) \\
&\theta^o = -\frac{\lambda_{11}}{\alpha}\frac{n\pi}{b}\cos(n\pi\frac{x_1}{b})
\end{aligned} \tag{42}$$

where

$$\bar{c}_{1111} = c_{1111} + \frac{\lambda_{11}^2}{\alpha}\ . \tag{43}$$

Following the steps used to obtain (28), we get

$$i\frac{\Delta\omega}{\omega^o} = -\frac{\kappa_{11}\lambda_{11}^2}{2\rho\alpha^2\Theta}\left(\frac{\rho}{\bar{c}_{1111}}\right)^{\frac{3}{2}}\left(\frac{n\pi}{b}\right) \tag{44}$$

From (44) we can see that $i\Delta\omega$ is negative which indicates a damped motion. We make the following observations.

(i) The damping increases as thermal conductivity κ_{11} and thermoelastic coupling λ_{11} increase, and decreases as α which is proportional to the specific heat increases.

(ii) For a particular mode, i.e., fixed n, the damping increases as ρ increases and $\bar{c}_{1111}$ decreases.

(iii) Higher order modes with larger n have higher temperature gradients which cause enhanced heat flow and hence more damping.

For an AT cut quarts at $\Theta = 298.2\,^{\circ}K$, we have [25]

$$\begin{aligned} \bar{c}_{1111} &= 86.75 \times 10^{10} dyn/cm^2 \\ \lambda_{11} &= 13.52 \times 10^{6} dyn/cm^2\,{}^{\circ}K \\ \kappa_{11} &= 6.65 \times 10^{5} erg/cm \cdot sec\,{}^{\circ}K \\ \rho &= 2.649 g/cm^3 \\ \alpha\Theta &= 1.948 \times 10^{7} dyn/cm^2\,{}^{\circ}K \end{aligned} \tag{45}$$

which gives, for $b = 1cm$

$$i\frac{\Delta\omega}{\omega^o} = -9.21 \times 10^{-11}(n\pi) \tag{46}$$

which is a very small number.

V.2 Perturbation with Respect to Thermoelastic Coupling Constants

The zeroth order problem is

$$\begin{aligned} & -c_{1111}u^o_{1,11} = \rho(\omega^o)^2 u^o_1\,, \quad 0 < x_1 < b \\ & i\omega^o\lambda_{11}u^o_{1,1} + i\omega^o\alpha\theta^o = \frac{\kappa_{11}}{\Theta}\theta^o_{,11}\,, \quad 0 < x_1 < b \\ & u^o_1 = 0\,, \quad x_1 = 0, b \\ & \theta^o_{,1} = 0\,, \quad x_1 = 0, b \end{aligned} \tag{47}$$

with solution

$$\begin{aligned} \omega^o &= \frac{n\pi}{b}\left(\frac{c_{1111}}{\rho}\right)^{\frac{1}{2}}\,, \quad n = 1, 2, 3, \cdots \\ u^o_1 &= \sin\left(n\pi\frac{x_1}{b}\right) \\ \theta^o &= -\frac{\frac{\lambda_{11}}{\alpha}\frac{n\pi}{b}\frac{\alpha\omega^o\Theta}{\kappa_{11}}}{\left(\frac{n\pi}{b}\right)^4 + \left(\frac{\alpha\omega^o\Theta}{\kappa_{11}}\right)^2} \\ &\quad \times \left[\frac{\alpha\omega^o\Theta}{\kappa_{11}} + i\left(\frac{n\pi}{b}\right)^2\right]\cos\left(n\pi\frac{x_1}{b}\right)\,, \end{aligned} \tag{48}$$

and

$$\begin{aligned} \mathrm{Re}\left(i\frac{\Delta\omega}{\omega^o}\right) &= -\frac{\kappa_{11}\lambda_{11}^2}{2\rho\alpha^2\Theta}\left(\frac{\rho}{c_{1111}}\right)^{\frac{1}{2}}\left(\frac{n\pi}{b}\right) \\ &\quad \times \frac{1}{1 + \left(\frac{\kappa_{11}}{\alpha\Theta}\right)^2\left(\frac{n\pi}{b}\right)^2\frac{\rho}{c_{1111}}} \end{aligned} \tag{49}$$

Comparing (49) with (44), we see that formally the main difference between the two expressions is a factor of $1/[1 + (\kappa_{11}/\alpha\Theta)^2(n\pi/b)^2(\rho/c_{1111})]$ which is very close to one in our example. Hence the two approaches give nearly the same relative shift in the frequencies for our example problem. We note that in (44) and (49) the unperturbed frequency ω^o is slightly different due to the difference between c_{1111} and $\bar{c}_{1111}$. For $b = 1cm$, (49) gives

$$\begin{aligned} \mathrm{Re}\left(i\frac{\Delta\omega}{\omega^o}\right) &= -9.26 \times 10^{-11} \\ &\quad \times (n\pi)\frac{1}{1 + 3.57 \times 10^{-13}(n\pi)^2} \end{aligned} \tag{50}$$

which is indeed very close to (46). As stated earlier, $\Delta\omega$ is a complex number.

VI. Conclusions

We have used two perturbation methods to ascertain the changes in the frequencies caused by thermal dissipation in a themopiezoelectric body. In one, the solution is perturbed with respect to heat conductivity, in the other with respect to the thermopiezoelectric coupling constants. Both methods imply that the relative frequency shift is imaginary thus implying the decay of the amplitude of vibrations. In the first method adiabatic material constants are used, but in the second isothermal material constants are employed. In the simple example of the thickness stretch vibrations of a quartz plate, the two methods give close results: the imaginary relative frequency shift which represents the effect of thermal damping is of the order of 10^{-10}.

Acknowledgements

This work was supported by the U.S. Army Research Office grant DAAH 04-93-G-0214 to the University of Missouri-Rolla.

References

[1] R. D. Mindlin, On the equations of motion of piezoelectric crystals, *Problems of Continuum Mechanics*, N. I. Muskhelishvili 70th Birthday Volume, pp. 282-290, 1961.

[2] R. D. Mindlin, Equations of high frequency vibrations of thermopiezoelectric crystal plates, *Int. J. Solids Structures*, Vol. 10, pp. 625-637, 1974.

[3] W. Nowacki, A reciprocity theorem for coupled mechanical and thermoelectric fields in piezoelectric crystals, *Proceedings of Vibration Problems*, Warsaw, 1, 6, pp. 3-12, 1965.

[4] W. Nowacki, *Dynamic Problems of Thermoelasticity*, pp. 323-341, PWN-Polish Scientific Publishers, Warszawa, 1975.

[5] W. Nowacki, Some general theorems of thermopiezoelectricity, *Journal of Thermal Stresses*, 1:171-182, 1978.

[6] W. Nowacki, Foundations of linear piezoelectricity, in H. Parkus (ed.), *Electromagnetic Interactions in Elastic Solids*, Chap. 1, Wien, Springer-Verlag, 1979.

[7] W. Nowacki, Mathematical models of phenomenological piezo-electricity, *New Problems in Mechanics of Continua*, pp. 29-49, University of Waterloo Press, 1983.

[8] W. Nowacki, *Thermoelasticity* 2nd ed., pp. 528-532, Polish Scientific Publishers, Warszawa, 1986.

[9] J. P. Nowacki, Steady-state problems of thermopiezoelectricity, *Journal of Thermal Stresses*, Vol. 5, pp. 183-194, 1982.

[10] K. Majorkowska-Knap, Uniqueness theorem of linear thermopiezoelectricity I, *Bulletin of the Polish Academy of Sciences, Technical Sciences*, Vol. 35, No. 3-4, pp. 163-169, 1987.

[11] K. Majorkowska-Knap, Uniqueness theorem of linear thermopiezoelectricity II, *Bulletin of the Polish Academy of Sciences, Technical Sciences*, Vol. 35, No. 3-4, pp. 170-177, 1987.

[12] J. L. Nowinski, Theory of Thermoelasticity with Applications, Sijthoff and Noordhoff International Publishers B. V., Alphen aan den Rijn, the Netherlands, pp. 741-751, 1978.

[13] D. Iesan, On some theorems in thermopiezoelectricity, *Journal of Thermal Stresses*, 12:209-223, 1989.

[14] D. S. Chandrasekharaiah, A temperature-rate-dependent theory of thermopiezoelectricity, *Journal of Thermal Stresses*, Vol. 7, pp. 293-306, 1984.

[15] D. S. Chandrasekharaiah, A generalised linear thermoelasticity theory for piezoelectric media, *Acta Mechanica*, Vol. 71, pp. 39-49, 1988.

[16] A. K. Pal, Surface waves in a thermo piezoelectric medium of monoclinic symmetry, *Czech. J. Phys.*, Vol. B 29, pp. 1271-1281, 1979.

[17] A. P. Mayer, Thermoelastic attenuation of surface acoustic waves, *Int. J. Engng Sci.*, Vol. 28, No. 10, pp. 1073-1082, 1990.

[18] H. S. Paul and K. Renganathan Free vibrations of a pyroelectric layer of hexagonal (6mm) class, *J. Acoust. Soc. Am.*, Vol. 78 (2), pp. 395-397, 1985.

[19] H. S. Paul and G. V. Raman, Vibrations of pyroelectric plates, *J. Acoust. Soc. Am.*, Vol. 90 (4), pp. 1729-1732, 1991.

[20] H. S. Paul and G. V. Raman, Wave propagation in a hollow pyroelectric circular cylinder of crystal class 6, *Acta Mechanica*, Vol. 87, pp. 37-46, 1991.

[21] F. Ashida, T. R. Tauchert and N. Noda, Response of a piezothermoelectric plate of crystal class 6mm subject to axisymmetric heating, *Int. J. Engng Sci.*, Vol. 31, No. 3, pp. 373-384, 1993.

[22] T. R. Tauchert, Piezothermoelastic behavior of a laminated plate, *Journal of Thermal Stresses*, Vol. 15, pp. 25-37, 1992.

[23] E. Radzikowska, Thermopiezoelectricity equations of plates, *Bulletin de L'Academie Polonaise des Sciences*, Serie des Sciences Techniques, Vol. XXIX, No. 9-10, pp. 195-203, 1981.

[24] G. A. Maugin, *Continuum Mechanics of Electromagnetic Solids*, pp. 235-236, Elsevier, New York, 1988.

[25] J. Tsai and G. Hermann, Thermoelastic dissipation in high-frequency vibrations of crystal plates, *The Journal of the Acoustical Society of America*, Vol. 36, pp. 100-110, 1964.

[26] M. C. Dömecki, Recent progress in the dynamic applications of piezoelectric crystals, *Shock Vib. Dig.*, Vol. 20, pp. 3-20, 1988.

[27] J. L. Nowinski, Theory of Thermoelasticity with Applications, Sijthoff and Noordhoff International Publishers B. V., Alphen aan den Rijn, the Netherlands, p. 140, 1978.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

COMPUTER ANALYSIS OF THE CHARACTERISTICS OF HIGH-FREQUENCY QUARTZ RESONATORS

I. I. Postnikov

RD Institute of Electronic Measurement Instruments
141018, Moscow Region, Mytishschi,
Novo-Mytishschinsky prospekt, 47-3-51, Russia

Abstract

Fundamentals of the theory of vibration of quartz resonators with piezoelectric elements (PE) of rotated Y-cuts operating in thickness-shear mode and features of obtaining their mathematical model are discussed, the results of numerous calculation experiments with resonators having axially symmetric piezoelectric elements of variable thickness aimed at determination of their characteristics and parameters are presented.

Introduction

A great number of works both theoretical and experimental is dedicated to the investigation of high-frequency quartz resonators operating in thickness-shear mode, which allow to resolve successfully the problems of designing and manufacturing quartz resonators with required electrophysical parameters. The theories available to-day, however, describe, as a rule, vibrations of finite plates and give exact results for plates of a simple kind only. The majority of solutions used are based on approximation of two-dimensional equations obtained by expansion in power series of selected functions [1,2] or their representation as Legendre polynomials [3,4] both for a purely elastic case [1,3] and that with the account of piezoelectric properties of the material [2,4].

Seeking to simplify more the equations of motion obtained, describing coupled vibrations of finite plates with monoclinic symmetry, Mindlin and Spencer [5] have neglected the piezoelectric effect and made some more number of simplifications connected with the evaluation of stresses and the second derivatives of the displacement function. Equations of coupled vibrations (thickness-shear, thickness-twist, flexural) of a plate obtained as a result have been later on considered as the initial ones in quite a number of works, (e.g.[5-9]) discussing AT-cut quartz crystal units. In the case of a weak coupling of operating mode of vibration with the flexural and face-shear ones, the basic type of unwanted vibrations will be anharmonic vibrations, representing higher order components of a thickness-shear mode, for approximate determination of which the formula proposed by R.A.Sykes is widely used [10].

With the elapse of time the equations of motion obtained and their solutions have been more exactly defined, taking into account the electrode mass and also the fact that electrodes cover only a part of a plate, as a rule [6-7,11,12]. The piezoelectric effect has been neglected in these works and the same elastic properties have been supposed in parts of a plate both covered and not covered by-metal. But in the papeis [13,14] the influence of piezoelectric effect on the vibration system of a plate under investigation has been taken into account, and the frequency equation of a partially metallized rectangular piezoelectric plate obtained as a result has been used for determination of electrode influence on the frequency spectrum of AT-cut quartz resonators [14]. By using analogous equations, mechanical displacements at separate points of a plate have been calculated and in such a way the mode of vibration has been determined [7]. In the above referenced papers the limitation of a plate and electrodes has been considered in one direction only, which was insufficient for practical application, therefore the tendency to take into account these limitations in the second direction has resulted in the appearence of an approximate solution for this case also [15,16].

The influence of a limited area electrodes on overtone suppression of the thickness-shear mode for the first time observed by Bechmann [17] and Mortley[18] has been strictly considered by Mindlin and Lee [19] Shockly et.al.[20] have obtained numerous experimental confirma-

0-7803-1945-1/94 $3.00 © 1994 IEEE

tions of this phenomenon called nowadays "energy trapping".

For suppression of the nearest higher order components of the thickness-shear mode Bechmann has proposed and Mindlin [21] has substantiated theoretically an intended dimension selection, including electrode thickness. The conditions of maximum unwanted response suppression formulated for this case are known as Bechmann's criterion. It is also known, that reduction of unwanted vibrations influence on the operating ones is ensured by PE countouring, due to which the rezonator quality factor Q becomes higher.

However, the calculation methods of required dimensions of plate diameter and those of a contour and its shape are yet insufficiently developed. There exist only approximate expressions for the calculation of a frequency spectrum of doubly-convex and plane-convex PEs both without account [22] and with the account [23] of the electrode mass influence. Little developed are calculation methods of resonators having PE and electrodes in a form of a step and also those of bevelled PEs. Problems connected with piezoelectric element concavity are little investigated. Development of calculation methods for multimode resonators is also at an initial stage, for which it is necessary to ensure not only a specified resonance spacing between operating vibrations, but also their sufficient intensity with the maximum suppression of unwanted modes at the same time.

It is also necessary to note, that the majority of resonator calculation methods employed today, consider piezoelectric elements in the Cartesian rectangular co-ordinate system and use variables division just in this system, which does not allow to describe precisely displacement distribution in PE and, consequently, to determine resonator equivalent electrical parameters. For example, experimental investigations [24] have shown, that nodes of displacement of many modes have a form of a family of orthogonal ellipses and hyperbolas rather than that of direct lines as it follows from the methods of solution used (see, e.g. [22, 23]).

As it will be shown below, numerical methods of thickness-shear mode investigation of axially symmetric piezoelectric elements of variable thickness (figure 1), proposed in this paper, allow to make not only a qualitative, but also a quantitative evaluation of the influence of PE- and electrode geometry changes, angle of cut of a crystal plate changes and temperature changes on spectral and frequency-temperature characteristics of rotated Y-cut resonators and their equivalent electrical parameters.

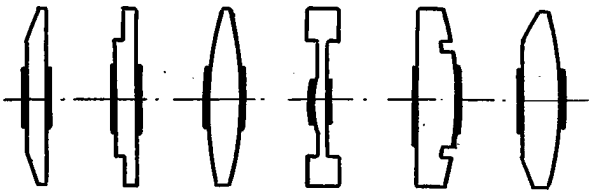


Figure 1 - Axially symmetric piezoelectric elements

1. Deduction of equations of motion

In the region of the fundamental frequency of the thickness-shear mode the exact set of three-dimensional differential equations can be replaced by an approximate system of two-dimensional equations [25]:

$$\left.\begin{aligned} &\frac{\partial M_x}{\partial x} - Q_x + \frac{\partial M_{xz}}{\partial z} + \rho_q \frac{t^3}{12}\omega^2\Psi_x = 0; \quad (1)\\ &\frac{\partial Q_x}{\partial x} + \frac{\partial Q_z}{\partial z} + \rho_q t\omega^2\eta = 0; \quad (2)\\ &\frac{\partial Q_x}{\partial z} - Q_z + \frac{\partial M_{xz}}{\partial z} + \rho_q \frac{t^3}{12}\omega^2\Psi_z = 0; \quad (3) \end{aligned}\right\} \quad (1.1)$$

where

$$M_x = D_1\frac{\partial\Psi_x}{\partial x} + D_2\frac{\partial\Psi_z}{\partial z}; \quad M_z = D_2\frac{\partial\Psi_x}{\partial x} + D_3\frac{\partial\Psi_z}{\partial z};$$

$$M_{xz} = D_5\left(\frac{\partial\Psi_z}{\partial x} + \frac{\partial\Psi_x}{\partial z}\right); \quad Q_x = D_6\left(\Psi_x + \frac{\partial\eta}{\partial x}\right);$$

$$Q_z = D_4\left(\Psi_z + \frac{\partial\eta}{\partial z}\right); \quad D_1=(C_{11}-C_{12}^2/C_{22})t^3/12;$$

$$D_2=(C_{13}-C_{12}C_{23}/C_{22})t^3/12; \quad D_5 = C_{55}t^3/12;$$

$$D_3=(C_{13}-C_{23}^2/C_{22})t^3/12; \quad D_6 = \pi^2 C_{66}t/12;$$

$$D_4 =\pi^2 t(C_{44}-C_{24}^2/C_{22}/12; \quad (1.2)$$

where C_{ij} are elastic constants
ω is the angular frequency
ρ_q quartz density
t PE thickness at the point with co-ordinates X, Z (fig.2).

Planar displacement functions $\Psi_x(X,Z)$, $\Psi_z(X,Z)$, $\eta(X,Z)$, given in (1.1) and (1.2), are connected with the displacement vector projections $U(X,Y,Z)$ by the following relations [25-26]:

$$U_x(X,y,Z) = y\Psi_x(X,Z); \quad (1.3)$$

$$U_z(x,y,z) = y\Psi_z(x,z); \qquad (1.4)$$

$$U_y(x,y,z) = \eta(x,z). \qquad (1.5)$$

We transform equation (2) of the set (1.1) with the account of (1.2) we to the following form [27,28]:

$$\frac{\partial}{\partial x}\left[D_6\left(\Psi_x + \frac{\partial \eta}{\partial x}\right)\right] + \frac{\partial}{\partial z}\left[D_4\left(\Psi_z + \frac{\partial \eta}{\partial z}\right)\right] + \rho_q t\omega^2\eta = 0. \qquad (1.6)$$

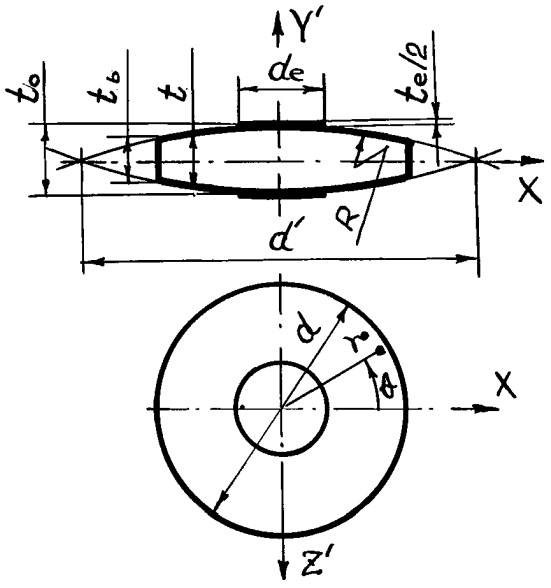


Figure 2 - Doubley-convex lens-type piezoelectric element and co-ordinate systems

The analysis of coupling of the thickness-shear mode with the flexural mode, made in [29,30] has shown, that with PE diameters, corresponding to gently sloping part of frequency branches, the flexural component of displacement is very small compared with the shear component, and one can write

$$\partial\eta/\partial x \ll \Psi_x; \qquad \partial\eta/\partial z \ll \Psi_z. \qquad (1.7)$$

The equation (1.6) will then take the form

$$\partial(D_6\Psi_x)/\partial x + \partial(D_4\Psi_z)/\partial z + \rho_q t\omega^2\eta = 0. \qquad (1.8)$$

By substituting the value of η from (1.8) in eq.(1) and (3) of equation set (1.1), we obtain:

$$\left.\begin{aligned} &\frac{\partial}{\partial x}\left(D_1\frac{\partial \Psi_x}{\partial x} + D_2\frac{\partial \Psi_z}{\partial z}\right) + \frac{\partial}{\partial z}D_5\left(\frac{\partial \Psi_z}{\partial x} + \frac{\partial \Psi_x}{\partial z}\right) + \rho_q\frac{t^3}{12}\omega^2\Psi_x = 0, \qquad (1)\\ &\frac{\partial}{\partial x}D_5\left(\frac{\partial \Psi_z}{\partial x} + \frac{\partial \Psi_x}{\partial z}\right) + \frac{\partial}{\partial z}\left(D_2\frac{\partial \Psi_x}{\partial x} + D_3\frac{\partial \Psi_z}{\partial z}\right) - D_4\Psi_z + \frac{D_4}{\rho_q t\omega^2}\left(\frac{\partial^2}{\partial z^2}D_4\Psi_z + \frac{\partial^2}{\partial x\partial z}D_6\Psi_x\right) + \rho_q\frac{t^3}{12}\omega^2\Psi_z = 0, \qquad (2) \end{aligned}\right\} \qquad (1.9)$$

With an optional form of a limiting profile $t(x,z)$, the main part of vibration energy concentates in a certain zone, where PE thickness is maximum. By supposing $t(x,z)=t_o$ and $\partial t/\partial x=\partial t/\partial z=0$, we shall obtain the set of equation system from (1.9), which describes vibrations of an infinitely thin plate having a constant thickness:

$$D_6\Psi_x - \rho_q\omega_1^2 t_o^3\Psi_x/12 = 0, \qquad (1.10)$$

$$D_4\Psi_z - \rho_q\omega_2^2 t_o^3\Psi_z/12 = 0. \qquad (1.11)$$

From (1.10) and (1.11) we obtain:

$$\omega_1 = (\pi/t_o)(C_{66}/\rho_q)^{1/2}; \qquad (1.12)$$

$$\omega_2 = (\pi/t_o)(C_{44}/\rho_q)^{1/2}. \qquad (1.13)$$

The frequencies ω_1 and ω_2 are inherent to the thickness-shear mode in xoy- and zoy-planes, respectively. A substantial diffe- rence in thise frequencies and relatively small frequency difference of an infinite plate and trapped anharmonics of shear vibrations in both planes shows, that while investigating spectral characteristics of contoured plates it is necessary to consider separately two frequency regions: those proximate to ω_1 and ω_2. For the first frequency region from eq.(2) of the set (1.9) and by negleting all the terms containing derivatives of the function Ψ_z with respect to arguments x and z (because they are small compared with $D_4\Psi_z - \rho_q t^3\omega^2\Psi_z/12$). We shall find the function Ψ_z expression and substitute it in eq.(1) of the set (1.9). The expression thus obtaned contains the fourth-order derivatives of the displacement function Ψ_x with respect to arguments X and Z, which can be neglected, because they are small compared to the terms containing second-order derivatives of Ψ_x [27]

Consequently, approximate equations of motion for the thickness-shear mode can be obtained from (1.9) by equating with zero the terms containing functions Ψ_z in eq. (1) and Ψ_x in eq.(2) of the set (1.9).

We obtain the following as a result:

$$\frac{\partial}{\partial x}\left(D_1\frac{\partial \Psi_x}{\partial x}\right) + \frac{\partial}{\partial z}\left(D_5\frac{\partial \Psi_x}{\partial z}\right) + \frac{D_6}{\rho_q t\omega_1^2}\frac{\partial^2}{\partial x^2}(D_6\Psi_x) + (\rho_q t^3\omega^2/12 - D_6)\Psi_x = 0 , \quad (1.14)$$

$$\frac{\partial}{\partial z}\left(D_5\frac{\partial \Psi_z}{\partial x}\right) + \frac{\partial}{\partial z}\left(D_3\frac{\partial \Psi_z}{\partial z}\right) + \frac{D_4}{\rho_q t\omega_2^2}\frac{\partial^2}{\partial z^2}(D_4\Psi_z) + (\rho_q t^3\omega^2/12 - D_4)\Psi_z = 0 , \quad (1.15)$$

The equations (1.14) and (1.15) are valid over the region of the fundamental frequency of the thickness- shear mode and represent equations of motion for the thickness-shear and thickness-twist vibrations of piezoelectric elements having an optional form.

2. Specifying problem

The majority of piezoelectric used in the mass production of high-frequency crystal units have not a rectangular, but a circular shape. Several possible configurations of such piezoelectric elements are shown in Figure 1. For piezoelectric elements with axial symmetry it is more convenient to consider the equation of motion (1.14), describing thickness-shear vibrations in cylindrical coordinates r, ϑ, y (Fig. 2), which are connected with the Cartesian coordinates by the following transformation formulae:

$$r=(x^2 + z^2)^{0,5}; \qquad x=r\cdot\cos\vartheta;$$

$$tg\vartheta=z/x; \qquad z=r\cdot\sin\vartheta. \quad (2.1)$$

The equation (1.14) with the account of (2.1) in cylindrical coordinates will then take the form:

$$(A_1+A_2\cos2\vartheta)\partial^2\Psi/\partial r^2 + (A_3+A_4\cos2\vartheta)\partial\Psi/\partial r + A_5\sin2\vartheta\,\partial^2\Psi/\partial r\partial\vartheta + A_5\sin2\vartheta\,\partial\Psi/\partial\vartheta \quad (2.2)$$
$$+(A_7+A_8\cos2\vartheta)\partial^2\Psi/\partial\vartheta^2 + (A_9+A_{10}\cos2\vartheta)\Psi = 0,$$

where $\Psi = \Psi_x(r,\vartheta)$.

The variable coefficients Aj in (2.2) are defined as follows:

$$\left.\begin{aligned} A_1 &= g_5 + g_6; \quad A_2 = g_1 - g_6;\\ A_3 &= A_1/r + \partial g_1/\partial r + 2g_6\partial D_6/\partial r;\\ A_4 &= -A_2/r + \partial g_1/\partial r + 2g_6\partial D_6/\partial r;\\ A_5 &= -2A_2/r^2;\\ A_6 &= 2A_2/r^2-(\partial g_1/\partial r+2g_6\partial D_6/\partial r)/r;\\ A_7 &= A_1/r^2; \quad A_8 = -A_2/r^2; \end{aligned}\right\} \quad (2.3)$$

$$\left.\begin{aligned} A_9 &= g_6(1/r\cdot\partial D_6/\partial r - \partial^2 D_6/\partial r^2) + t^3\rho_q\omega^2/6 - 2D_6;\\ A_{10} &= g_6(\partial^2 D_6/\partial r^2 - 1/r\cdot\partial D_6/\partial r);\\ g_6 &= D_6^2/\rho_q t\omega^2; \quad g_1 = D_1 - D_5;\\ g_5 &= D_1 + D_5. \end{aligned}\right\}$$

We shall consider PEs, at which a vibration localisation in their central part is ensured at the expense of a variable profile and also PEs fixed at the periphery, the boundary conditions for which will be those, when the planar displacement function $\Psi(r,\vartheta)$ goese to zero at the contour limiting PE in the xoz plane:

$$\Psi(r,\vartheta)\Big|_{r=d/2} = 0 , \quad (2.4)$$

where d is the PE diameter (Fig.2).

The solution of the boundary-value problem (2.2), (2.4) will make it possible to define eigenvalues and eigenfunctions of axially symmetric piezoelectric elements with variable thickness in the region of the fundamental frequency of the thickness-shear vibrations.

3. Recurrence relations for equations of motion

The analysis of the known PE vibration topogrammes [24,31] and also the results of numerical studies of piezoelectric elements, operating in thickness-shear mode and having displacement antinodes in the centre [32-34] confirm, that a good approximation to the solution of the equation set (2.2) is the representation of displacement function as a trigonometric series:

$$\Psi(r,\vartheta) = P_o(r) + \sum_{m=1}^{M}\Big(P_m(r)\cos(m\vartheta) + Q_m(r)\sin(m\vartheta)\Big), \; M\to\infty \quad (3.1)$$

While substituting (3.1) in (2.2) the initial set of differential equations breaks down into four independent sets with respect to the expansion coefficients (P,Q) with even and odd indices, which [9] describe all possible modes of thickness-shear vibrations. All the four sets of differential equations can be conveniently represented as a generalized recurrence relation (strokes mean differentiating over to r):

$$2[A_1F''_m + A_3F'_m + (A_9 - mA_7)F_m] + A_2F''_{m+2} + [A_4-(m+2)A_5]F'_{m+2} + [A_{10}-(m+2)A_6-(m+2)^2A_8]F_{m+2}$$

$$+A_2F''_{m-2}+ [A_4+(m-2)A_5]F'_{m-2}$$

$$+ [A_{10}+(m-2)A_6-(m-2)^2A_8]F_{m-2}$$

$$+æ\{A_2F''_{2-m}+ [A_4+(m-2)A_5]F'_{2-m}$$

$$+ [A_{10}+(m-2)A_6-(m-2)^2A_8]F_{2-m}\}=0. \quad (3.2)$$

For obtaining a set of differential equations of motion required we make substitution of necessary designations of expansion coefficients (P or Q), corresponding to the indices m and the value of multiplear æ for F in (3.2):

1st set (P_m)- m=0, 2, 4, ... } æ=1, with m⩽2;

2nd set (P_m)- m=1, 3, 5, ... } æ=0, with m>2;

3rd set (Q_m)- m=1, 3, 5, ... } æ=-1, with m=1;

4th set (Q_m)- m=2, 4, 6, ... } æ=0, with m>1;

The first set of differential equations describes vibrations having displacement antinodes in PE centre, i.e. having odd indices along the X and Z- axes (111, 113, 131, 115, 151, ...); the second one -even indices along the X-axis and those odd along the Z-axis (121, 123, 141, 125, 143, ...); the third one - even indices along the Z-axis and those odd along the X-axis (112, 114, 132, 116, 134, ...) and, finally, the fourth set - even indices along both axes (122, 124, 142, 126, 144, ...).

Let us write the set (3.2) in a matrix form:

$$\sum_{n=1}^{M} [a_{mn}F_{2n-k} + b_{mn}F'_{2n-k} + d_{mn}F''_{2n-k}]=0, \quad m=1, 2, 3, \ldots M. \quad (3.3)$$

$$k = \begin{cases} 2 - \text{for the 1st set of equations;} \\ 1 - \text{for the 2nd and 4th set of equations;} \\ 0 - \text{for the 3rd equation system;} \end{cases}$$

where M is the number of terms of the (3.1) series;

a_{ij}, b_{iji}, d_{ij} are elements of three-diagonal matrices of MxM dimension;

F''_j, F'_j, F_j are vectors of M-dimension.

Writing matrix elements a_{ij}, b_{ij}, d_{ij} by way of coefficients Aj is described in detail in monograph [40] with the account of dimensionless designations:

$$\Omega=\omega/\omega_o; \quad T=t/t_o; \quad r_m=d/2t_o; \quad r_*=2r/d;$$

$$R^*=R^o/t_o; \quad (3.4)$$

where $\omega_o = (\pi/t_o)(C_{66}/\rho_k)^{0.5}$ is the frequency of thickness-shear vibrations of an infinite plate with the thickness t_o;

r is the current radius of a piezoelectric element;

$R^o=R\cdot k_f$ is the reduced radius for sphere of a PE ($k_f=1$ for a biconvex PE, $k_f=2$ for a planoconvex PE).

The boundary conditions for the modes vibration discussed have the form:

$$F_{2m-k}\Big|_{r=1} = 0. \quad m=1, 2, 3, \ldots M \quad (3.5)$$

General solution of a set with M second-order differential equations (3.3) having variable coefficients can be written in the form of a sum of partial solutions with 2M indefinite coefficients:

$$F_{2m-k} = \sum_{j=1}^{2M} C_jY_{mj}, \quad m=1.2.3.\ldots M$$

where C is a vector of constant coefficients;

Y is a matrix of 2M partial solutions.

It is known, that partial solutions Ymj are linearly independent then and only then, when Vronsky determinant $\Delta(Y_{11}, Y_{M,2M})$ differs from zero [35]. Since partial solutions of differential equation set (3.3) are completely defined by specifying initial values in 2M points, the unknown solution of (3.3) can be represented in the following way.

We write the function and the first derivative values at the point r=0 in accordance with the chosen type of vibration ($\Psi'|_{r=0}=0$; $\Psi|_{r=0}=\text{const}$ for the first set and $\Psi|_{r=0}=0$, $\Psi'|_{r=0}=\text{const}$ for the remaining of the sets).

For the first system:

$$\left.\begin{aligned} P_{2m-k}(0) &= \sum_{j=1}^{2M} C_jY_{mj} = g_m; \\ P'_{2m-k}(0) &= \sum_{j=1}^{2M} C_jY'_{mj} = 0; \end{aligned}\right\} \quad (3.6)$$

$$m=1.2.3.\ldots M$$

For the 2nd, 3rd, 4th sets:

$$\left.\begin{aligned} F_{2m-k}(0) &= \sum_{j=1}^{2M} C_jY_{mj} = 0; \\ F'_{2m-k}(0) &= \sum_{j=1}^{2M} C_jY'_{mj} = g_m; \end{aligned}\right\} \quad (3.7)$$

$$m=1, 2, 3, \ldots M$$

We shall write the initial values of coefficients of algebraic sets of equations (3.6), (3.7) as a unit matrix

$$(Y_{11}, Y_{M,2M}) = E, \quad (3.8)$$

then the Vronsky determinant

$$\Delta(Y_{11}, Y_{M,2M})=1$$

and partial solutions will be linearly independent [35].

From (2.25), (2.26) and (2.27) it

follows that:

$$C_{M+1} = C_{M+2} = \ldots = C_{M+M} = 0.$$

Thus, the general solution of a linear set of M second-order differential equations (3.3) consists of a sum M of linearly independent solutions with M of indefinite coefficients

$$F_{2m-k} = \sum_{j=1}^{M} C_j Y_{mj}, \quad m=1,2,3,\ldots M \qquad (3.9)$$

The values of C_j -coefficients are defined after the boundary conditions (3.5) from the set of linear algebraic equations (SLAE) are fulfilled

$$F_{2m-k} = \sum_{j=1}^{M} C_j Y_{mj} = 0, \quad m=1,2,3,\ldots M \qquad (3.10)$$

An uniform SLAE (3.10) has a nontrivial solution only in the case, when its determinant is equal to zero

$$\Delta Y = \left| \begin{array}{cccc} Y_{11} & Y_{12} & Y_{13} \ldots & Y_{1m} \\ Y_{21} & Y_{22} & Y_{23} \ldots & Y_{2m} \\ Y_{31} & Y_{32} & Y_{33} \ldots & Y_{3m} \\ \ldots & \ldots & \ldots & \\ Y_{m1} & Y_{m2} & Y_{m3} & Y_{mm} \end{array} \right|_{r=1} = 0,$$

which is a condition of eigenvalues (Ω_{1mp}) determination.

4. Numerical solution of a problem

For solution of the 2nd- order set of differential equations with variable coefficients (3.3), having a complicated dependence on radius it is reasonable to use numerical methods. For this purpose we transform (3.3) and (3.4) to a three-point difference circuit. Then the edge problem will be written in difference relations as follows:

$$[\alpha]_i[F]_{i-1} - [\gamma]_i[F]_i + [\beta]_i[F]_{i+1} = 0, \quad [F]_N = 0,$$

$[P]_o = [P]_1$ - for the 1st set of equations;

$[F]_o = -[F]_1$ - for the 2nd-4th sets of equations.

$i=1,2,3,\ldots N.$ (4.1)

Where, α, β, γ are matrices (of the order MxM) of the coefficients at point i;

P, F -eigenfunction vectors at points i-1, i, i+1(Fig.3).

Matrice elements a , β, γ at point i are connected with matrice elements a,b,d, (3.3) at the same points by the following relations:

$$\alpha_{mn} = d_{mn} - 0{,}5pb_{mn};$$

$$\beta_{mn} = d_{mn} + 0{,}5pb_{mn}; \quad \gamma_{mn} = 2d_{mn} - p^2 a_{mn};$$

where p=2(2N-1) is a subdivision step of integration range into N sections.

For resolving a homogenons SLAE (4.1) it is necessary to equate to zero its determinant, which corresponds to defined

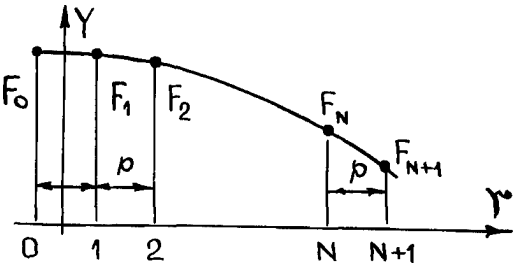


Figure 3 - Splitting of integrations interval

eigenvalues (Ω). However, the calculation of eigenfunctions [F] with the required accuracy can ensure solution of large sets of equations M.(N-1). It is more convenient, therefore to go over to the method of "shooting" [36], which is usually applied to small SLAE. Let us consider it applicably to the solution of the present problem.

We substitute (3.9) in (4.1)

$$[Y]_{i+1} = [\beta]_i^{-1}\{[\gamma]_i[Y]_i - [\alpha]_i[Y]_{i-1}\}, \qquad (4.2)$$

$[Y]_1 = [Y]_o$ - for the 1st group of vibrations;

$[Y]_1 = -[Y]_o$ - for the remaining groups of vibrations.

$$[Y]_o = E \qquad (4.3)$$

$$\Delta Y = |Y|_N = 0. \qquad (4.4)$$

The edge problem can be brought to the Cauchy problem [37]. By specifying boundary conditions with r=0 (4.3) we define consecutively the values of $[Y]_j$ at all points of the integration range over (4.2) and check fulfillment of the boundary condition at $r^*=1$ (4.4). The eigenvalues of Ω are with this a parameter which is defined by a shooting method.

By substituting the matrix coefficients $[Y]_N$ in (3.10) and making $C_1=1$, we define remaining coefficients C_j, from (3.10) and then eigenfunctions F_{2m-k} from (3.9) in any node of a net. In such a way by employing this algorithm one can solve the set M of algebraic equations in any node of net, i.e. the order of SLAE which can be solved, is by N times less than in the case of the finite difference method.

5. Stable algorithm of numerical problem solution

The algorithm considered above is applicable only for PE modilling, having small surface curvature. For resonators, using "trapped energy" principle (PEs with great curvature, etc.), the localization of maximum displacement amplitudes towards the PE centre is characteristic with their abrupt attenuation towards the periphery. Therefore the initial equations should

have a solution typical for an exponential form, and at a large integra- tion segment the increasing exponential component becomes soon a dominating one [38]. Consequenty, if one integrates over (4.2) from PE centre towards the periphery proceeding from independent initial data sets (4.3), then after passing a certain segment it can turn out, that the integration process of each of M solutions will result in the same solution. Moreover, if the exponential component increases too quickly, then the solution can overflow the limits of the capability net of a computer. Analogous arguments are applicable to integration from the periphery towards the centre. In this case it is necessery to solve the set of equations discussed by using a run method [36] in accordance with which the solution is represented in the following form:

$$[Y]_{i+1} = [\xi]_{i+1}[Y]_i + [\tau]_{i+1}. \qquad (5.1)$$

Substitute (5.1) in (4.2)

$$[\alpha]_i - [\gamma]_{i-1} - \{[\gamma]_i - [\beta]_i[\xi]_{i+1}\}[Y]_i + [\beta]_i[\tau]_{i+1} = 0. \qquad (5.2)$$

By expressing [Y] from (5.2) and substituting in (4.2) we shall obtain

$$\{[\alpha]_i - [\gamma]_i[\xi]_i + [\beta]_i[\xi]_{i+1}[\xi]\}[Y]_{i-1} - [\gamma]_i[\tau]_i + [\beta]_i[\xi]_{i+1}[\tau]_i + [\beta]_i[\tau]_{i+1} = 0. \qquad (5.3)$$

The set of equations (5.3) can be fulfilled for every $[Y]_{i-1}$, when

$$[\alpha]_i + \{[\beta]_i[\xi]_{i+1} - [\gamma]_i\}[\xi]_i = 0,$$

$$[\beta]_i[\tau]_{i+1} + \{[\beta]_i[\xi]_{i+1} - [\gamma]_i\}[\tau]_i = 0. \qquad (5.4)$$

Hence we obtain reccurrence formulae for defining matrix coefficients $[\xi]_i$, $[\tau]_i$:

$$[\chi]=[\gamma]_i - [\beta]_i[\xi]_{i+1};$$

$$[\xi]_i=[\alpha]_i/[\chi]_i; \quad [\tau]_i =[\beta]_i[\tau]_{i+1}/[\chi]. \qquad (5.5)$$

For calculation of matrix coefficients, $[\xi]_i$, $[\tau]_i$ it is necessery to know the values of $[\xi]_{i+1}$ and $[\tau]_{i+1}$, which can be conveniently set in the following way:

$$[\xi]_N = 0, \qquad [\tau]_N = 0 \qquad (5.6)$$

Then

$$[Y]_N\, C=\{[\xi]_N[Y]_{N-1}+[\tau]_N\}C=0, \qquad (5.7)$$

is feasible with any $[Y]_{N-1}$ matrix and, consequently, the first edge condition (4.1) is also feasible

$$[Y]_N \cdot C=0. \qquad (5.8)$$

In the works [39,40] it is shown, that the conditions of stable solution of run method will be fulfilled only then, when the value of the coefficient is $a<0$, which corresponds to $\Omega H<1$.

6. Calculation of resonance frequencies and displacement distributions of a piezoelectric element

By employing the above algoritm a computer programme has been developed, the complete text of which is given in the work [40] enabling to calculate a complete spetrum of thickness-shear vibrations of axially symmetric piezoelectric elements with variable thickness and displacement distribuions in them for any resonance vibration. Figure 4 depicts the calculated dependence of fundamental frequency on the radius of sphere and PE diameter; whereas Figure 5 depicts displacement distribution at the surface of a lens-type PE (LPE) for fundamental and several anharmonic vibrations. The calculated and experimental [22] frequencies of several resonance vibrations of a biconvex LPE (R=520mm, t_o=0,335мм) given in Table 1 confirm their good agreement.

Numerous results of a quantitative evaluation obtained by using the method of PE eigenvalues and PE eigenfrequencies calculation described in [40,41] confirm also a good agreement of calculated and experimental values. For quartz crystal resonators with LPEs of AT-cut simple generalized analytical expressions have been obtained for calculation of a complete frequency spectrum of thickness-shear vibrations (up to anharmonic n51) [42].

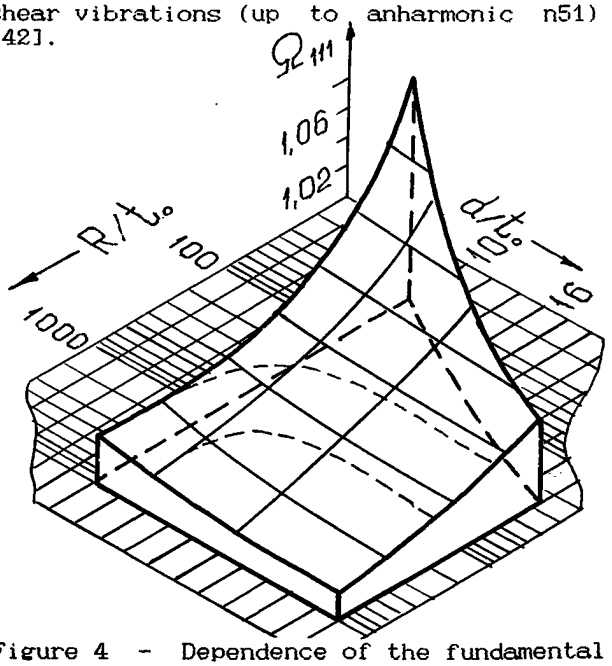


Figure 4 - Dependence of the fundamental frequency of a lens-type PE on its radius of sphere and diameter

Figure 6 shows the influence of electrode thickness on the displacements distribution in a piezoelectric element. The electrode thickness (Te^*) is evaluated in equivalent crystal PE thickness. In the case of disk-shaped PE without electrodes ($Te^*=0$) displacements of PE surface particles for the funda- mental and anharmonic vibrations are dispersed within all the PE volume. With the increase of driving electrode thickness the fundamental frequency 111 is localizes under the electrode first, then anharmonic vibration 115. Ferther electrode thickness vibration 113 follows and, finaly,

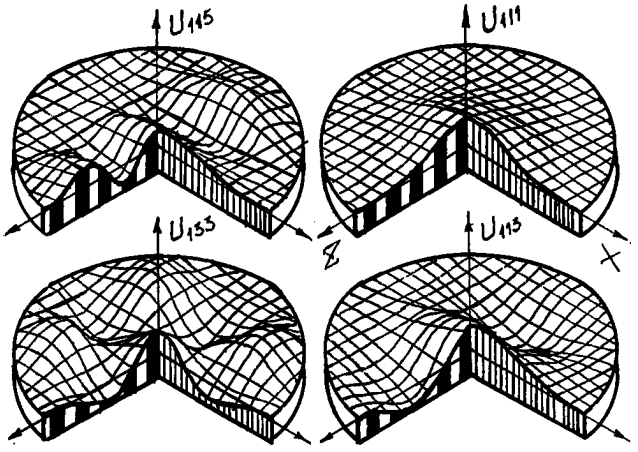


Figure 5 - Calculated displacement distributionsat the surface of a lens -type PE for a group of resonance vibrations

Table 1

Mode index	Measured values [22]		Calculated values
	f_{1mp}, kHz	f_{1mp}/f_{111}	$\Omega_{1mp}/\Omega_{111}$
111	5000	1.0	1.0
112	5084	1.0168	1.0174
121	5110	1.0220	1.0219
113	5171	1.0342	1.0344
131	5227	1.0454	1.0439
114	5261	1.0522	1.0517

increase will change (localize) the vibration energy distribution only insignificantly and with $T^*=0,03$ olready all the three modes are lumped in underelectrode area. With the electrode thickness of $Te^*=0,02$ ("thick" electrode) the intensity of all three modes of vibration will have the same order. Is it possible to disperse vibrations with "thick" electrodes? Let us consider concavity influence on the displacement distribution in piezoelectric elements.

Figure 7 represents displacement distributions at the surface of a PE with different radii of sphere for a concave PE (for Δ=0.03 - R/to=333333; 0.05-200000; 0.07-142857; 0.09-111111). We shall consider the silver electrode thickness of 0.2 μm to be permissible, then the

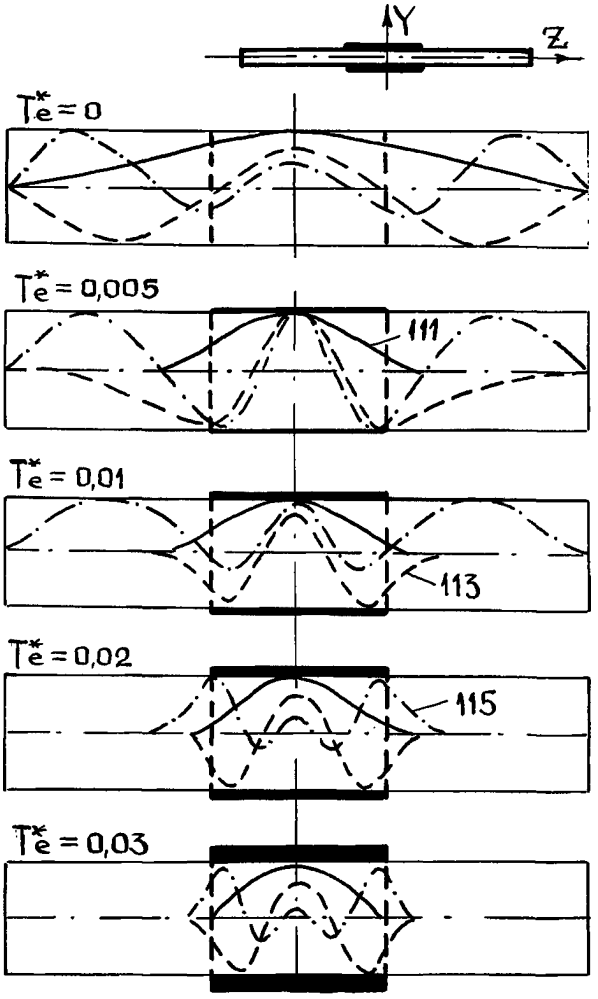


Figure 6 - Influence of electrode thickness (mass) on localization of rezonance vibrations in a disk-type PE d/to=120; de/d=0.3.

thickness (t) of a crystal plate in PE centre (Figure 7) will be equal to t=46 μm which corresponds to vibrations in the region of 35 MHz. With this electrode thickness in the case of a disk-shaped PE anharmonic vibrations 113, 131 are "trapped" and will have an amplitude of vibration comparable with the amplitude of the fundamental mode [40]. With the change of the radius of sphere of a concave PE redistribution of displacements both for the main and anharmonic vibrations occurs. The curves of displacement distribution dependence $U_{111}(X)$, $U_{111}(Z)$ begin to move from PE centre (curve 2) with the decrease of radius of sphere (increase of Δ) which results in inductance L_{111} desrease [40], after that the maximum L_{111} moves from PE

centre to the Z-axis and a deep of the displacement distribution function is formed (curve 3). At the extent of Δ increase the radius corresponding to the minimum displacement amplitude along the Z-axis increases the displacement amplitude in PE centre decreases (curves 4,5) and with a further Δ increase vibrations are completely forced out from the underelectrode area [40].

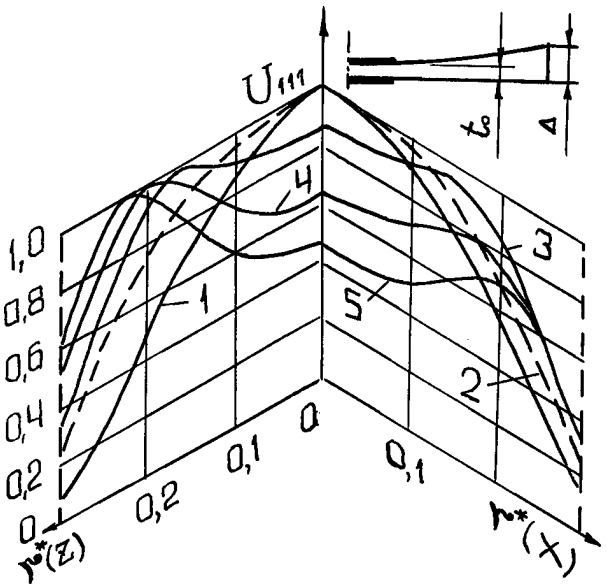


Figure 7 - Influence of PE concavity on displacement distributions at PE surface for the fundamental frequency (d/to=200; de/d=0.3; te*=0.035): 1 curve - Δ/to=0; 2-0.03; 3-0.05; 4-0.07; 5-0.09.

It is important to note that if at the initial stage of Δ change, the inductance of the fundamental mode was increased compared with the inductance of a disk-shaped PE having the same electrode, then after passing the minimum value it begins to increase and, at the extent of forcing out vibrations from the underelectrode area it tends to infinity. Analogous conversions undergo also anharmonic vibrations for exaple those of anharmonic 113, with Δ=0,05 which will be maximally attenuated . Special interest presents PE with Δ=0,07 at which not only simultaneous maximum attenuation of anharmonics 113 and 131 is obtained, but the fundamental mode will have a minimum inductance (figure 8). In the works [32-34] generalized analytical expressions for the calculation of resonance vibration distributions vat n11, n13, n31, n33, n51 of quartz resonators with LPE of AT-cut are given.

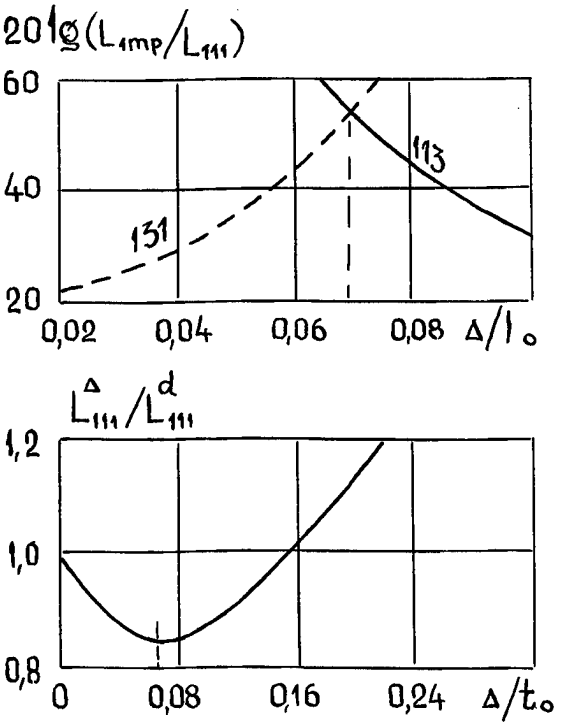


Figure 8 - Influence of PE concavity (d/to=200; de/d=0.3; te*=0.035) on the motional inductance at fundamental frequency and anharmonic vobrations

7. Calculation of equivalent electrical motional parameters of quartz resonators

Equivalent electrical circuit of a quartz resonator can be represented as shown in Figure 9. For the calculation of its equivalent motional parameters L and C it is convenient to use the following relations:

$$L_{1mp} = k_o \frac{F_v^{1mp}}{(F_s^{1mp})^2}; \quad C_{1mp} = \frac{(F_s^{1mp})^2}{k_o F_v^{1mp} \omega_{1mp}^2}; \tag{7.1}$$

$$F_s = \sum_{i=1}^{K} \int_{S_e^i} \left\{ P_o(r) + \sum_{m=1}^{M} \left[P_m(r)\cos m\vartheta + Q_m(r)\sin m\vartheta \right] \right\} dS, \tag{7.2}$$

$$F_v = \int_0^1 \left\{ 2P_o^2(r) = \sum_{m=1}^{M} \left[P_m^2(r) + Q_m^2(r) \right] \right\} H^3 r dr. \tag{7.3}$$

where K_o is coefficient depending on the effective electrical constant, crystal density and piezoelectric element dimensions;

K is the number of elementary land areas into which the area of a composite electrode is split.

The coefficient K_o in (7.1) is constant for all the modes, at the same time the function F_v^{1mp} depends on displacement distributions and for a specified vibration $1mp_{1mp}$ it is a finite value. The function F_s , in its turn, depends not only on the type of vibration, but also on geometrical dimensions and electrode configuration and can take for anharmonic vibrations even zero values. Just this function defines a qualitative character of equivalent motional inductances and capacitances of resonance vibrations and also, as it will be shown below, the intensity of vibration. Subroutines of computer calculation of equivalent motional inductances and capacitances of resonance vibrations of contoured axially symmetric PE with variable electrode configurations are presented in the book [40], where a good concidence of calculated and experimental results obtained on these parameters is also noted.

Figure 10 depicts calculated values of inductance variation for the 1st group of vibrations of biconvex LPE: $R/h_o=370$; $d/h_o=30$, illustrating the character of anharmonic vibration intensity vs electrode diameter. As it follows from Fig.10, with $d_e/d=0.27$ not only anharmonic vibrations 113,131 are maximally simultaneously attenuated (interception point of curves L_{113},L_{131}) but remaining anharmonics will have a sufficient attenuation. Just this is a condition for obtaining single frequency resonators.

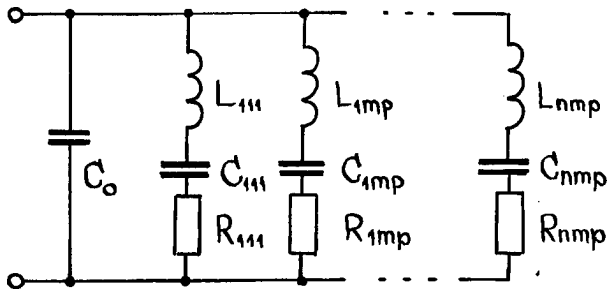


Figure 9 - Equivalent electrical circuit of a quartz resonator

However, a simultaneous maximum attenuation of anharmonic vibrations 113, 131 has a definite limit (see Fig.10). For improvement of resonator spectral characteristic it is necessary to use electrodes of a complicated configuration. Let us show whet by using a composite axially symmetric electrode having a circle and a ring configuration (Fig.11), one can obtain considerable attenuation of these unwanted responses with retaining

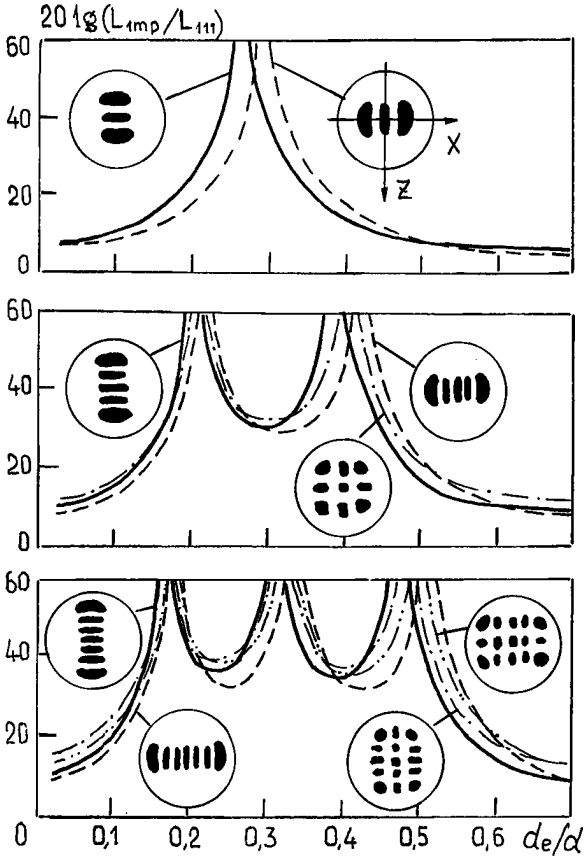


Figure 10 - Dependence of anharmonic vibration intensity attenuation on electrode diameter for LPE (R/to=370; d/to=30)

sufficient intensity of the operating mode of vibration [43]. Due to the fact, that the ring PE surface between diameters d_1 and d_2 (Fig.11) does not take part in inducing piezoelectric charge at an active electrode, it is possible by changing diameters of such an electrode to obtain simultaneous zero charge for both vibrations. For fixed $d_2/d=0.3$ values with the change of d_1/d - values, represented in Fig.11, the dependencies A_{113}, A_{131} [$A_{1mp}=20\lg(L_{1mp}/L_{111})$] begin to shift towards each other, and with the coincidence of their extremum values (for example $d_1/d=0.21$ and $d_3/d=0.348$) a maximum simultaneous attenuation of these anharmonic vibrations is observed, which is greater than in the case of optimum diameter of a circular electrode (Fig.11). It is necessary to note that with this eletrode shape the main vibration is only insignificantly attenuted (see Fig.11 - dashed line).

Especially useful are calculation methods of equivalent motional inductance when evaluating composite electrode configurations for multimode resonators, when with simultaneous unwanted rezonances attenuation of it is necessary to ensure a sufficient intensity of two or more of working resonances.

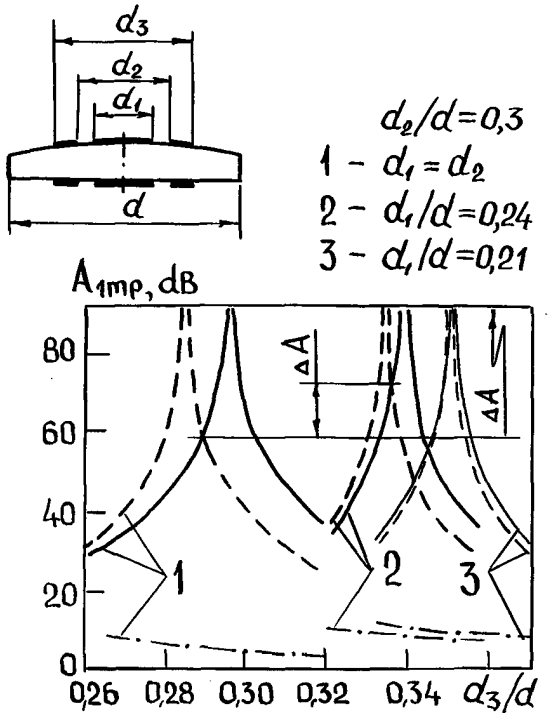


Figure 11 -Dependence of anharmonic vibrations attenuation on geometric dimensions of a composite electrode

Especially useful are the calculation methods of equivalent motional indactance developed for evaluation of composite electrode configurations for multimode resonators, when with simultaneous unwanted resonances attenuation it is necessary to ensure a sufficient intensity of two or more of working resonances.

The above calculation procedure of equivalent motional capacitance enabled us to reveal the influence of PE geometrical dimensions on capacitance relation of resonator (m) KP ($m_{nmp}=C_{nmp}/C_o$) [44-46]. ($m=1/r$, where r is a capacitance ratio). Figure 12 shows the dependence of m_{111} on r_e for LPEs with various radii of sphere, and Figure 13 shows the dependence character of r_e on capacitance ralatios at with anharmonic vibrations 113.131. Theoretical studies carried out, enabled us not only to describle more exactly the character of capacitance ratio dependencies, but also to obtain for LPEs their generalized dependencies on r_e both for harmonic and anharmonic vibrations.

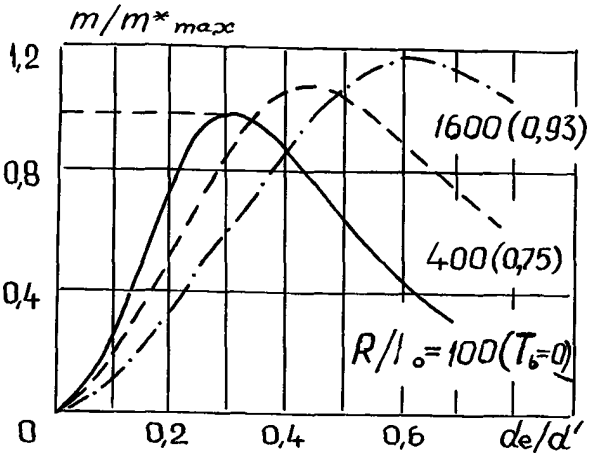


Figure 12 - Dependence of capacitance relation of a resonator with a lens-type PE ($d/t_o=20$) on electrode diameter

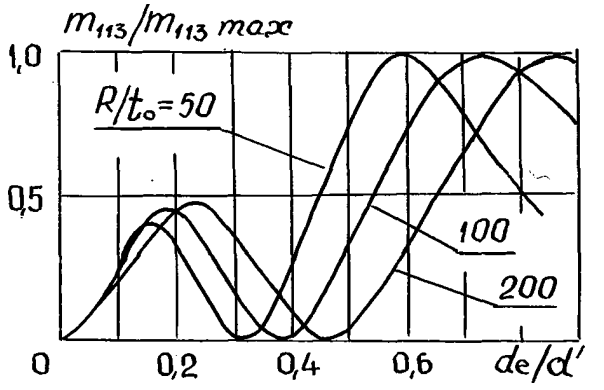


Figure 13 - Dependence of capacitance relations for anharmonic vibrations 113.131 of resonators with LPEs (d/to=14)

8. Determination of resonator frequency-temperature characteristics

For determination of frequency-temperature characteristics (FTC)) of quartz resonators with axially symmetic PEs of viriable thickness it is necessary to take into account the dependence of coefficients ρ, C_{ij} and PE geometrical dimensions on temperature (t) and angle of cut (β) in equation of motion with thickness-shear vibrations. This enables to make an analysis of FTCs of the complete spectrum of thickness-shear vibrations of axially symmetric PEs of rotated Y-cuts with a variable thickness. Figure 14a illusrates the calculated FTCs ($\beta=\beta_o+\Delta\beta$) of PEs having different radii of sphere beginning with $R=\infty$ (disk-type PE) and ending with a minimum permissible radius when $t_b=0$. The effect of a bevel on

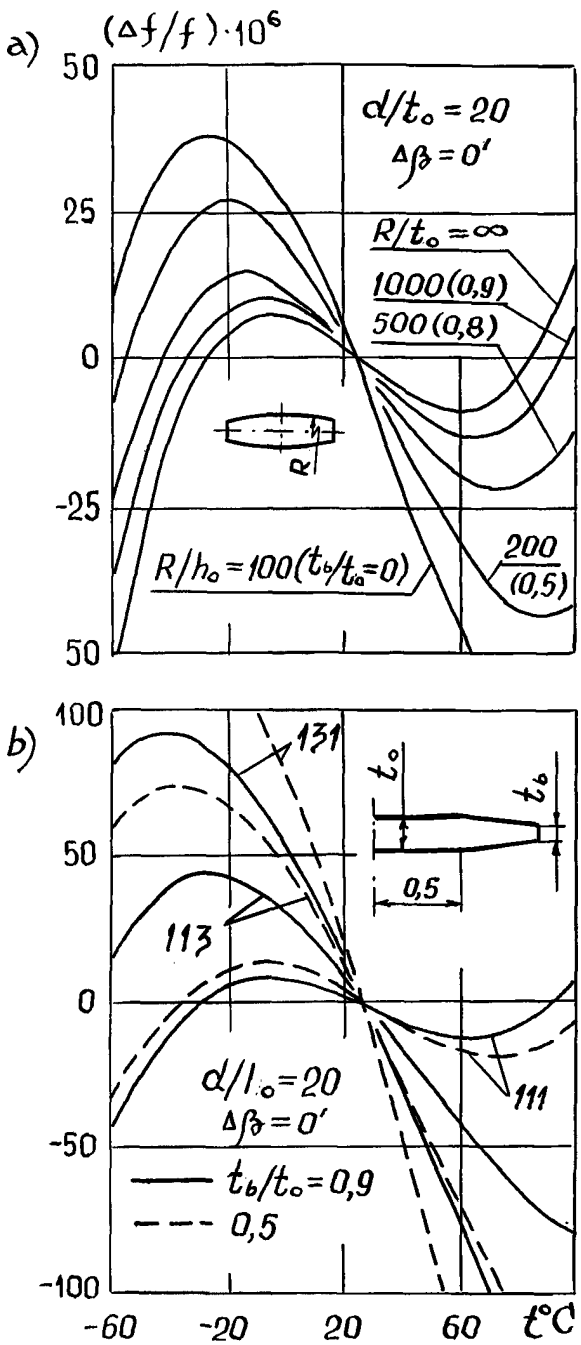


Figure 14 - Frequency- temperature characteristic of AT cut piezoelectric elements (β=35.0862°) with different surface curvature: a) LPE with different radii of sphere; b) disk-type PE with bevels

FTC of a harmonic and two anharmonic vibrations is shown in Figure 14b. In Figure 15 an example is given illustrating determination of local PE thickness changes on resonator FTC. Orientation error of a crystal plate for AT-cut LPE within the limits of ($\Delta\beta=\pm 1'$) causes FTC change shown in Figure 15:

curve 1-$\Delta\beta=-1'$; curve 2-$\Delta\beta=0'$; curve 3-$\Delta\beta=1'$ The ring notch of PE with the dimensions shown in the figure changes FTC equivalently to the change of crystal plate orientation for one minute (Fig. 15b). The curves of Figure 15c illustrate the effect of local electrode thickness change on FTC.

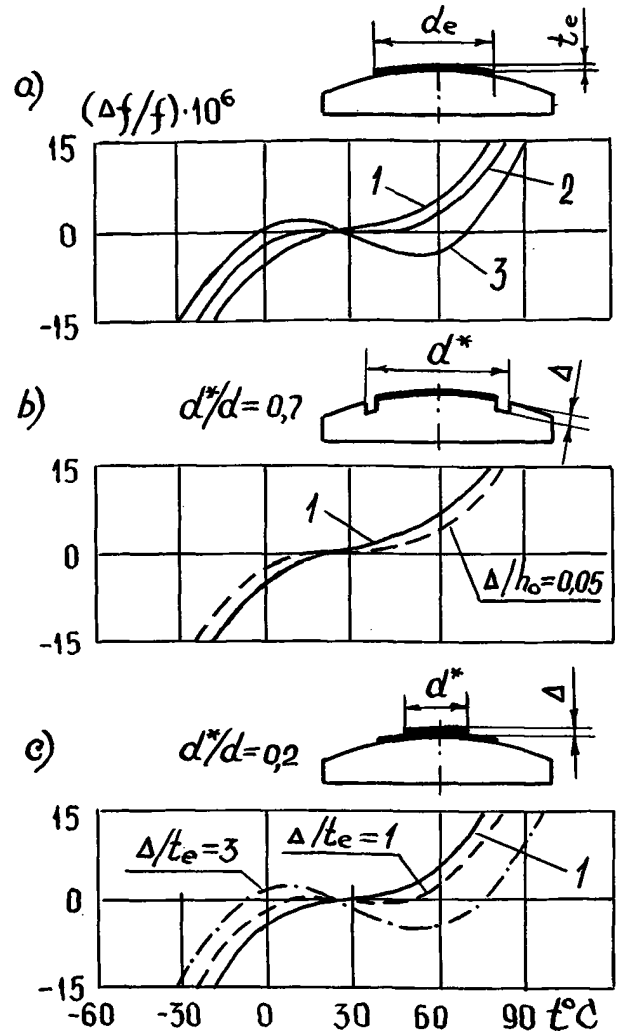


Figure 15 - Dependence of frequency-temperature characteristic on local electrode thickness changes and PE (d/to=20; R/to=250; te*=0.01; de/d=0.58)

The numerical methods of resonator FTC calculation developed enable to make a complex estimation of the influence of angle of cut, changes PE and geometry variations, local PE and electrode thickness changes on the resonator FTC, which is especially important for

high-frequency quartz resonatorss, having small PE thicknesses due to which all the above-mentioned local PE geometry and electrode changes become especially appreciable. They are also useful in designing multimode resonators, when it is necessary to ensure simultaneously FTC for each of the working vibrations.

Conclusion

The methods of numerical study of main quartz resonator characteristics and parameters confirm effectiveness of developed computer models of high-frequency resonators which enables to make theoretical investigations of not only the widely used, but also of new PE configurations, which can be easily obtained by modern progressive shaping techniques, for which a very precise experiment is necessary and, which cannot be always implemented in practice. Evaluattion of the complex effect of PE and electrode geomet- ries angle of cut and environmental influences makes it possible to improve resonator characteristics employing known traditional PE configurations and to design resonators with PEs of new geometries with the best qualities.

Acknowledgement

The author thanks Vitaly Maksimenko and his colleagues for the constant support in making numerous computational experiments.

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 Rectanqular 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 discontinuously 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 rectanqular 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 2249933,1941.

[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 Freqency 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 I.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 rezonators, 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 crystzal unit characteristics", Radioengineering, No.12, pp. 85-86, 1989 (in Russian)
[41]Vesselov G.I., Postnikov I.I., Samojlov V.S. "Modelling of lens-type piezoelectric elements for HF resonators", "Problem of microelectronics", MIET, pp.133-139, 1979 (in Russian)
[42]Postnikov I.I. "Frequency distances 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 virations of quartz crystal units with lens-type piezoelectric elements", Radioengineering, No.11, pp.58-60, 1986 (in Russian)

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

RELIABLE QUADRATIC FOR FREQUENCY-TURNOVER TEMPERATURE VS ORIENTATION OF ROTATED Y-CUT QUARTZ PLATE RESONATOR OSCILLATING IN C-MODE

MITSUO NAKAZAWA*, ATSUSHI ARAKI*, SEIICHI TAKAKUWA**,
AND
MOTOYASU HANJI**
*Shinsyu University , 500 Wakasato , Nagano-shi 380 , Japan ,
**Kinseki Co.Ltd. , Tokyo 201 , Japan

Abstract

This paper describes the reliable quadratic for frequency-turnover temperature vs orientation of c-mode in the temperature range $-196 \sim 300\,^{\circ}\mathrm{C}$. From the relation we can show the useful piezoelectric cuts of quartz for the resonator and sensor at the elevated temperature region. In particular, the experiments are carried out in the vicinity of AT-cut.

Introduction

As we know , the relationship between the frequency–turnover temperature and angle of cut for the rotated Y-cut quartz plate resonator oscillating in thickness vibration c-mode shows the nonlinear phenomenon . Bechmann, Ballato and Lukaszek have been investigated the frequency-turnover temperature vs orientation angle for AT-, BT-, CT-, and DT- cuts. [1-2] It is of great interest to know the theoretical and experimental behaviors of nonlinear phenomena for the frequency turning point of temperature vs the orientation in the rotated Y-cut, in particular, near the AT-cut, in view of the need for such plates in advanced frequency control, digital communication, etc. In the 1992 and 1993 Frequency Control Symposium an investigation has been presented [3-4] with respect to the nonlinear effect as mentioned above.

We now propose to show how the nonlinear effect of the frequency-temperature characteristics vs the angle of cut in the vicinity of AT-cut may be taken into account and we deduce the simple analytic and experimental formula for them. The formula shows approximately the quadratic and in this paper the reliable coefficients of quadratic are shown experimentally.

Analysis

According to references [3] and [4], the frequency of crystal resonator can be expressed by Taylor expansion theorem as follows:

$$f(\theta_0+\Delta\theta, T_0+\Delta T) = \sum_{n=1}^{n}\frac{1}{n!}(\Delta\theta\frac{\partial}{\partial\theta}+\Delta T\frac{\partial}{\partial T})^n f(\theta_0,T_0) + R_n, \quad (1)$$

where

θ : any angle of rotated Y-cut,

θ_0 : reference angle,

T : any temperature,

T_0 : reference temperature,
$\Delta\theta = \theta - \theta_0$, $\Delta T = T - T_0$

and R_n: residual term.

From eqn.(1) we have the following eqns.:

$$f(\theta_0+\Delta\theta, T_0+\Delta T) \cong f(\theta_0,T_0)(1+A\Delta T+B\Delta T^2+C\Delta T^3), \quad (2)$$

where

$$A = (d_T + d_{T\theta}\theta + \frac{1}{2}d_{T\theta\theta}\theta^2),$$
$$B = \frac{1}{2}(d_{TT} + d_{TT\theta}\theta + \frac{1}{2}d_{TT\theta\theta}\theta^2),$$
$$C = \frac{1}{6}(d_{TTT} + d_{TTT\theta}\theta + \frac{1}{2}d_{TTT\theta\theta}\theta^2),$$

and

$$d_T = (\frac{1}{f}\frac{\partial f}{\partial T})_{T_0,\theta_0} = \alpha,$$
$$d_{TT} = (\frac{1}{f}\frac{\partial^2 f}{\partial T^2})_{T_0,\theta_0} = \beta,$$
$$d_{TTT} = (\frac{1}{f}\frac{\partial^3 f}{\partial T^3})_{T_0,\theta_0} = \gamma,$$
$$d_{T\theta} = (\frac{1}{f}\frac{\partial^2 f}{\partial T\partial\theta})_{T_0,\theta_0}, etc. \quad (3)$$

0-7803-1945-1/94 $3.00 © 1994 IEEE

Now, for the first-order approximation we set eqn. (3) as follows:

$$A = d_T + d_{T\theta}\theta + \frac{1}{2}d_{T\theta\theta}\theta^2,$$

$$B = \frac{1}{2}(d_{TT} + d_{TT\theta}\theta),$$

and

$$C = \frac{1}{6}(d_{TTT} + d_{TTT\theta}\theta). \quad (4)$$

From eqn.(2) we have the following equation of frequency-turnover temperature T:

$$\frac{\partial f}{\partial T} = 0 :$$

$$3C\Delta T^2 + 2B\Delta T + A = 0. \quad (5)$$

Hence,

$$\Delta T = (T - T_0) = \frac{-B \pm \sqrt{B^2 - 3AC}}{3C}. \quad (6)$$

Similarly, we have the following quadratic: [3-4]

$$a\theta^2 + 2h\theta\Delta T + \Delta T^2 + 2g\theta + 2e\Delta T + c = 0, \quad (7)$$

where

$$\Delta\theta = \theta - \theta_0 = \theta, \theta_0 = 0,\ \Delta T = T - T_0, d_{TTT\theta} = 0\ ,$$

$$a = \frac{d_{T\theta\theta}}{d_{TTT}}\ , c = \frac{2d_T}{d_{TTT}}\ , e = \frac{d_{TT}}{d_{TTT}}\ ,$$

$$g = \frac{d_{T\theta}}{d_{TTT}},\ and\ \ h = \frac{d_{TT\theta}}{d_{TTT}}. \quad (8)$$

From eqn.(2) we obtain the following eqns.of inflection temperature T:

$$\frac{\partial^2 f}{\partial T^2} = 0 :$$

$$T = T_0 - \frac{B}{3C}, \quad (9)$$

and

$$T = T_0 - e - h\theta. \quad (10)$$

Experiments

The typical specifications of the quartz crystal resonator in this paper are as follows:

- Orientation: (YXl)θ, $\theta = 36.80° \pm 15"$,
- Frequency constant for fundamental c-mode: $N_0 = f_0 t_0 = 1.6588$MHz-mm at 25 ℃,
- Frequency: $f_0 = 9.99280$MHz at 25 ℃,
- Diameter of plate: d=12.5mm,
- Diameter of electrodes: $d_e = 6.0$ mm,
- Surface polish: FO ♯4000,
- Electrode films: Ni Cr+Au,
- Holder : HC-47/U,
- Annealing/baking: 12 hours at 120-130 ℃.

The resonator is measured to have the following nominal equivalent circuit values at room temperature and fundamental harmonic:

- Capacitance ratio: $r = \frac{C_0}{C_1}$=246,
- Motional resistance: $R_1 = 5\ \Omega$,
- Quality factor: Q=91920,
- Motional time constant: $\tau_1 = (2\pi f_0 Q)^{-1} = 0.173$ps,
- Motional capacitance: $C_1 = \frac{\tau_1}{R_1} = 34.65$ fF,
- Motional inductance: $L_1 = 7.32$ mH,
- Static capacitance: $C_0 = 8.53$ pF.

Figure 1 shows the block diagram for the frequency-temperature measuring system. Figure 2 shows the frequency-temperature characteristics of the quartz-plate vibrating in c-mode (θ=36.80°).

Figure 3 shows the frequency-turnover temperature vs angle of cut characteristics for the rotated Y-cut plates oscillating in c-mode.

In the figure solid line shows the quadratic curve drawn by using the reliable quadratic coefficients which were determined by experiments.

Figure 4 shows the resonance frequency characteristics of the plate resonator at room temperature (θ=36.80°). It is found that the thickness-shear c-mode is operating at 9.992803MHz and the unwnated modes keep away from the principal mode.

Quadratic

As we know,the mode change between the b- and c- modes for the quartz plate resonator occure at approximately $\theta = -24°$ and their elastic velocities change one another. Therefore, we must take care of this point. We now discuss the quadratic in the vicinity of AT-cut.

Table 1 is made by our experimental results, as shown in Fig.3.

In the Fig.2 we can calculate the first-, second- and third- order frequency-temperature coefficients as, by use of the least squares method,

$$d_T = -7.2390 \times 10^{-6}\ (\mathrm{K}^{-1}),$$

$$d_{TT} = -1.3050 \times 10^{-8}\ (\mathrm{K}^{-2}),$$

and

$$d_{TTT} = 4.8890 \times 10^{-10}\ (\mathrm{K}^{-3}),$$

where $T_0 = 25$ ℃.

In the quadratic formula (7) we assume that constant c is invariant for the c-mode near AT-cut.

Hence,

$c = \dfrac{2d_T}{d_{TTT}} = -29618\ (\mathrm{K}^2)$.

With c value known, only a,h,g, and e in eqn.(7) remain to be determined. We thus obtain the following eqn.(11), where

$A_i = -(\Delta T_i^2 + c)$,

$\Delta T_i = T_i - T_0$, $T_0 = 25$ ℃, and i=1,2,3 and 4.

$$\begin{bmatrix} {\theta_1}^2 & \theta_1\Delta T_1 & \theta_1 & \Delta T_1 \\ {\theta_2}^2 & \theta_2\Delta T_2 & \theta_2 & \Delta T_2 \\ {\theta_3}^2 & \theta_3\Delta T_3 & \theta_3 & \Delta T_3 \\ {\theta_4}^2 & \theta_4\Delta T_4 & \theta_4 & \Delta T_4 \end{bmatrix} \begin{bmatrix} a \\ 2h \\ 2g \\ 2e \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix}. \tag{11}$$

Hence, we have the following results:

$$\begin{aligned} a &= -521.3801\ (K^2), \\ b &= 1, \\ c &= -29618\ (K^2), \\ e &= 632.4305\ (K), \\ g &= 9606.8865\ (K^2), \\ h &= -17.8785\ (K), \\ d_T &= -7.2396 \times 10^{-6}\ (K^{-1}), \\ d_{TT} &= 30.9195 \times 10^{-8}\ (K^{-2}), \\ d_{TTT} &= 4.8890 \times 10^{-10}\ (K^{-3}). \end{aligned} \tag{12}$$

From these values, it follows that

$$\Delta = \begin{vmatrix} a & h \\ h & b \end{vmatrix} < 0 \quad and \quad D = \begin{vmatrix} a & h & g \\ h & b & e \\ g & e & c \end{vmatrix} \neq 0\ . \tag{13}$$

Therefore, eqn.(11) represents a hyperbola as shown in Fig.3 (solid line).

Appendix

The authors want to change the Figs.5 and 6 which were represented in our paper of 1993 Frequency Control Symposium[4] into the Fig.5 of this paper.

Conclusions

The authors performed both an analysis and experiments with respect to the frequency-turnover temperature vs orientation for the rotated Y-cut oscillating in c-mode in the vicinity of AT-cut. From these results the authors have obtained new relationships between the cut angles and the frequency-turnover temperatures on the elevated temperatures. These results can be applied to chemical sensors, resonators and transducers, etc.

Acknowledgements

This work was supported by Shinsyu University, Kinseki Co.,Ltd., Miyota Co.,Ltd., and San-ei Denshi Kogyo Co.,Ltd., Japan. The authors wish to thank Dr.Arthur Ballato and Mr.J.Gualtieri, USA for their help and suggestions, and also wish to thank Messrs. T.Sugiura , S.Kitamura , T.Umeda and W.kato of Shinshu University, and Messrs. A.Chiba, H.Akagawa, and T.Kakizawa, Japan for their assistance.

References

1 R.Bechmann, A.D.Ballato and T.J.Lukaszek: Higher-order temperature coefficients of the elastic stiffnesses and compliances of alpha-quartz, PIRE, Vol.50, pp.1812-1822, 1962.

2 R.Bechmann, A.D.Ballato and T.J.Lukaszek: Higher order temperature coefficients of the elastic stiffnesses and compliances of alpha-quartz, United States Army, Electronics Research And Development Laboratory , Fort Monmouth, N.J., 1963.

3 M. Nakazawa , M. Takeuchi , T. Moriizumi , H.Niiyama, A.Ballato, and T.Lukaszek: A theoretical study of quartz crystal resonators for chemical sensors, vibrating in c-mode, 1992 IEEE Frequency Control Symposium, pp.610-613, 1992.

4 M.Nakazawa, M.Takeuchi, H.Yamaguchi and A.Ballato : Thin rotated Y-cut quartz resonators vibrating in b-mode over a wide temperature range, 1993 IEEE Frequency Control symposium, pp.541-547, 1993.

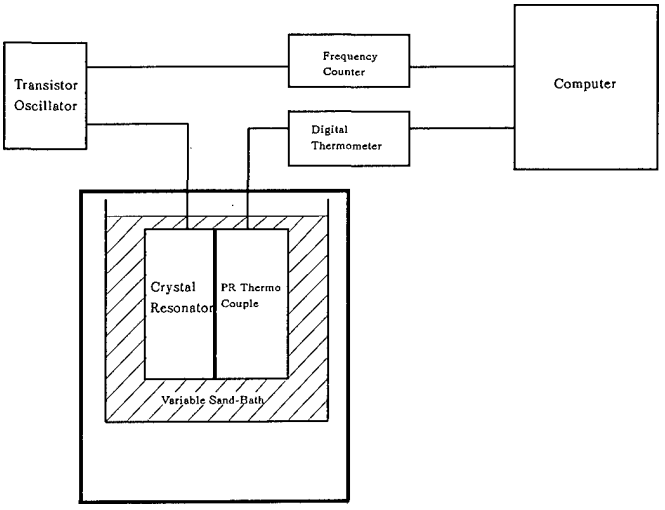


Fig.1. Circuit diagram for frequency-temperature characteristics.

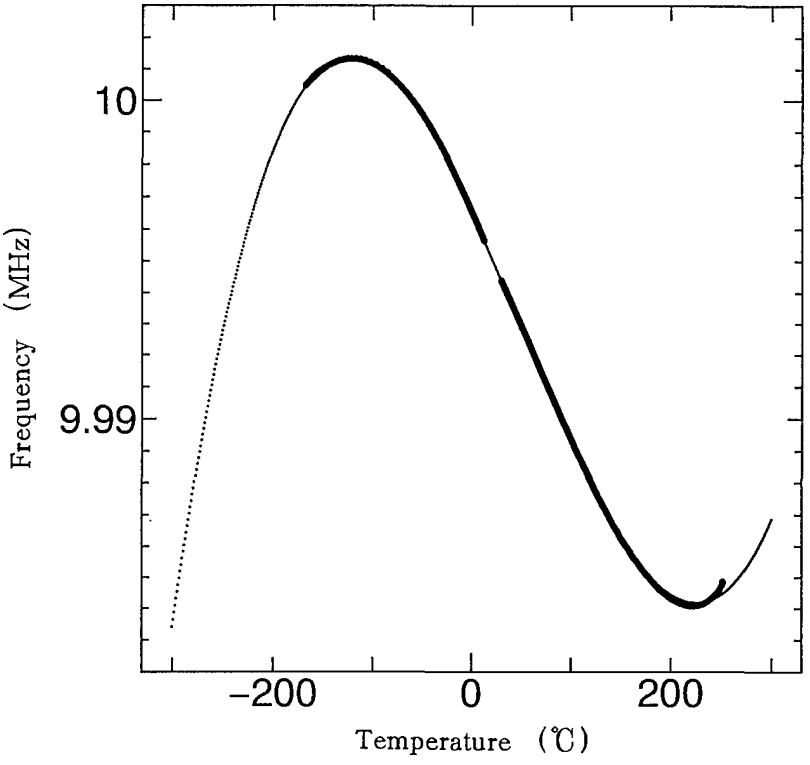


Fig.2. Frequency-temperature characteristics for the rotated Y-cut plate oscillating in c-mode in the vicinity of AT (θ =36.80 °).Solid line shows the experimental result and the dotted line shows the calculated value of least squares method.

Table 1. Frequency-turnover temperatures and angles of cut for rotated Y-cuts in the vicinity of AT

θ_i (°)	T_i (℃)	$\Delta T_i = T_i - T_0$ ($T_0 = 25$℃)
35.25	36.33	11.33
35.55	-46.00	-71.00
36.80	221.60	-196.60
36.80	-120.60	-145.60

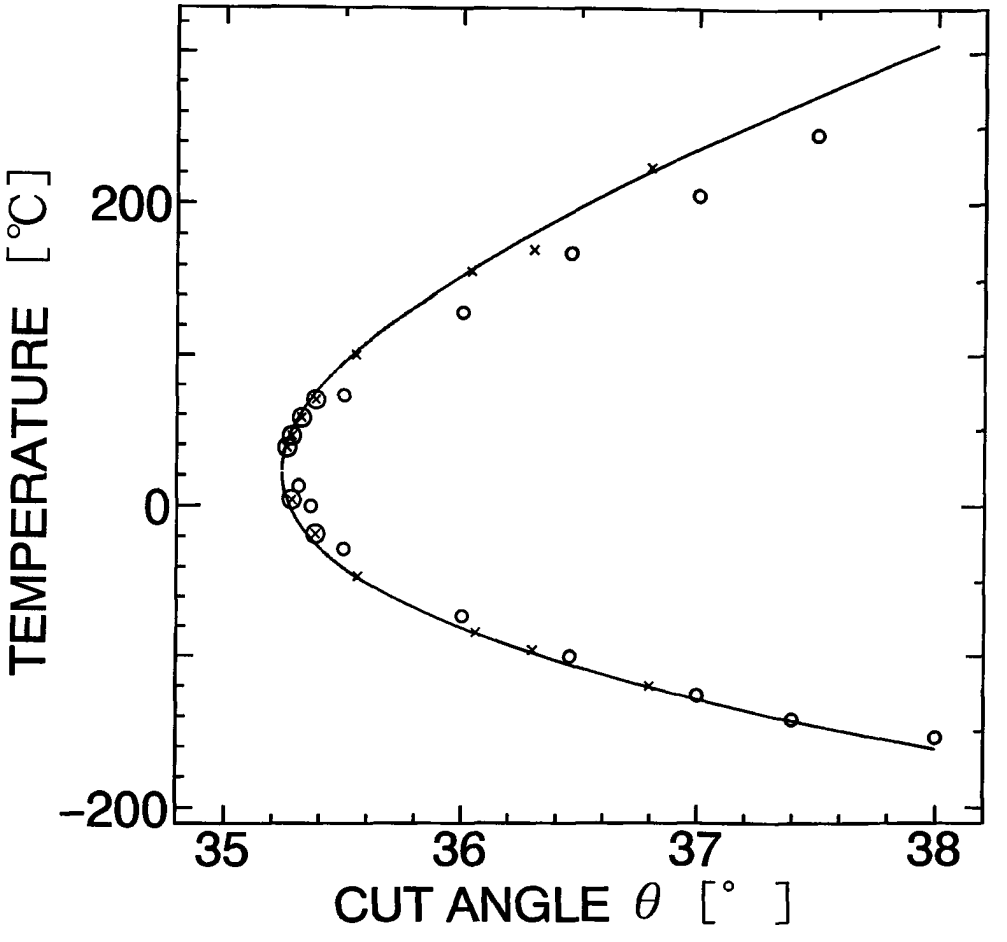
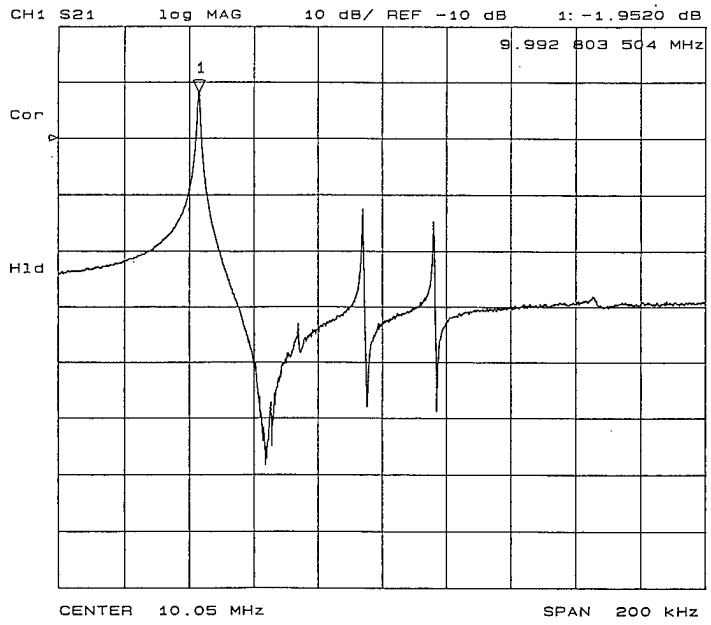


Fig.3. Frequency-turnover temperature vs angle of cut for the rotated Y-cut plate oscillating in c-mode in the vicinity of AT(x:present value, circle: BBL value[1,2],cross - circle:value obtained by Koga and Shinada, solid line: quadratic curve of present.).

Fig.4. Resonance frequency chracteristics for the rotated Y-cut plate vibrating in c-mode in the vicinity of AT(θ =36.80 °).



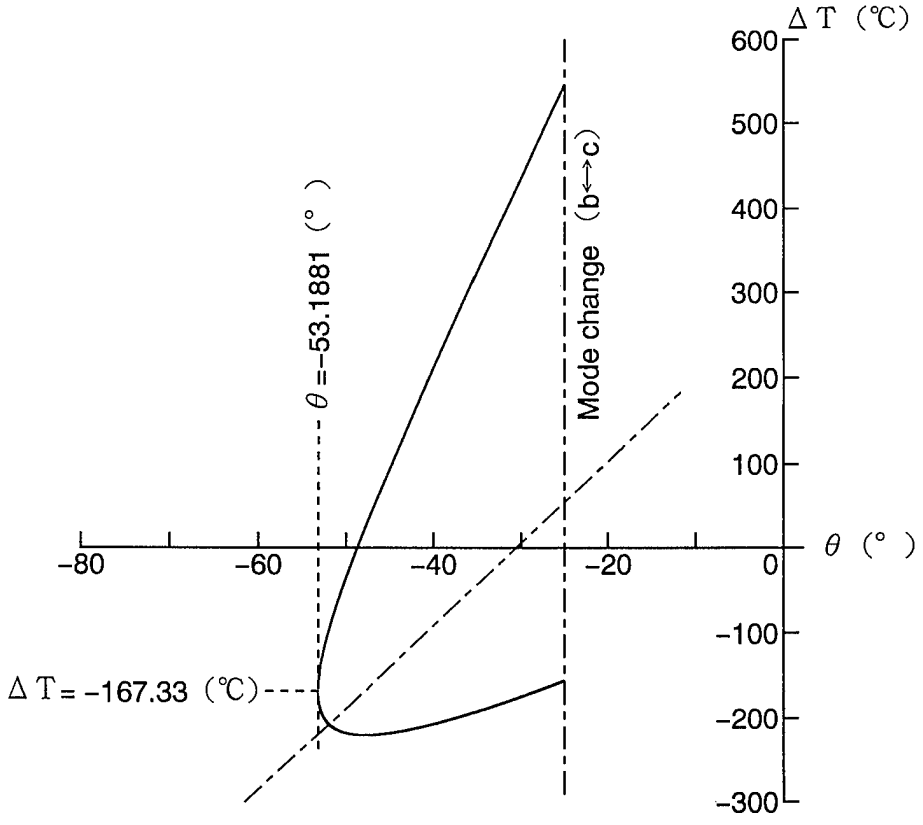


Fig.5. Frequency-turnover temperature vs angle of cut characteristics for the rotated Y-cut vibrating in b-mode in the vicinity of BT-cut. Solid line shows the quadratic calculated by using the quadratic coefficients which are determind by experiments

$(\Delta T=T-T_0,\ T_0=30℃)$.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

INCLUSION OF NON-UNIFORM DISTRIBUTION OF MOTION EFFECTS IN THE TRANSMISSION-LINE ANALOGS OF THE PIEZOELECTRIC PLATE RESONATOR: THEORY AND EXPERIMENT

John A. Kosinski, Arthur Ballato, Irina Mateescu*, and Ioan Mateescu*

US Army Research Laboratory
ATTN: AMSRL-EP-MD
Fort Monmouth, NJ 07703-5601
USA

*Institute of Physics and Technology of Materials
P.O. Box MG-7
Bucharest-Magurele
Romania

Abstract

The inclusion of the effects of a non-uniform distribution of vibratory motion in Ballato's exact transmission-line analogs of the piezoelectric plate resonator is addressed on both theoretical and experimental bases. It is shown that both the mass loading (series inductance) and piezoelectric coupling (transformer) terms are modified by the non-uniform distribution of motion. Typical magnitudes of the required correction factors are discussed, and representative experimental data are presented.

Introduction

Transmission-line equivalent electrical circuits for the piezoelectric plate resonator have been developed by a number of researchers. The equivalent circuits of greatest utility are Ballato's exact transmission-line analogs [1], since they are 1) exact solutions of the one-dimensional cases considered (and thus valid at all frequencies), and 2) provide descriptive analogs of the physical processes occurring at each point in the resonator structure. They are particularly useful in analyzing stacked structures and structures where multiple harmonics of the resonator are to be excited, such as in the dual-mode temperature sensing scheme originated by Schodowski [2].

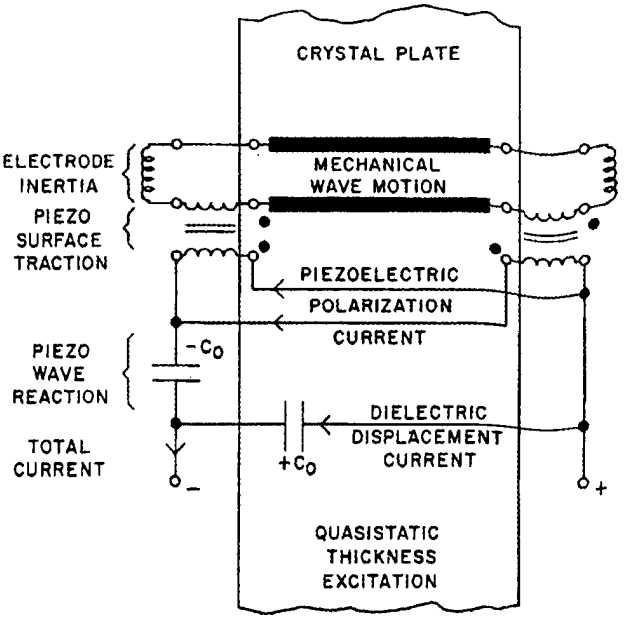


Fig. 1. Exact electrical equivalent network for a single simple thickness mode driven by thickness-field excitation in a laterally unbounded piezoelectric resonator.

0-7803-1945-1/94 $3.00 © 1994 IEEE

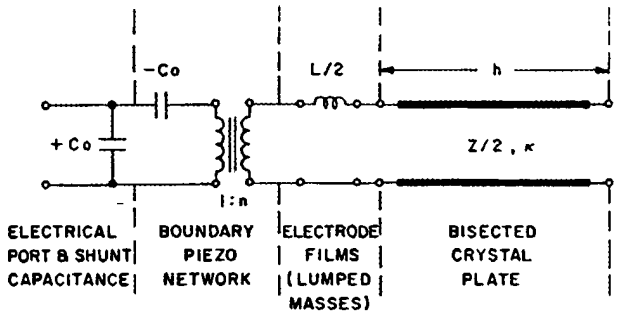


Fig. 2. Bisected version of the exact equivalent network shown in Fig. 1.

The correspondence between physical processes and equivalent circuit elements is illustrated in Fig. 1 for the case of thickness excitation (TE) of a single simple thickness mode. For a physically symmetric plate resonator, the equivalent circuit of Fig. 1 may be bisected yielding the simpler form shown in Fig. 2. In the one-dimensional approximation, the various circuit elements are readily calculated from the piezoelectric material constants and resonator dimensions as

$$C_0 = \frac{\varepsilon_{22}^S A_e}{2h} , \tag{1}$$

$$n^2 = \frac{C_0 A_e c_m k^2}{2h} , \tag{2}$$

$$L = mA_e , \tag{3}$$

$$Z = A_e \sqrt{\rho c_m} , \tag{4}$$

and

$$\kappa = \omega \sqrt{\rho / c_m} . \tag{5}$$

In (1) through (5), a uniform plane wave of modal stiffness c_m and piezoelectric coupling k is considered to propagate in a piezoelectric plate of mass density ρ. The resonator geometry is characterized by the plate thickness 2h, electrode area A_e, and electrode mass per unit area m.

The governing equations for the critical frequencies may be found by examination of the resonator impedance or admittance functions using either short-circuited (resonance) or open-circuited (antiresonance) conditions at the electrical port. The TE resonance frequencies are related as

$$\tan X = \frac{X}{k^2 + \mu X^2} \tag{6}$$

with

$$X = \kappa h = \frac{\pi}{2} \frac{f_{R\mu}^{(M)}}{f_{Ao}^{(1)}} . \tag{7}$$

The TE antiresonance frequencies are related as

$$\tan X = \frac{1}{\mu X} \tag{8}$$

with

$$X = \kappa h = \frac{\pi}{2} \frac{f_{A\mu}^{(M)}}{f_{Ao}^{(1)}} . \tag{9}$$

In (6) through (9), μ denotes the normalized electrode mass loading calculated as

$$\mu = \frac{m}{\rho h} . \tag{10}$$

As may be seen from (6) through (9), the harmonic frequency spectrum of a TE-driven, simple thickness mode in a piezoelectric plate resonator is determined by three quantities: the zero mass loading, fundamental antiresonance frequency $f_{Ao}^{(1)}$, the TE piezoelectric coupling k, and the electrode mass loading μ.

Theory

In an earlier paper, two of the authors had discussed the use of measurements of multiple harmonics in order to uniquely and precisely determine the electrode mass loading [3]. Discrepancies were noted between the "effective" mass loading determined from the measured frequency spectrum, and the "theoretical" mass loading determined from the physical and geometric

characteristics of the substrate and electrodes. Based on these results, a correction to the mass loading term of the TL analogs was proposed in order to account for the non-uniform distribution of motion found in practical plate resonators.

Subsequently, an experimental investigation of the proposed correction was performed by the other authors [4,5]. These results, along with observations made during the course of a recently published $Li_2B_4O_7$ material constants determination have led to a substantially improved understanding of the required corrections, as both the mass loading (series inductance) and piezoelectric coupling (transformer) terms are modified by the non-uniform distribution of motion.

"Effective" Mass Loading

Consider rotated Y-cut quartz, c-mode operation. The essentially thickness shear solution may be written

$$\hat{u}_1 = H\, f(x_1, x_3) \sin\eta x_2 \exp(i\omega t) \ ,$$
$$\hat{u}_2 \langle\langle \hat{u}_1 \ ,$$
$$\hat{u}_3 \approx 0 \ , \qquad (11)$$

where H is the modal amplitude and $f(x_1, x_3)$ describes the non-uniform distribution of motion (NUDM). From the boundary conditions for electrode films without intrinsic stress given by

$$\bar{c}_{66}\, \hat{u}_{1,2} - (e_{26}^2 / \varepsilon_{22}^S\, h)\, \hat{u}_1 = \mp 2\rho' h' \ddot{\hat{u}}_1 \quad \text{at} \quad x_2 = \pm h \ , \qquad (12)$$

one may obtain a transcendental frequency equation similar to (6):

$$\tan\eta h = \frac{\eta h}{k_{26}^2 + \mu_{eff}(\eta h)^2} \qquad (13)$$

where

$$\mu_{eff} = \left\{\frac{\rho\omega^2}{\bar{c}_{66}\eta^2}\right\}\mu \qquad (14)$$

with the term $\rho\omega^2$ determined by the dispersion relation.

"Effective" Piezoelectric Coupling

Two approaches may be used to examine the "effective" piezoelectric coupling.

1. Following Bechmann [6],

$$k_{eff}^2 = \Phi k^2 \ , \qquad (15)$$

$$\text{with } \Phi = \frac{\varphi^2}{\psi} \ , \qquad (16)$$

$$\varphi = \frac{1}{A_e}\iint_{A_e} f(x_1, x_3)\, dx_1 dx_3 \ , \qquad (17)$$

and

$$\psi = \frac{1}{A_q}\iint_{A_q} \left[f(x_1, x_3)\right]^2 dx_1 dx_3 \ . \qquad (18)$$

The term A_e in (17)represents the area of the driving electrodes, while the term A_q in (18) represents the area of the major plate surfaces.

2. Following Tiersten [7-9],

$$C_1 = -\frac{e_{26}\, H\, (-1)^{(n-1)/2}}{V\, h}\iint_{A_e} f(x_1, x_3)\, dx_1 dx_3 \ . \qquad (19)$$

The effects of the NUDM may be found by comparing C_1 for a given $f(x_1, x_3)$ to that obtained for the infinite flat plate with $f(x_1, x_3) = 1$.

Results of Calculations

The "effective" mass loading and piezoelectric coupling have been calculated for three types of AT-cut resonator: the infinite, flat plate (no energy trapping), the flat, trapped energy resonator (trapping only due to electrode mass loading), and the contoured, trapped energy resonator (trapping primarily due to the contour). The results of the calculations are summarized in Table I. The notation used in Table I is patterned after [7-9], in which the various quantities ($\hat{k}$, M_n, etc.) are defined. The calculations indicate that, for stress-free electrodes, μ_{eff} of the trapped energy resonator should be 1) less than the "true" μ calculated from the densities and thicknesses of the quartz plate and electrodes, and 2) should vary with harmonic with larger μ_{eff} on higher

TABLE I

COMPARISON OF "EFFECTIVE" MASS LOADING AND PIEZOELECTRIC COUPLING FOR THREE TYPES OF AT-CUT RESONATOR

	Infinite, Flat Plate	Flat, Trapped Energy Resonator	Contoured, Trapped Energy Resonator
$\hat{u}_1$	$H_n f(x_1,x_3)\sin\eta x_2 \exp(i\omega t)$	$H_{n\kappa\gamma} f(x_1,x_3)\sin\eta x_2 \exp(i\omega t)$, $x_1 \le l_1$, $x_3 \le l_3$ ≈ 0 elsewhere	$H_{nmp} f(x_1,x_3)\sin\eta x_2 \exp(i\omega t)$
$\hat{u}_2$	0	$\ll \hat{u}_1$	$\ll \hat{u}_1$
$\hat{u}_3$	0	≈ 0	≈ 0
H	$\dfrac{4(-1)^{(n-1)/2} e_{26} V}{c_{66}^E n^2\pi^2[1-(\omega_n^2/\omega^2)]}$	$\dfrac{16(-1)^{(n-1)/2} e_{26} V \sin\xi l_1 \sin\nu l_3}{c_{66}^E n^2\pi^2[1-(\tilde{\omega}_{n\kappa\gamma}^2/\omega^2)]\xi\nu L_{n\kappa\gamma}}$	$\dfrac{4(-1)^{(n-1)/2} e_{26} V h_o J_{nmp}}{c_{66}^E n^2\pi^2[1-(\tilde{\omega}_{nmp}^2/\omega^2)]L_{nmp}}$
$f(x_1,x_3)$	1	$\cos\xi x_1 \cos\nu x_3$	$\exp\left[-\alpha_n\left(x_1^2/2\right)\right] H_m\left[(\alpha_n)^{1/2} x_1\right] \exp\left[-\beta_n\left(x_3^2/2\right)\right] H_p\left[(\beta_n)^{1/2} x_3\right]$
$\rho\omega^2$	$\bar{c}_{66}\eta^2$	$M_n\xi^2 + c_{55}^E\nu^2 + \hat{c}_{66}\eta^2$	$\dfrac{n^2\pi^2\hat{c}_{66}}{4h_o^2}\left\{1+\dfrac{1}{n\pi}\left(\dfrac{2h_o}{R}\right)^{1/2}\left[\left(\dfrac{M_n}{\hat{c}_{66}}\right)^{1/2}(2m+1)+\left(\dfrac{c_{55}^E}{\hat{c}_{66}}\right)^{1/2}(2p+1)\right]\right\}$
μ_{eff}	μ	$\left\{1-\dfrac{8k_{26}^2}{n^2\pi^2}-2\mu+\dfrac{M_n}{\bar{c}_{66}}\left(\dfrac{\xi}{\eta}\right)^2+\dfrac{c_{55}^E}{\bar{c}_{66}}\left(\dfrac{\nu}{\eta}\right)^2\right\}\mu$	$\left\{1-\dfrac{8k_{26}^2}{n^2\pi^2}-2\mu\right\}\left\{1+\dfrac{1}{n\pi}\left(\dfrac{2h_o}{R}\right)^{1/2}\left[\left(\dfrac{M_n}{\hat{c}_{66}}\right)^{1/2}(2m+1)+\left(\dfrac{c_{55}^E}{\hat{c}_{66}}\right)^{1/2}(2p+1)\right]\right\}\mu$
Φ	1	$\dfrac{\sin^2\xi l_1 \sin^2\nu l_3}{(\xi l_1)^2(\nu l_3)^2\left(1_2+\dfrac{\sin 2\xi l_1}{4\xi l_1}\right)\left(1_2+\dfrac{\sin 2\nu l_3}{4\nu l_3}\right)}$	(n00) mode $\dfrac{4\pi}{\hat{A}_e\sqrt{\alpha_n\beta_n}}\cdot\dfrac{\operatorname{erf}^2\left(\sqrt{\dfrac{\alpha_n}{2}}l_1\right)\operatorname{erf}^2\left(\sqrt{\dfrac{\beta_n}{2}}l_3\right)}{\operatorname{erf}\left(\sqrt{\alpha_n}l_1\right)\operatorname{erf}\left(\sqrt{\beta_n}l_3\right)}$
C_1	$\dfrac{\varepsilon_{22}^S A_e}{2h}\cdot\dfrac{8\hat{k}_{26}^2}{n^2\pi^2}$	$\dfrac{\varepsilon_{22}^S A_e}{2h}\cdot\dfrac{8\hat{k}_{26}^2}{n^2\pi^2}\cdot\left\{\dfrac{4\sin^2\xi l_1 \sin^2\nu l_3}{\xi^2 l_1 \nu^2 l_3 L_{n\kappa\gamma}}\right\}$	$\dfrac{\varepsilon_{22}^S \hat{A}_e}{2h_o}\cdot\dfrac{8\hat{k}_{26}^2}{n^2\pi^2}\left\{\dfrac{h_o}{\hat{A}_e}\cdot\dfrac{J_{nmp}^2}{L_{nmp}}\right\}$

harmonics. Further, for stress-free electrodes, the amount of variation in μ_{eff} with harmonic number is determined by the magnitude of the piezoelectric coupling.

Experiment

The "effective" mass loading and piezoelectric coupling of flat, trapped energy AT-cut resonators have been examined experimentally. The resonators used in the experiments were fabricated using Sawyer cultured quartz plates with an orientation angle of θ=35°17'. The plates were approximately 0.34mm thick, 14mm in diameter, and optically polished.

Circular electrodes of 7mm diameter were deposited on the plates. Several different electrode metallizations were investigated, including evaporated CrNiAu, CrAu, and CrNi, and sputtered CrNi. For each type of metallization, a series of different electrode thicknesses between 1000Å and 3500Å were examined. The electrode thicknesses were monitored during deposition using a MIKI-MSV 1841 thickness monitor, and afterwards checked by interferometric means.

The initial deposition of the evaporated CrNiAu electrodes consisted of 100Å Cr, 500Å Ni, and 1000Å Au. A subsequent deposition added an additional 800Å Au. The initial deposition of the evaporated CrAu electrodes consisted of 200Å Cr and 1000Å Au, with an additional 1000Å Au added in a subsequent deposition. For the evaporated CrNi electrodes, the initial deposition consisted of 100Å Cr and 1000Å Ni. Subsequent depositions of Ni yielded total thicknesses of 2000Å, 3000Å, and 3500Å for the Ni layer. Sputtered CrNi electrodes of 1000Å, 2000Å, and 3000Å were formed using a 20%Cr/80%Ni target.

TABLE II
EXPERIMENTAL DATA - MASS LOADING

	Nominal Mass Loading	Effective Mass Loading (percent)			
Electrode Type	(percent)	M=1	M=3	M=5	M=7
CrNiAu	0.58	0.14	0.38	0.42	0.45
	0.95	0.35	0.64	0.69	0.71
CrAu	0.49	0.25	0.54	0.58	0.61
	0.95	0.62	0.93	0.98	1.01
sputtered CrNi	0.20	0.16	0.38	0.41	0.44
	0.41	0.45	0.73	0.77	0.79
	0.61	0.55	0.85	0.89	0.91
CrNi	0.23	0.13	0.39	0.43	0.45
	0.44	0.39	0.35	0.72	0.74
	0.65	0.60	0.58	0.96	0.98
	0.76	0.71	1.04	1.09	1.11

Sample Data

For each resonator, the resonance and antiresonance frequencies with and without electrodes of the M=1, 3, 5, 7 harmonics were measured using an HP4194A Impedance/Gain-Phase Analyzer. The measured frequencies were then used in conjunction with the transcendental frequency equations to determine the "effective" mass loading and piezoelectric coupling values for each harmonic. Typical results for the "effective" mass loading and piezoelectric coupling are shown in Tables II and III, respectively.

The equivalent circuit parameters were also measured, and the "effective" piezoelectric coupling values were also determined by comparing the measured capacitance ratios to the theoretical values for the infinite, flat plate. The values determined in this way are also listed in Table III. In order to use the measured C_1 values to determine k_{eff}, the electrically driven area must be precisely known. Using the measured capacitance ratio C_0/C_1 to determine k_{eff} eliminated uncertainties due to fringing fields at the electrode edges.

Data Analysis

Analysis of the "effective" mass loading data is facilitated by normalization of the data for the various electrode types as shown in Table IV. The value of μ=0.60% is chosen since it is common to all the data considered. We observe that μ_{eff} increases with increasing harmonic as expected, however the magnitude of the increase is greater than expected. Further, μ_{eff} is less than μ_{nom} at all harmonics only for the CrNiAu electrodes.

An analysis of the "effective" piezoelectric coupling data is presented in Table V. With the exception of a few outliers, the values of Bechmann's factor Φ are quite consistent for all of the electrode metallizations considered. From the average values of Bechmann's factor, the spatial extent of an equivalent isotropic cosine amplitude distribution can be calculated as $0.96d_e$, $0.85d_e$, $0.81d_e$, and $0.80d_e$ for the M=1, 3, 5, 7 harmonics respectively.

TABLE III

EXPERIMENTAL DATA - PIEZOELECTRIC COUPLING

Electrode Type	Nominal Mass	Measurement	Measured Effective Piezoelectric Coupling (%)			
	Loading (%)	Technique	M=1	M=3	M=5	M=7
CrNiAu	0.58	tanX	6.37	5.32	4.61	4.46
		C_0/C_1	6.38	5.24	4.62	4.35
	0.95	tanX	6.79	5.32	4.86	4.82
		C_0/C_1	6.81	5.43	4.87	4.89
CrAu	0.49	tanX	6.77	6.05	5.28	6.14
		C_0/C_1	6.80	6.06	5.29	6.14
	0.95	tanX	6.93	5.69	6.00	8.01
		C_0/C_1	6.97	5.68	6.01	8.00
sputtered CrNi	0.20	tanX	6.48	5.19	4.61	3.64
		C_0/C_1	6.50	5.20	4.59	4.65
	0.41	tanX	6.79	5.19	4.62	4.47
		C_0/C_1	6.83	5.25	4.56	4.66
	0.61	tanX	6.93	5.33	4.87	5.17
		C_0/C_1	6.96	5.27	5.00	5.36
CrNi	0.23	tanX	7.03	5.83	5.10	4.81
		C_0/C_1	7.05	5.78	5.11	4.83
	0.44	tanX	6.97	5.58	4.87	4.47
		C_0/C_1	6.99	5.56	4.80	4.43
	0.65	tanX	6.97	5.32	4.87	5.17
		C_0/C_1	7.02	5.39	4.80	5.11
	0.76	tanX	7.07	5.46	4.62	4.84
		C_0/C_1	7.11	5.40	4.64	4.70

TABLE IV

"EFFECTIVE" MASS LOADING DATA NORMALIZED TO NOMINAL MASS LOADING OF 0.60%

	Effective Mass Loading (percent)			
Electrode Type	M=1	M=3	M=5	M=7
CrNiAu	0.14	0.39	0.43	0.47
CrAu	0.31	0.66	0.71	0.74
sputtered CrNi	0.54	0.84	0.88	0.90
CrNi	0.55	0.54	0.89	0.90

Conclusions

The exact transmission-line analogs of the infinite, flat plate piezoelectric resonator with a uniform distribution of vibratory motion can be applied to finite plates with non-uniform distribution of motion provided that μ_{eff} and k_{eff} are properly accounted for.

The nature of μ_{eff} and k_{eff} have been examined both theoretically and experimentally. For the trapped energy resonator, the "effective" mass loading depends upon the dispersion relation, leading to a harmonic dependence of μ_{eff}, as well as μ_{eff} being less than μ_{nom}. For an arbitrary non-uniform distribution $f(x_1,x_3)$, the effective piezoelectric coupling can readily be calculated using Bechmann's factor Φ as outlined in (15)-(18). The

TABLE V

PIEZOELECTRIC COUPLING DATA ANALYSIS

Electrode Type	Nominal Mass	Bechmann's Factor Φ			
	Loading (%)	M=1	M=3	M=5	M=7
CrNiAu	0.58	0.53	0.37	0.28	0.26
	0.95	0.60	0.37	0.31	0.30
CrAu	0.49	0.59	0.47	0.36	0.49
	0.95	0.62	0.42	0.47	0.83
sputtered CrNi	0.20	0.54	0.35	0.28	0.17
	0.41	0.60	0.35	0.28	0.26
	0.61	0.62	0.37	0.31	0.35
CrNi	0.23	0.64	0.44	0.34	0.30
	0.44	0.63	0.40	0.31	0.26
	0.65	0.63	0.37	0.31	0.35
	0.76	0.65	0.39	0.28	0.30
Average Value		0.60	0.39	0.31	0.30

factor Φ has been determined experimentally for the case of the AT-cut c-mode plano-plano resonator. The values Φ=0.60, 0.39, 0.31, 0.30 are observed for the M=1, 3, 5, 7 harmonics respectively.

References

[1] A. Ballato, "Transmission-Line Analogs for Stacked Piezoelectric Crystal Devices," in Proceedings of the 26th Annual Frequency Control Symposium, 1972, pp. 86-91.

[2] S. Schodowski, "Resonator Self-Temperature-Sensing Using a Dual-Harmonic-Mode Crystal Oscillator," in Proceedings of the 43rd Annual Frequency Control Symposium, 1989, pp. 2-7.

[3] J. Kosinski, A. Ballato, and S. Mallikarjun, "Mass Loading Measurements of Quartz Crystal Plates," in Proceedings of the 43rd Annual Frequency Control Symposium, 1989, pp. 365-371.

[4] I. Mateescu and E. Candet, "Non-Uniform Distribution of Motion Influence on the Effective Mass-Loading in AT-cut Quartz Resonators," in Proceedings of the 1992 IEEE Frequency Control Symposium, 1992, pp. 561-566.

[5] I. Mateescu and I. Mateescu, "Complex Mass-Loading Effects in AT-cut Quartz Crystal Resonators," in Proceedings of the 7th European Frequency and Time Forum, 1993, pp. 63-66.

[6] R. Bechmann, "Über Dickenschwingungen piezoelektrischer Kristallplatten," Archiv. der elekr. Übertragung, vol. 16, pp. 361-368, 1952.

[7] H. F. Tiersten, "Analysis of Intermodulation in Thickness-Shear and Trapped Energy Resonators," Journal of the Acoustical Society of America, vol. 57, no. 3, pp. 667-681, March 1975.

[8] H. F. Tiersten, "Analysis of Trapped-Energy Resonators Operating in Overtones of Coupled Thickness Shear and Thickness Twist," Journal of the Acoustical Society of America, vol. 59, no. 4, pp. 879-888, April 1976.

[9] H. F. Tiersten and R. C. Smythe, "An Analysis of Contoured Resonators Operating in Overtones of Coupled Thickness Shear and Thickness Twist," Journal of the Acoustical Society of America, vol. 65, no. 6, pp. 1455-1460, June 1979.

APPENDIX - NONZERO $\hat{u}_2$

For the flat, trapped energy resonator, $\hat{u}_2 << \hat{u}_1$, but $\hat{u}_2 \neq 0$. When the complete displacement solution

$$\hat{u}_1 = \left(B_1^{(1)} \sin \eta x_2 + B_1^{(2)} \sin \kappa\eta x_2\right) \cos \xi x_1 \cos \nu x_3 \exp(i\omega t) \ , \qquad \text{(A1)}$$

$$\hat{u}_2 = \left(B_2^{(1)} \sin \eta x_2 + B_2^{(2)} \sin \kappa\eta x_2\right) \sin \xi x_1 \cos \nu x_3 \exp(i\omega t) \tag{A2}$$

with

$$B_2^{(1)} = r\xi B_1^{(1)} / \eta \ , \tag{A3}$$

$$B_1^{(2)} = -r\xi B_2^{(2)} / \kappa\eta \ , \tag{A4}$$

and

$$B_2^{(2)} = (-1)^{(n+1)/2}\left(c_{22}^E r + c_{12}^E\right)\xi B_1^{(1)} / c_{22}^E \kappa\eta \sin \kappa\eta h \tag{A5}$$

is considered, the transcendental frequency equation becomes

$$\frac{\sin \eta h + \delta}{\cos \eta h + \varepsilon} = \frac{\eta h}{k_{26}^2 + \mu_{eff}(\eta h)^2} \tag{A6}$$

with

$$\delta \equiv -\frac{(-1)^{(n+1)/2}\left(c_{22}^E r + c_{12}^E\right)}{\bar{c}_{66}} \cdot \left(\frac{\xi}{\eta}\right)^2 \tag{A7}$$

$$\varepsilon \equiv \left\{\left(r - k_{26}^2\right)\cos \eta h + \left(1 - r - k_{26}^2\right)\left[\frac{\left(c_{22}^E r + c_{12}^E\right)(-1)^{(n+1)/2}}{\kappa c_{22}^E}\right]\cot \kappa\eta h\right\}\left(\frac{\xi}{\eta}\right)^2 \tag{A8}$$

For AT-cut quartz, n=1, both δ and ε are approximately $(\xi/\varepsilon)^2$. The lateral wavenumber ξ is related to the electrode diameter, while the thickness propgagtion wavenumber η is related to the plate thickness. As a result, the error terms δ and ε cuased by the nonzero $\hat{u}_2$ displacement are on the order of (plate thickness/electrode diameter)^2.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

IMPROVED Ω-SCAN FOR SEPARATE MEASUREMENT OF TRUE AT-CUTTING ANGLES AND X-MISCUTTING ANGLES FOR ROUND QUARTZ BLANKS

Bernhard Morys, Hans Bradaczek and Gerhard Hildebrandt

Institut für Kristallographie der Freien Universität Berlin

Abstract

The Improved Ω-Scan is a new method for measuring the AT-cutting angle of round quartz blanks uninfluenced by an X-tilt. From this true AT-cutting angle the frequency/temperature curve of the quartz oscillator can be calculated. This method combines all positive properties of the known Ω-Scan with the possibility to select even strongly miscutted blanks correctly.

To get more information about the cutting orientation of a quartz blank than the simple Ω-Scan was able to obtain, four - instead of two - reflections at two different lattice planes - instead of one - are measured. From the position of these four reflections the true AT-cutting angle and the X-tilt can be determined separately.

Cutting errors in the production process of AT-cutted quartz blanks and their influence on the frequency/temperature curve of quartz oscillators

In the production process of AT-cutted quartz blanks from Y-bars mainly two different kinds of cutting errors occur: a difference of the cutting angle from the nominal value of 35° 15' on the one hand and an X-miscutting angle unequal to zero - as well called X-tilt - an the other hand. These two cutting errors describe differences from the correct cutting orientation in two perpendicular directions. Figure 1 shows the correct cutting orientation on the left side (1), an cutting orientation with incorrect cutting angle in the middle (2) and with an X-tilt on the right side (3). While even small variations of the cutting angle influence the frequency/temperature stability of the

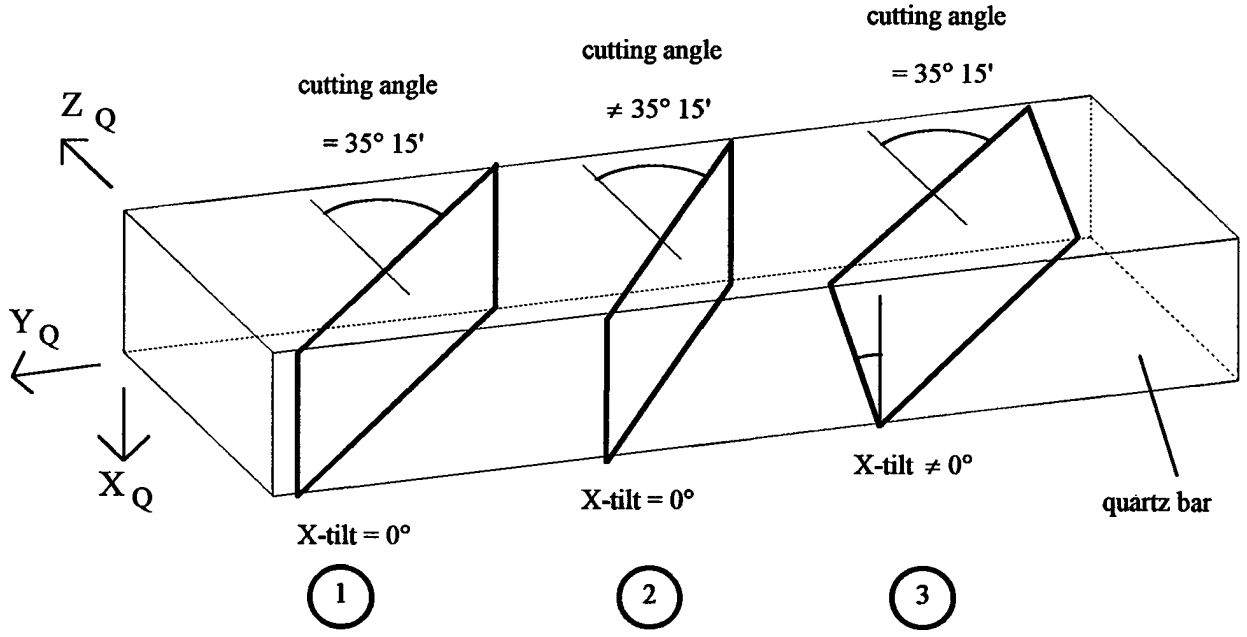


Figure 1. Orientation of the correct AT-cut and two different cutting errors in a quartz bar

0-7803-1945-1/94 $3.00 © 1994 IEEE

quartz oscillator strongly, an X-tilt - in a certain range - changes this property much less. The X-miscutting angle should be as small as possible.

To reduce costs, the cutting angle of round quartz blanks is measured after the cutting, rounding and polishing process, to select only those blanks for assembly, which meet the quality requirements. For this purpose the Ω-Scan is known as a reliable method for measuring big quantities of AT-cutted round quartz blanks.

Measurement principle of the simple Ω-Scan

The Z_Q-axis of the quartz co-ordinate system encloses with the $(01\bar{1}1)$ lattice plane of quartz an angle of 38° 13', and with the sectional plane the cutting angle of 35° 15' ± 30'. The difference of these two angles is called true AT-cutting angle α_Y and is equal to about 2° 58'. As shown in figure 2 the angle α between the blank surface and the $(01\bar{1}1)$ lattice plane consists of two components: the true AT-cutting angle α_Y and the so called X-miscutting angle α_X.

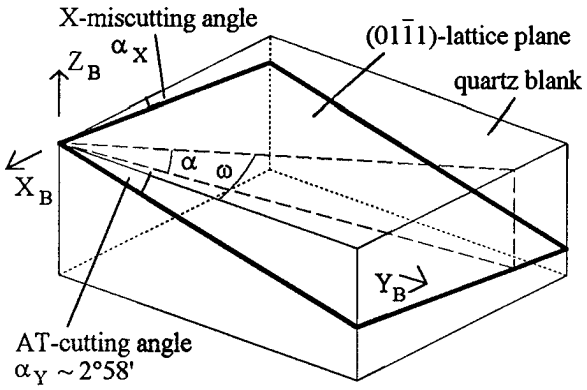


Figure 2. Quartz blank with an X-tilt

For the Ω-Scan a blank is rotated around its Z_B-axis of the blank co-ordinate system, which is perpendicular to the blank surface, while being exposed to a narrow X-ray beam. The angle between the incident X-ray beam and the surface of a reference blank - with known cutting angle - is adjusted in such a way that two Bragg-reflections from the $(01\bar{1}1)$ lattice plane appear in a distance of ± 35° rotation from the correct position. After this adjustment the AT-cutting angle α_Y of other blanks is calculated from the actual distance of the measured two reflections relatively to the cutting angle α_{YRef} of the reference blank.

For quartz blanks with an X-miscutting angle α_X smaller than 5', the Ω-Scan delivers results with extremely high accuracy better than ± 6" in a short measurement time of 2 seconds per blank. Bigger X-miscutting angles however enlarge the measured AT-cutting angle and falsify the result of this method strongly. Hereby the accuracy is decreased to an intolerable level. Figure 3 shows the enlargement of the measured AT-cutting angle in comparison with the true AT-cutting angle in dependence on an X-tilt.

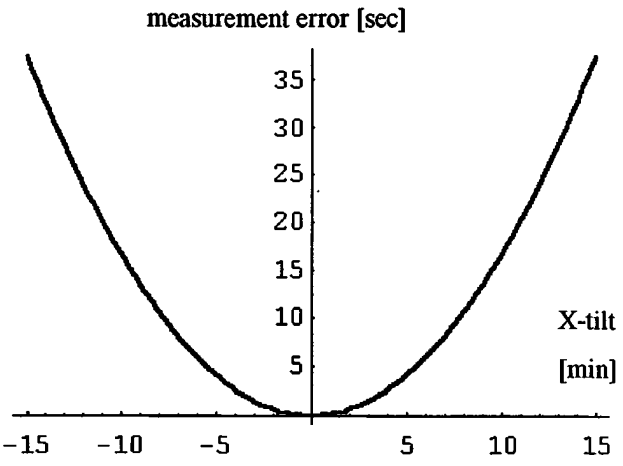


Figure 3. Enlargement of measured AT-cutting angle of simple Ω-Scan in dependence on an X-tilt

The Ω-Scan determines the angle α between the blank surface and the $(01\bar{1}1)$ lattice plane. As mentioned above and shown in figure 2, α consists of two perpendicular components α_Y and α_X. Only α_Y influences the frequency/temperature curve. As the reason for the loss of accuracy, the Ω-Scan is unable to determine these two components of α separately! With one known variable it is impossible to determine three unknown variables in the system of equations (1) definitely. This system of equations shows also, that α_Y can be calculated definitely until the X-miscutting angle α_X is zero.

$$\begin{pmatrix}\alpha_Y\\ \alpha_X\end{pmatrix} = \alpha\begin{pmatrix}\cos\omega\\ -\sin\omega\end{pmatrix} \qquad (1)$$

Measurement principle of the Improved Ω-Scan

To solve the described problem of the simple Ω-Scan and to determine the AT-cutting angle and the X-miscutting angle of a quartz blank separately, the „Improved Ω-Scan" was developed. With one single narrow incident X-ray beam four Bragg-reflections at two lattice planes, $(01\bar{1}1)$ and $(02\bar{2}3)$, whose relative position in the structure of quartz is known, are measured. Figure 4 shows the new measurement arrangement, figure 5 a quartz blank and both lattice planes $(01\bar{1}1)$ and $(02\bar{2}3)$. The reflected beams are detected by one and the same scintillation counter, which has a three hole plumb mask to screen other reflections and noise.

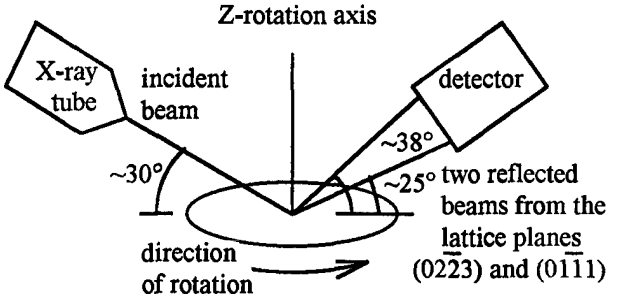


Figure 4. Arrangement of Improved Ω-Scan

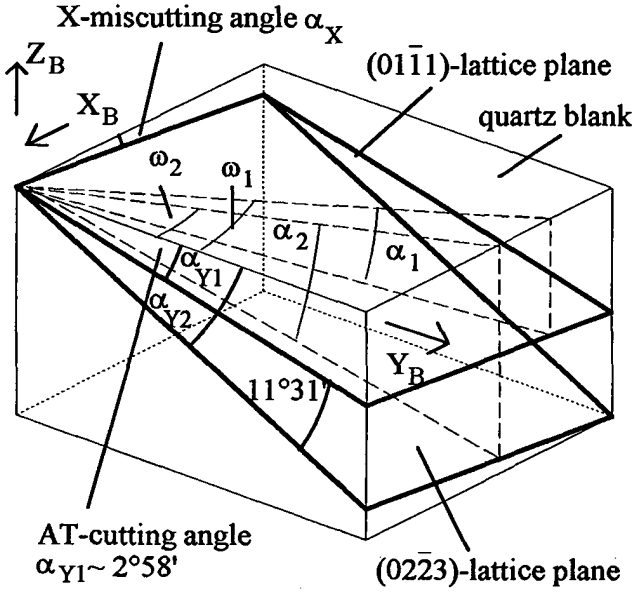


Figure 5. Quartz blank with an X-tilt

The four measured reflections are shown in figure 6. From these reflections the two angles α_1 - in simple Ω-Scan called α - and α_2 between the blank surface and the two lattice planes can be determined, as explained for the simple Ω-Scan. Each of these two measured angles consists, as described above, of two perpendicular components, α_{y1} - in simple Ω-Scan called α_Y - and α_{y2} in Y direction and the X-miscutting angle α_X in X direction, which is the same for both lattice planes. The angle between the planes, $\alpha_{y2} - \alpha_{y1}$, is a known quartz constant and does not depend on cutting errors.

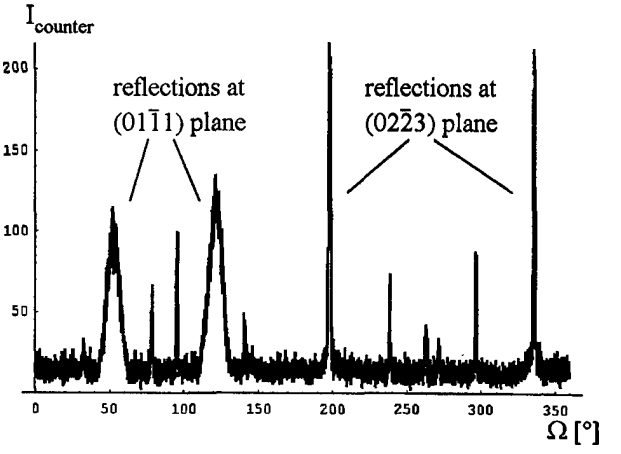


Figure 6. Four measured X-ray reflections at lattice planes $(01\bar{1}1)$ and $(02\bar{2}3)$

By solving the system of equations (2), the true AT-cutting angle α_{Y1} and the X-miscutting angle α_X can be determined separately and definitely.

$$\begin{pmatrix}\alpha_{Y1}\\ \alpha_X\end{pmatrix} = \alpha_1\begin{pmatrix}\cos\omega_1\\ -\sin\omega_1\end{pmatrix}$$

$$\begin{pmatrix}\alpha_{Y2}\\ \alpha_X\end{pmatrix} = \alpha_2\begin{pmatrix}\cos\omega_2\\ -\sin\omega_2\end{pmatrix} \qquad (2)$$

$$\alpha_{Y2} - \alpha_{Y1} = 11°31'$$

Confirmation of the Improved Ω-Scan

The function of the "Improved Ω-Scan" was confirmed by the following test with square blanks. The blank support was artificially moved up and down at one of three support points to simulate different cutting angles α_1. The blank was positioned to change only the X-tilt α_X - without influencing the true AT-cutting angle α_{y1} - by moving the support. To find the correct position, a square blank with known Y and X direction was used. The test proved, that the measured true AT-cutting angle α_{Y1} was not influenced by moving the support in the whole tested range of ± 1° and that the measured X-tilt was proportional to the support disadjustment. Both results verify the theory of the "Improved Ω-Scan".

Whereas the simple Ω-Scan method was unable to deliver correct values for the true AT-cutting angle when larger X-tilts were present, the new "Improved Ω-Scan" arrives at the true AT-cutting angle uninfluenced by the X-tilt.

References

[1] B. Nestler, H-J Kuhr, G. Hildebrandt and H. Bradaczek : Novel use of a commercial goniometer for sorting round quartz blanks Meas. Sci. Technol. 2 (1991) 528-531

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ALTERNATIVES TO OZONE DEPLETING SUBSTANCES IN ELECTRONIC CLEANING

Dr. John R. Stemniski

Charles Stark Draper Laboratory, Inc.

Evidence indicating that the earth's ozone layer is being depleted by man-made chemicals such as chlorofluorocarbons (CFCs) and other halogen containing solvents has challenged the scientific community to seek alternatives to precision and electronic cleaning. The crisis reached a climax when it was declared at a meeting of concerned nations in Montreal, Canada that phase-out of production of these substances was the only way to effectively reduce the impact of ozone-depleting substances (ODSs). Initially the Protocol dictated that there would be a progressive reduction in production with complete phase-out at the turn of the century. Present evidence indicates that the ozone layer is being depleted at a rate faster than anticipated. Subsequent meetings of the Parties to the Montreal Protocol have resulted in an accelerated phase-out and effectively production would cease 1 January 1996, if not sooner.

Adjustments to the Protocol

CFCs

(a) Seventy-five percent reduction of its calculated level of consumption by the 1986 level by 1994.

(b) One hundred percent phase-out by 1996.

- Parties operating under par. 1 of Article 5 (i.e., parties where consumption does not exceed 300 gm/capita) may exceed by 10% of (a).
- Parties operating under par. 1 of Article 5 may exceed the limit in (b) by not more than 15% of (b) where essential use is determined.

Methyl Chloroform (1,1,1-trichloroethane)

- Consumption shall not exceed level of 1989 by 1993.
- Fifty percent of its calculated level of consumption of 1989 by 1994.
- One hundred percent phase-out by 1996.

Amendments to the Protocol

Hydrochlorofluorocarbons (HCFCs)

(a) Freeze in 1996 based on (b).
(b) 3.1% of calculated level of CFC consumption plus calculated level of HCFCs in 1989.
(c) 35% reduction by 2004.
(d) 65% reduction by 2010.
(e) 90% reduction by 2015.
(f) 99.5% reduction by 2020.
(g) 100% phase-out by 2030.

Halons

- One hundred percent phase-out by 1994.

Carbon Tetrachloride

- Eighty-five percent reduction of its calculated level of consumption of 1989 by 1995.
- One hundred percent phase-out by 1996.

0-7803-1945-1/94 $3.00 © 1994 IEEE

Alternatives to Ozone Depleting Substances Achieving the Goals of Protocol Resolutions

The task of achieving the cleaning results realized by ODSs using alternatives is indeed formidable. CFCs and other ODSs possessed many desirable properties for cleaning with the only exception of the property that they deplete the ozone layer. Replacement in kind would not be easy. In fact no "drop-in" replacement for cleaning solvents has been made available to date.

A partial list of requirements for acceptable replacements should include at least the following:

- Effective cleaner
- Zero or low ozone depleting potential, low low global warming potential, and be a low volatile organic compound
- Non-toxic
- Non-flammable
- Chemically inert
- Low cost

To realize these goals several approaches can be taken. New solvents can be developed. A return to the pre-CFC era using old solvents and methods could be re-established. New solvent systems could be developed. New processes involving other than solvents or solvent systems could be employed.

Instituting new techniques not requiring solvents for cleaning is possible. An example is the "no-clean" solder process. New designs not requiring the rigorous cleaning so accustomed to being employed would alleviate the demand placed on the new alternatives' chemistry and physics.

New Solvents

There are no drop-in replacements for the presently used solvents for electronic cleaning. Alternatives being synthesized and evaluated must possess the properties previously listed. The following is a partial list of new solvents being considered for electronic cleaning applications with examples.

- Hydrochlorofluorocarbons HCFCs - HCFC-123, HCFC-141b, HCFC-225 ca/cb
- Hydrofluorocarbons HFCs - DuPont 43-10, Allied-Signal to announce HFC soon
- Perfluorocarbons PFCs - 3M PF-5050, 3M PF-5060, DuPont Vertrel 245
- Ethyl lactate - ester compound not previously used as a cleaning solvent
- Volatile methyl siloxanes (VMS) - Dow Corning OS fluids
- Fluorinated alcohols - pentafluoropropanol

Old Solvents

Return to traditional solvents remains a possibility but the drawbacks and hazards that prompted the departure to new solvents such as CFCs must now be addressed with increased awareness. The following list of organic classes of compounds is presented as a review, for consideration and as examples.

- Hydrocarbons - aliphatic, acylic, aromatics
- Alcohols - methyl, ethyl, isopropyl
- Esters - ethyl acetate, butyl acetate
- Ketones - acetone, methyl ethyl ketone
- Chlorinated - methylene chloride, trichloroethylene, perchloroethylene

New Solvent Systems

For the purposes of this dissertation, the term solvent system refers to aqueous and semi-aqueous cleaning processes. The common literature and trade journals have unlimited references to both the aqueous and semi-aqueous processes. In an aqueous system a detergent in water is the processing ingredient. In essence it is a process involving a single cleaning step. In the semi-aqueous process an organic moiety presumably dissolves the soil and this is sequentially rinsed in water and ultimately dried in air. Since there are so many commercial systems available it is unnecessary to discuss this further.

New Processes/Methods

While the cleaning process is usually considered in terms of solubility, solutions and rinsing, cleaning surfaces can be accomplished by means other than the classical approaches to contaminant removal. Methods usually involve a more complex series of procedures to surface cleanliness and also include in some cases costly apparatus or equipment.

Ultraviolet Ozone/UV/03

A cleaning method that has excellent utility in removing small concentrations of organic contamination from surfaces is that of ultraviolet/ozone cleaning (UV/03). This method is particularly useful in cleaning inorganic substrates of organics. The mechanism by which UV/03 operates is in that the organic contamination is oxidized and subsequently volatilized.

Supercritical Fluid - SCF

Some gases when compressed and heated to near their critical points exhibit a phase change which is referred to as supercritical. Supercritical fluids have desirable properties of exhibiting near zero viscosity and have potential for dissolving selected organic contamination without ill effects to the substrate materials. The undesirable feature of SCF is that a complex pressure and temperature vessel must be employed.

CO_2 Pellet/Snow

CO_2 pellet/snow procedure has demonstrated utility in cleaning surfaces susceptible to damage by solvents and many other procedures. Spraying a stream of fine pellets or snow on a surface has been effective in cleaning quartz and glass surfaces. Since the pressure of impact can be controlled, the effect on the substrate can be minimized.

Plasma Cleaning/Glow Discharge

Plasma cleaning has been effective in removing organic contamination from electronic circuitry. It, however, has disadvantages that include complex equipment and potential for substrate destruction. Glow discharge cleaning is closely related to plasma cleaning in effectivity and hazards.

New Techniques

Complexity involved in the new alternative solvents and procedures has challenged the ingenuity of designers. No clean solder technology obviates the need for cleaning. Many no-clean solders are now available for commercial application. Since there is no cleaning involved the need for solvent or any other cleaning is eliminated.

Water is without a doubt the most common solvent known and widely used in commercial applications. An area that has been receiving much attention is that of water-soluble fluxes. No attempt will be made to discuss this further except to say that using water-soluble fluxes may be useful in many but not all cases.

Essential Use Exemptions to the Montreal Protocol

The Parties to the Montreal Protocol recognize that there may be some extenuating circumstances where no alternatives to ODSs may be available or acceptable. With this in mind a process exists where essential use exemptions may be petitioned. This is a complex and time consuming process and must comply at a minimum with the following requirements.

- It is necessary for the health, safety or is critical for the functioning of society (encompassing cultural and intellectual aspects).
- There are no available technically and economically feasible alternatives or substitutes that are acceptable from the standpoint of environment and health.

Summary

The phasing-out of production of substances that deplete the ozone layer have presented a complex challenge to technology. The substance of this paper is to present some preliminary information that will be useful in making a judicious alternative selection to an existing soon to be phased-out solvent or solvent system.

References and Suggested Readings

C. Lea, After CFCs?. Isle of Man, British Isles: Electrochemical Publications, 1992

K. L. Mitall, Treatise on Clean Surface Technology, vol. 1., New York: Plenum Press, 1987

_____, "CFC Replacement Critical Area Response", NASA Marshall Space Flight Center, 1994

J. R. Stemniski and R. L. King, "Ultraviolet/Ozone Cleaning", presented at Adhesives for Industry, El Segundo, CA, June 24-25, 1980

J. R. Stemniski, "Gyroscope Design and Manufacture to Avoid Ozone Depleting Substances", presented at International NATO/CCMS Conference, Brussels, Belgium, January 24, 1994

J. R. Stemniski, "Solvents and Alternatives Overview", presented at the Japan-Thai-US Trilateral Conference, Bangkok, Thailand, March 11-12, 1992

J. R. Stemniski, "The Cleaning Process - Choosing the Proper Solvent", presented at the Japan-Thai-US Trilateral Conference, Bangkok, Thailand, March 11-12, 1992

J. R. Stemniski, "Precision Cleaning", presented at the International Conference on CFC and Halon Alternatives, Beijing, China, April 20-23, 1993

_____, Protecting the Ozone Layer, Solvents, Coatings and Adhesives, UNEP/PAC, 75739 Paris Cedex 15, France

J. R. Vig, "UV/Ozone Cleaning of Surfaces", presented at the Fourth Annual International Workshop on Solvent Substitution, Phoenix, Arizona, December 7-10, 1993

_____, "Fourth Meeting of the Parties to the Montreal Protocol on Substances that Deplete the Ozone Layer", Copenhagen, Denmark, November 23-25, 1992

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

POLISHING AND ETCHING LANGASITE AND QUARTZ CRYSTALS

Sally Laffey*, Mary Hendrickson** and John R. Vig**

*Vitronics, Inc. and **U.S. Army Research Laboratory
Fort Monmouth, NJ 07703-5601, U.S.A.

Abstract

Langasite is a promising new piezoelectric material which is similar to quartz in its acoustic behavior. Little was known until recently about the methods needed to mechanically polish and chemically polish/etch this material. The results of polishing and etching experiments will be described. Conventional quartz polishing methods, with polishing agents such as a rare earth slurry or ruby powder, did not produce well polished surfaces. Polishing with a colloidal silica suspension has produced the best results. Polishing quartz with colloidal silica was also investigated. Colloidal silica was found to be capable of producing defect-free polished quartz surfaces as well.

Y-cut langasite surfaces which had been polished with the colloidal silica suspension were etched in a variety of etchants. The conventional quartz etchants destroyed the polished surfaces. Other etchants formed a coating on the surfaces. An etching process was found which was able to etch without leaving a film on the surfaces. When colloidal silica polished surfaces were etched deeply with this etchant subsequent to polishing, the etching did not degrade the polished surfaces. Similar results were obtained with polished quartz. After deep-etching colloidal silica polished quartz surfaces, the surfaces remained polished (except for the defects which were brought out by the etchant).

Introduction

The phase transition of quartz at 573°C limits the processing temperatures one can use in the fabrication of quartz resonators. It also prevents the use of quartz crystals in high temperature environments, e.g., as sensors. Langasite is a promising new piezoelectric material which is similar to quartz in its acoustic behavior, but unlike quartz, it has no phase transition - up to its melting point of 1470°C [1-3]. This may allow higher stabilities through higher-temperature processing, and may also allow sensors for high-temperature environments, such as deep oil and gas wells.

Langasite's acoustic attenuation has been reported to be three to five times lower than that of quartz [1-3]. This suggests that higher device Qs should be possible, however, the Qs of langasite resonators have been reported to be significantly lower than the Qs which can be achieved with quartz resonators of the same frequency. This discrepancy has been attributed to the lack good polishing methods for langasite [4]. Therefore, the experiments described below were undertaken with the primary goal of developing polishing and etching methods which are capable of producing defect-free langasite surfaces. A secondary goal was to test whether or not the method that produces defect-free langasite surfaces can also produce defect-free quartz surfaces.

Description of the Experiments

Polishing

The langasite polishing experiments were performed on Y-cut langasite wafers, each having an area of approximately 1.9 cm^2. The wafers were lapped on an Elgin Model 1-12 lapping machine, and the polishing was performed on an R. Howard Strasbaugh Model 6BQ polishing machine, at 60 RPM, with a swing of 65 CPM. The pressure during polishing was ~90 g/cm^2. Three langasite plates were lapped and then polished together. The polishing was done in an ordinary "dirty" environment (i.e., in a room without filtration of dust particles), and as a result, it was difficult to obtain completely scratch-free wafers.

The polishing processes attempted, and the results, are listed in Table I. (These initial experiments were aimed primarily at developing a process that works, rather than minimizing the polishing times.) Cerium oxide, which is widely used for polishing quartz, failed to polish langasite, as did 1 μm aluminum oxide. The best results, by far, were obtained with colloidal silica slurries.

Colloidal silica slurries from two different sources [5,6]

U.S. Government work not protected by U.S. copyright

Table I - Polishing experiment results.

Polishing Process	Results
Cerium oxide on pellon paper	Scratches, insufficient polishing
1 μm aluminum oxide on pellon paper	Scratches, insufficient polishing
1 μm aluminum oxide on polyurethane impregnated polyester felt (Suba™ 500 [5])	Scratches, insufficient polishing
Colloidal silica (Syton™ HT-50 [6]) on wool felt (Buehler AB Miramet [7])	Good polish
Colloidal silica (Syton HT-50) on polyurethane impregnated polyester felt	Excellent polish
Colloidal silica (Nalco 2360 [8]) on polyurethane impregnated polyester felt	Excellent results, with no edge rounding visible
Colloidal silica (Nalco 2360) on Politex Supreme pad [5]	Best results, with no edge rounding visible

were tried. Although both were capable of producing excellent polished surfaces, we were able to produce polished langasite faster with one than the other (and one was also more resistant to evaporation and crystallization than the other). Our evaluation of various slurry and polishing pad combinations is not yet complete, however, the processes that produced the best results to date consisted of an initial polish with Nalco 2360 [8] undiluted colloidal silica polishing slurry, followed by polishing with the same slurry diluted with 15 parts water, both with a Suba™ 500 polishing pad [5], or with the Suba™ 500 followed by Politex Supreme [5] for the 15:1 solution.

According to its supplier [5], the Suba 500 polishing pad has (Shore A) hardness of 72, compressibility of 7%, density of 0.35 g/cm^3, and a thickness of 1.27 mm (0.050"). The supplier recommended that the Suba 500 be used for "stock removal" and initial polish, followed by final polishing with their Politex Supreme polyurethane poromeric pad. We obtained a well-polished surface with the Suba 500 without any "final" polishing, however, final polishing with the Politex Supreme did improve the appearance of the polished surface somewhat. Polishing with the Politex Supreme and the 15:1 diluted slurry was about ten times slower than polishing with the full strength slurry and the Suba 500, i.e, 0.9 μm/h vs. 9 μm/h. Starting with a 3 μm lapped surface, the Nalco 2360 with Suba 500 process was capable of producing a well-polished surface in about 90 minutes per side.

Colloidal silica has been used for polishing silicon wafers since the 1960s [9]. Today, it is used for polishing a variety of materials, including both hard and soft materials, e.g., it can polish both sapphire and lead [10]. (Reference 10 is a review of the colloidal silica polishing literature, and reference 11 is a review of the silicon cutting, lapping and polishing literature - up to 1970.) Colloidal silica has also been used for polishing quartz wafers for surface acoustic wave devices [12]. Oxide and non-oxide ceramics have also been polished with this material [13].

Colloidal silica polishing slurries consist of discreet submicron amorphous silica particles dispersed in water, and, usually, some additives, e.g., for controlling pH. The particles are nearly spherical in shape. The particle sizes are very small, typically on the order of 50 nm. Figure 1 shows a transmission electron micrograph of one of the commercially available colloidal silicas, Syton HT-40 [6], and Figure 2 shows a histogram of particle size distributions for this material.

Etching

When quartz polishing with a cerium oxide slurry and pitch was compared to polishing with a colloidal silica slurry and a polyurethane foam pad, the quartz removal rates were found to be slower with the colloidal silica; e.g., at a 200 g/cm^2 load, the cerium oxide removed about 17 μm in 4 h, whereas the colloidal silica removed only about 8.5 μm [12]. As polishing with colloidal silica is a relatively slow process, we thought it to be highly desirable to find a langasite etchant which can remove the damaged layer produced by

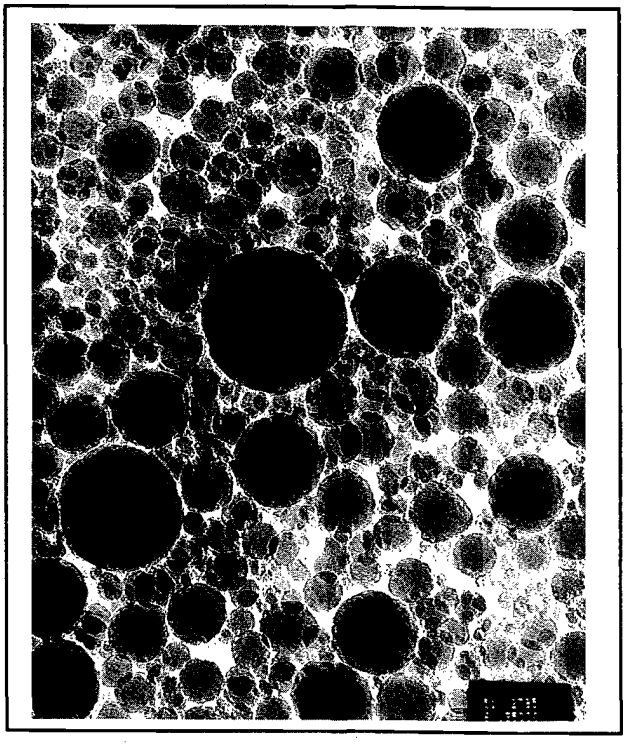


Figure 1 - Colloidal silica particles magnified 135,000X. Average particle size is 0.033 μm. From [10].

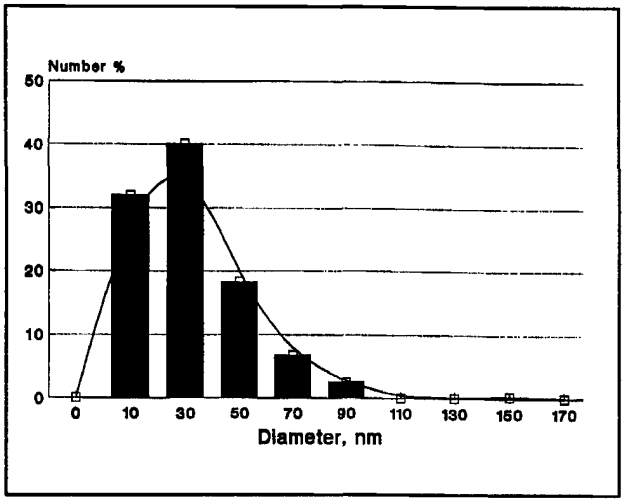


Figure 2 - Particle size distribution of the colloidal silica shown in Fig. 1. From [10].

lapping while leaving a chemically polished surface. Chemically polished surfaces are atomically smooth but undulating. For example, the mean surface roughness for a 3 μm aluminum oxide lapped and chemically polished AT-cut quartz plate is about 0.1 μm [14].

Starting with such a chemically polished surface instead of a 3 μm lapped surface can reduce the polishing time, and can also eliminate scratching by particles which break loose from the lapped surfaces.

A variety of etchants were investigated (empirically). Table II shows the etching attempts and the results.

The etchant we selected to use is a 1 : 50 : 150 solution of HCl : HF : H_2O, at 70°C, as this solution provides a reasonable etching rate, while being able to chemically polish langasite. Figure 3 shows the etching rate of langasite at laboratory ambient, 40°C and 70°C. Several

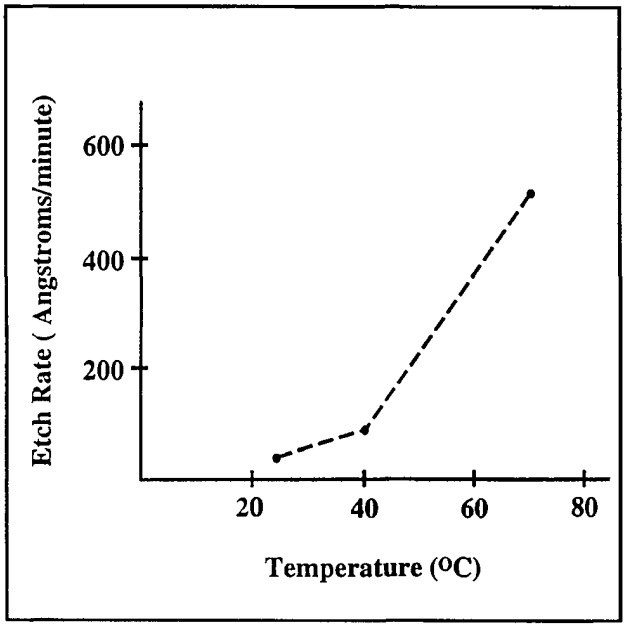


Figure 3 - The etching rates of langasite in HCl : HF : H_2O, 1 : 50 : 150, as a function of temperature.

of the other etchants tried produced a thin (~10 nm) film on the surfaces. Figure 4 shows an SEM micrograph of such a film. Auger electron spectroscopy revealed this film to consist of lanthanum and fluorine. The film could be etched away in dilute HCl, which is the reason for incorporating a small amount of HCl into the preferred etchant. Figure 5 shows an SEM micrograph of an etched surface. This surface was lapped with 3 μm aluminum oxide, polished for 4 hours with Syton HT-50 on a felt, and deep-etched in the preferred (1:50:150) etchant for 2 hours at 70°C. The plate was initially 0.77 mm thick. About 43 μm was removed with the etching. The surface roughness after etching, measured with an Alpha Step 200 profilometer [15] was 5 nm, which is the resolution of the profilometer.

Table II - Etching experiment results.

Etchant	Results
NH_4F : H_2O various ratios and temperatures	Film formation
HF : H_2O various ratios and temperatures	Film formation
HNO_3 : H_2O various ratios and temperatures	Film formation
HCl : H_2O various ratios and temperatures	High etching rates, rough surfaces
HF : HNO_3 : H_2O various ratios and temperatures	Film formation
HCl : HF : H_2O various ratios and temperatures	Acceptable etch rates, smooth surfaces, no film

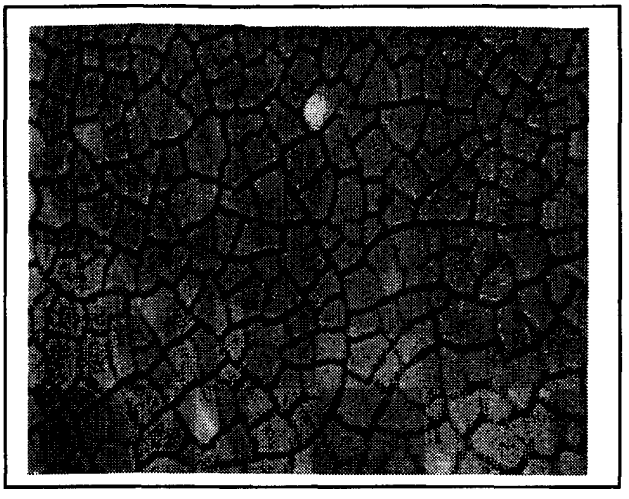


Figure 4 - SEM micrograph of film formed on langasite by etching (~500X magnification)..

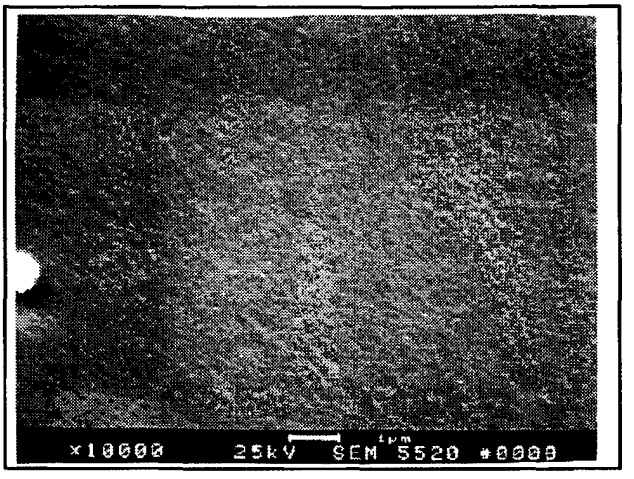


Figure 5 - SEM of colloidal silica polished langasite after 2 hours of etching at 70°C in 1: 50 : 150 solution of HCl : HF : H^2O (~10^4 X magnification).

Polishing Followed by Etching

Two polished surfaces may appear to be equally damage-free when examined under a microscope, however, one may possess subsurface damage, while the other may be truly damage-free. An effective way to separate a "good" polished surface from a "bad" one is to observe what happens upon etching. Etching can quickly reveal subsurface damage. The desirable polishing process is one which can produce a polished surface in a reasonable time, and moreover, when this polished surface is etched, no subsurface damage is revealed. The Nalco 2360 and Suba 500 combination can allow such a polishing process, as can the Syton HT-50 on Suba 500 or wool felt process.

Some langasite wafers were lapped to a 3 μm finish, cleaned, and etched for one hour. The pieces were polished using the Nalco 2360 on the Suba 500 pad for one hour. They were then polished with dilute (15:1) 2360 on a Politex Supreme polyurethane poromeric pad one more hour. The pieces were cleaned and etched in the 1:50:150 HCL:HF:H_2O at 70°C. One was etched for one minute, one for three minutes, one for one hour and another for two hours. The surfaces were compared with an unetched control sample. Although some flaws were enlarged during the etching, the surfaces of all the samples stayed polished, i.e. no fog or haze formation was noted. Using the Alpha step profilometer, the surface roughness of the unetched sample was 15 nm. The etched pieces, including the one etched for one hour and the one etched for two hours, all measured 5 to 10 nm, which is close to the resolution of the profilometer.

Polishing Quartz with Colloidal Silica

Several AT-cut quartz plates were processed by lapping on cast iron to a 3 μm finish and etching in saturated ammonium bifluoride for 20 minutes at 70°C. This etching removed the mechanical damage caused by the lapping. The pieces were again mounted and polished on a Suba 500 pad for one hour and then final-polished on a Politex Supreme pad for an additional hour using Nalco 2360, first full strength then diluted with 15 parts of water. The resulting polish was a bright clean surface free of haze.

The samples were then etched again in saturated ammonium bifluoride at 70°C: one for 3 minutes and one for 15 minutes. Under the microscope there was no discernable difference between the etched samples and the unetched polished sample, except that flaws were magnified by the etching. The profilometer indicated no degradation of surface roughness by the etching; the surface roughnesses were ~10 nm for all three samples.

Discussion

There is strong evidence that the process of polishing quartz (or silicon) with colloidal silica (or with cerium oxide) differs significantly from the process of lapping with a hard material such as aluminum oxide or diamond [10,11]. Lapping is primarily a mechanical process of abrasion, whereas polishing is a result of combined chemical and mechanical action. When diamond and aluminum oxide were used to "polish" silicon, both produced a damaged surface, and diamond produced more severely damaged surfaces for the same particle size [11]. Colloidal silica has been found to be capable of producing defect-free silicon surfaces. Colloidal silica particles are very small, nearly spherical, and softer than quartz, silicon or sapphire, yet colloidal silica can polish all three. The polishing rates depend on pH. When either oil or ethanol is substituted for water, poor polishing results. Water also plays an essential role in the polishing of glasses [16,17].

Based on research done on polishing glasses and silicon, the chemistry of polishing quartz is complicated (as langasite polishing probably is also). A simple description of the quartz polishing process may be that water, or the combined action of water and the colloidal silica particles, continuously react with the surface to form a soft, hydrated, silica-gel-like material, and the combined action of the colloidal silica slurry and the polishing pad continuously remove this material. The underlying crystal, which is much harder than the colloidal silica, is not abraded or damaged by the process.

Many possibilities of polishing with silica slurries remain to be explored. A variety of slurries and polishing pads are available from the semiconductor polishing suppliers. Silicas are available in a range of particle diameters, and at least three types of silica slurries are available: colloidal, precipitated and fumed silicas [5]. To date, we have tried only the colloidal silicas.

Summary and Conclusions

Colloidal silica has been shown to be able to polish both langasite and quartz without producing a subsurface damaged layer. Etching the polished surfaces with chemical polishing etchants produced no degradation of the polished surfaces (except that the etching does reveal defects - both scratches and crystallographic defects). A good etchant for langasite is a solution of 1 part HCl to 50 parts HF to 150 parts H_2O, at 70°C.

That etching does not degrade the polished surfaces (except for revealing flaws in the material) suggests that polishing resonator plates with colloidal silica followed by deep-etching can allow the micromachining of quartz and langasite structures [18], including the production of very thin and very small resonator plates, provided that the plates are made of sufficiently defect-free materials.

Acknowledgements

The authors wish to thank S. A. Sakharov of Firm FOMOS and A. N. Gotalskaya of LANTAN, for providing langasite samples, and George F. Vander Voort of Carpenter Technology for providing and giving permission to reproduce the transmission electron micrograph of colloidal silica particles, and of the particle size distribution curve.

We also thank Timothy Monahan, Donald Eckart and Luis Casas for their expert assistance with SEM, AES and SIMS analyses, and we also thank Donald Eckart for the photography.

References

1) V. B. Grouzinenko and V. V. Bezdelkin, "Piezoelectric Resonators from $La_3Ga_5SiO_{14}$ (Langasite) - Single Crystals,"Proc. 1992 IEEE Frequency Control Symposium," pp. 707-712, 1992.

2) S. A. Sakharov, I. M. Larionov, and A. V. Medvedev, "Application of Langasite Crystals in Monolithic Filters Operating in Shear Modes," 1992 IEEE Frequency Control Symposium," pp. 713-723, 1992.

3) I. M. Silvestrova, Yu. V. Pisarevsky, V. V. Bezdelkin and P. A. Senyushenkov, "New Piezoelectric Materials," Proc. 1993 IEEE Int'l Frequency Control Symposium, pp. 351-352, 1993.

4) Private communication, Yu. V. Pisarevsky and S. A. Sakharov, October 1993.

5) Rodel, 9495 East San Salvador Drive, Scottsdale, AZ 85258.

6) Remet Chemical Corp., P. O. Box 278, Chadwicks, NY 13319.

7) Buehler, LTD 41 Waukegan Rd., Lake Bluff, Ill. 60044

8) Nalco Chemical Co., One Nalco Center, Naperville, IL 60566-1024; distributed by Rodel, 9495 East San Salvador Drive, Scottsdale, AZ 85258.

9) R. J. Walsh and A. H. Herzog, "Process for Polishing Semiconductor Materials," U. S. Pat. No. 3,170,273, Feb. 23, 1965.

10) G. F. Vander Voort, "Polishing with Colloidal Silica," Structure - Struers Journal of Materialography, pp. 3-11, Feb., 1992.

11) T. M. Buck and R. M. Meek, "Crystallographic Damage to Silicon by Typical Slicing, Lapping, and Polishing Operations," in Silicon Device Processing, C. P. Marsden, ed., National Bureau of Standards Special Publ. No. 337, pp. 419-430, 1970.

12) G. W. Fynn and W.J.A. Powell, The Cutting and Polishing of Electro-optic Materials, John Wiley & Sons, New York, pp. 178-184, 1979.

13) H. Vora, "Study of Mechano-Chemical Machining of Ceramics and the Effect of Thin Film Behavior," Technical Report 1, February 1984, AD-A142 542.

14) J. R. Vig, J. W. LeBus, and R. L. Filler, "Chemically Polished Quartz," Proc. 31st Ann. Symp. on Frequency Control, pp. 131-143, 1977.

15) Tencor Instruments, 2400 Charleston Rd., Mountainview, California 94043.

16) N. J. Brown and B. A. Fuchs, "Shear Mode Grinding," Proc. 43rd Ann. Symp. on Frequency Control, pp. 606-610, 1989.

17) L. M. Cook, "Chemical Processes in Glass Polishing," J. of Non-Crystalline Solids, 120, pp. 152-171, 1990.

18) R. T. Howe, Applications of Silicon Micromachining to Resonator Fabrication," Proc. 1994 IEEE Freq. Ctrl. Symp., 1994.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Submicron Fabrication Technology Using Electro-Chemical Effects and Application to GHz-range Unidirectional SAW Transducers

K.YAMANOUCHI, H.ODAGAWA, T.MEGURO, Y.WAGATSUMA and K.YAMAMOTO

Research Institute of Electrical Communication, Tohoku University,
Sendai 980, Japan

ABSTRACT

The frequency ranges of mobile communication systems are now in 1 GHz and in extending to 2～4 GHz. Moreover SAW devices require the frequency range of around 10 GHz. The submicron technology below 0.1μm is required. In this paper, we propose a new lithography technique that can make below 0.05μm electrode width. The electrodes are fabricated by using very thin anodic oxidation films as resists and wet or dry etching techniques. Also, new lithography technologies using electro-chemical effects are described in the fabrication process of the unidirectional SAW IDT with a different thickness and a material electrode using a single photo-lithography method. The results show the 0.075μm width IDT with thickness about 0.03μm on 128° Y-X $LiNbO_3$.SAW experimental results show 17 GHz and 10 GH_z-range SAW conventional interdigital transducers , filters and the 1 GHz low loss filters with the insertion loss around 5 dB.

1. Introduction

Submicron lithography technology is very important for high density integrated circuits and higher frequency-range SAW devices.

Conventional etching methods have some problems for submicron fabrication as follows.

(1) They require a thick resist to protect the metal films against an etchant, though a thin resist is much better for narrow lithography.

(2) They have the problem that the etchant soaks into the contact of the resist and the metal.

On the other hand, the new method in this paper has no such problems[1]. The key technique of this method is to change a thin photo resist or E-beam resist pattern into an oxidized resist pattern by anodic oxidation before etching.

We describe the etching process using anodic oxidation for submicron fabrication, some experimental results and applications to high frequency SAW devices. New lithography technologies using electro-chemical effects are also described in the fabrication process of the unidirectional SAW IDT with a different thickness and a material electrode using a single photo-lithography method.

2. Etching process using anodic oxidation resist - I

The thickness of photo or E-beam resist is very important, especially in the fine lithography. It is difficult to use thick resists in order to obtain a fine resist pattern because of scatterings and reflections of beams. However thin resists cannot be used for etching process because the thin resist film cannot protect metal films toward etchant. In contrast, the new etching process using anodic oxidation resist can use the thin resist. Because the photo or E-beam resist, which is weak for most of etchant, is changed into an oxidized film by anodic oxidation before etching. A neutral or non-aqueous solution can be used as the electrolyte of anodic oxidation so that thin resist, less than 50 nm of PMMA, shuld be enough to protect the metals against the electrolytes. In addition, this method has a feature that it realizes an ideal contact of the resist and a metal film. Figure 1 shows one of the process of the etching method using the anodic oxidation resist.

(1) Al film is deposited on a substrate and photo or E-beam resist is coated.
(2) Exposure and developmet processes are done and the resist pattern is obtained.
(3) Anodic oxidation is done, so that only the surface of Al without resist is changed into Al_2O_3. In this case, the thickness of Al_2O_3 is less than 5nm.
(4) Resists are removed and Al_2O_3 resist pattern is obtained.
(5) Dry or wet etching is done, so that Al electrodes are fabricated.

In neutral or non-equious solution the thickness of Al_2O_3 film is proportional to applied voltage and easily controlled; the thickness of 1.4 nm/V has been obtained, and anodic oxidation can be operated by a simple equipment [2].

Figure 2 shows an etching rate of Al and Al_2O_3 films in Al etchant based on the phosphoric acid. Since the ratio of Al to Al_2O_3 is more than 100, Al_2O_3 can be used as the resist of the etchant. This Al_2O_3 is an anodic oxidation film made in ammonium tetraborate solution in ethiren glycol.

Experimental results in this method are shown in Fig.3(a),(b). They are SEM image of an E-beam resist pattern before anodic oxidation and an electrode pattern after etching, respectively. They show that 200 nm periodic electrodes (they correspond to the width of 100nm) are obtained. The conditions are as follows; resist is PMMA of 35nm, Al thickness is 30nm, Al_2O_3 thickness is about 3nm (oxidation voltage 3V), oxidation time is 300s.

0-7803-1945-1/94 $3.00 © 1994 IEEE

3. Etching process using anodic oxidation resist - Ⅱ

Figure 4 shows another process using the anodic oxidation resist.

(1) After deposition of an Al film on a substrate, the surface is anodic oxidized less than 5 nm below the surface and photo or E-beam resist is coated.
(2) Exposure and development processes are done and the resist pattern is obtained. After that, an ion beam shower hits against the Al_2O_3.
(3) As a result, only the Al_2O_3 that is not covered by a resist is removed.
(4) Photo or E-beam resist is removed and Al_2O_3 resist patern is obtained.
(5) Dry or wet etching is done, so that Al electrodes shoud be fabricated.

In the second process, even if the contact between the photo or E-beam resist and the metal is not perfect, fine electrodes can be obtained.

A SEM image of electrode pattern by using the second method is shown in Fig.5. It shows that 150 nm periodic electrodes (they correspond to the width of 75nm) are obtained. The conditions are the almost same as the case of Fig.2. In this case, argon ion shower of 30 seconds is used.

4. Application to 17 GHz-range SAW filter

Figure 6 shows 10 GHz range SAW filter characteristics made by the first method on 128° Y-X $LiNbO_3$. IDT periode is 100nm with the pair number of 20 and aperture of 30λ. Since the filter is not in matching condition with bi-directional loss of 6dB, insertion loss of 15.7 dB should not be large in 10 GHz-range.

Figure 7 shows the 11.7GHz SAW propagation on 128° Y-X $LiNbO_3$. The IDT period is 80nm. The insertion loss is about 31dB. It shows that the coherent SAW generation, and the propagation at 11.7 GHz is realized. The IDTs are not perfect in this case. If we obtain the perfect electrodes, the insertion loss should be decreased.

Figure 8 shows the 17GHz SAW filter on 128°Y-X $LiNbO_3$. The insertion loss is about 60dB.

5. New Lithography Techniques Using Electrochemical Effects

The Interdigital Transducer with different thicknesses or different materials of submicron electrodes can easily be fabricated by using the electrochemical effects by only one photo-process. Figure 9 shows the fabrication process using the electro-chemical effect. In the process-(a), the Cr-thin films with an applied voltage are removed by the etching process. On the otther hand, in the process-(b), the Cr-thin films without an applied voltage are removed. Al and Cr-metals are evaporated on 128° Y-X $LiNbO_3$ and coated with the resist. The exposure and development processes are done and the resist patterns are obtained. Cr/Al thin films electrodes are obtaind by etching process. One of the electrodes is connected to the electrical DC-source and applied a few volts. The etching processes are performed within ammonium tetraborate solution in ethylene glycol or cerium diamonium nitrorate with perchloric acid. Finally the elletrodes with Al and Cr/Al metal films are obtained.

Figure 10 shows one of the experimental results. The IDT with Al and Cr/Al electrodes are obtained.

6. Application to SAW Unidirectional Transducers of Bi-metal Elecrodes

The stracture of the $\lambda/6$ type unidirectional transducer with the Al and Cr/Al electrodes is shown in Fig 9(b). The directivity is caused by a different reflectivity. The frequency response of the filter is shown in Fig.11. The directivity with 10 dB is obtained. The minimun insertion loss of about 5dB at 1.0 GHz with matching condition is also obtained.

7. Propagation loss of GHz-range SAW

In SAW device application, the propagation loss is very important. The loss for isotropic materials is given as follows,

$$\alpha s = \alpha \ell + \alpha t(1-A)$$

where αs:SAW loss , $\alpha \ell$:longitudinal wave loss, αt:transverse wave loss, and A: Poisson ratio.

The propagation loss for SAW is caused by the scattering loss of thermal phonon, imperfection of single crystals, surface roughness and surface defects due to mechanical polishing. Figure 12 shows the propagation loss of SAW at GHz-range for 128° Y-X $LiNbO_3$ at the room temperature[3]. We can see that the loss at 5GHz is about $0.56dB/100\lambda$. The propagation loss is not serious for the application of GHz-range SAW devices, because it is not so large. Also, the direct through signals can be suppressed to over 50dB under the condition of the distance of 100λ between input and output transducers. The insertion loss for GHz-range SAW devices therefore depends on the IDT loss.

8.Conclusion

The etching processes using anodic oxidation resist has been presented. It is very useful for fabrication of fine electrode such as high frequency low loss SAW IDTs. The periodic electrodes with the width of 75 nm have been obtained by this method. 10GHz and 17GHz SAW coherent propagation have been obtained.New Etching Process Using Electro-chemical Effects is proposed and applied to unidirectional SAW transducers and good results are obtained. We are now trying fabrication of the narrower width electrode than that of 0.05 μm and above 20 GHz-range SAW generation-detection. We are also making a unidirectional low loss filter in 10GHz-range.

Reference

[1] K.Yamanouchi,"Nano-meter Fabrication Processing technology and GHz-range Unidirectional low-loss SAW Filters", Proc.Int.SAW Devices for Mobile Comm., Dec.-1992,pp.123-133, Edited by K.Shibayama and K.Yamanouchi.

[2] K.Yamanouchi, T.Meguro and K.Matsumoto, "Surface Acoustic Wave Unidirectional Transducers Using Anodic Oxidation Thechnology and Low-loss Filters", Electron. Lett., vol.25, No.15, pp.958-960, 1989.

[3] K.Yamanouchi,Y.Cho and T.Meguro,"SHF-range surface acoustic wave interdigital transducers using electron beam exposure", Proc. of IEEE Ultrason.Symp.,p.115,(1988)

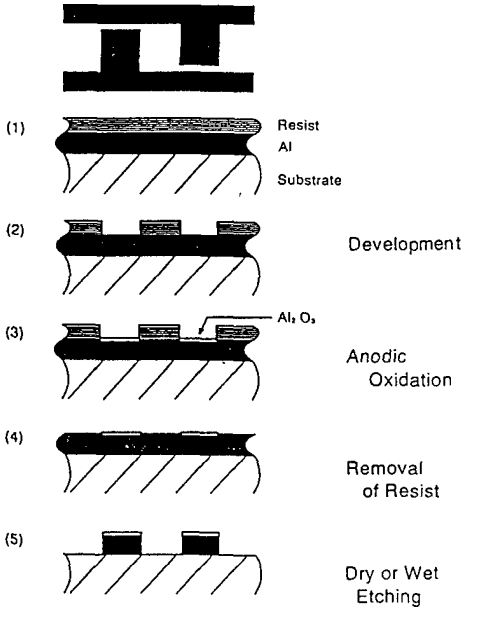


Fig. 1. Fabrication process of fine electrode using anodic oxidation resist 1.

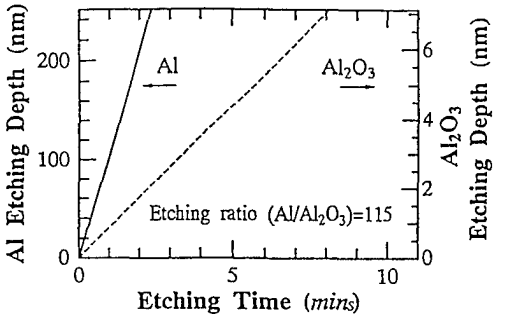


Fig. 2. Etching rates of Al and Al_2O_3 films.

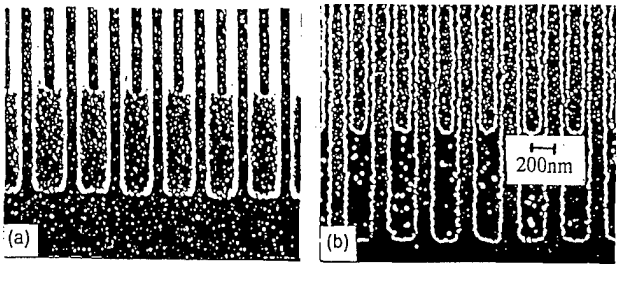


Fig. 3. SEM images of resist pattern and electrodes. (a) PMMA resist pattern, (b) Electrodes.

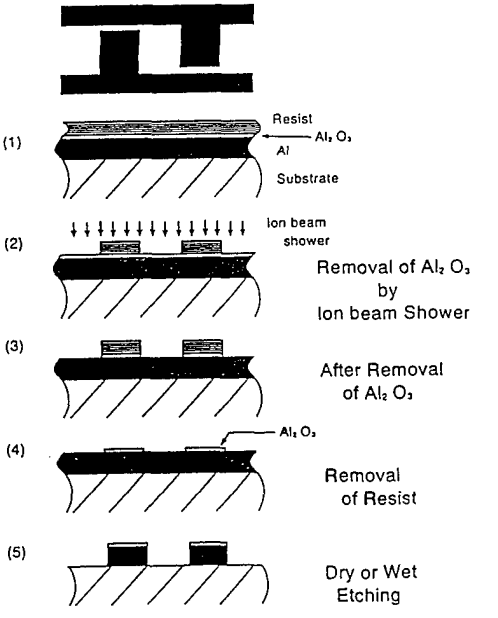


Fig. 4. Fabrication process of fine electrode using anodic oxidation resist 2.

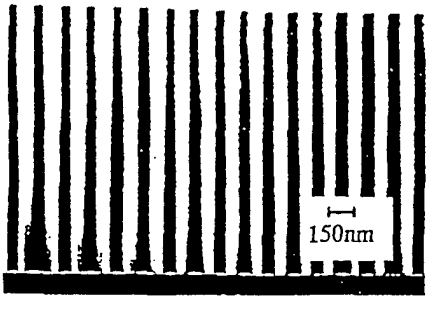


Fig. 5. SEM images of electrodes.

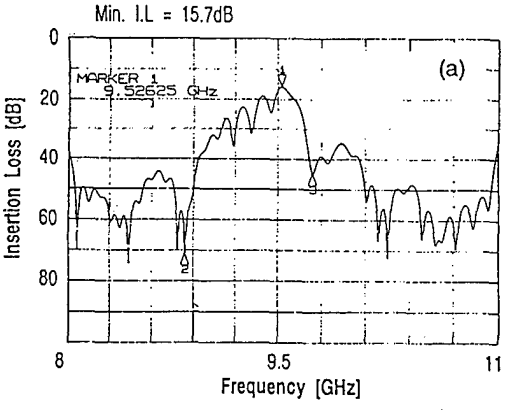


Fig. 6. Experimental frequency response of 9.5 GHz SAW filter

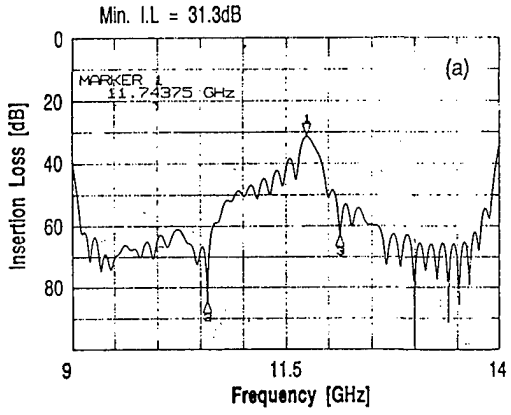


Fig. 7. Frequency response of 11.7GHz SAW filter

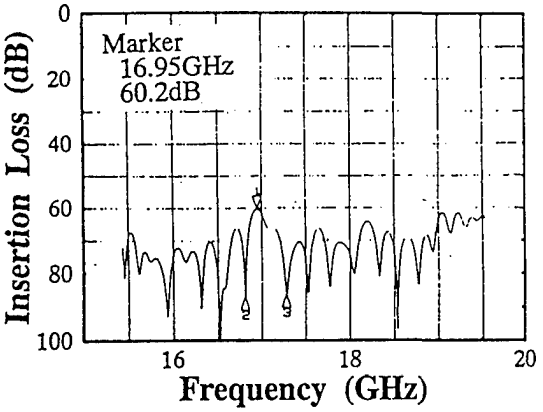


Fig. 8. Frequency response of 17 GHz SAW filter.

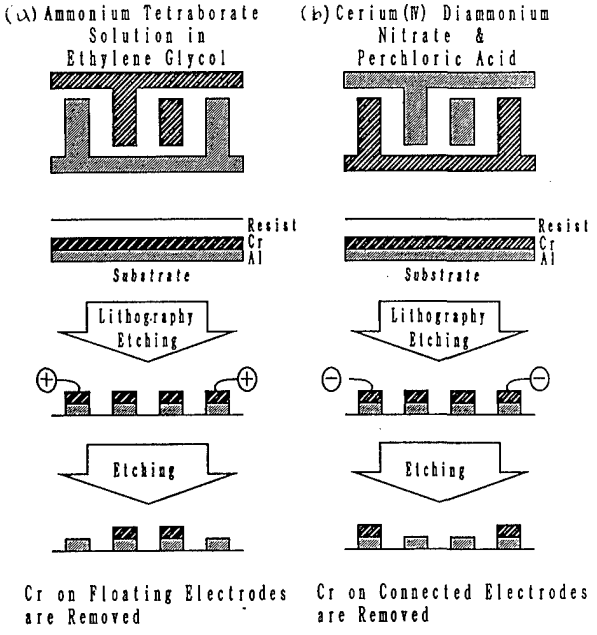
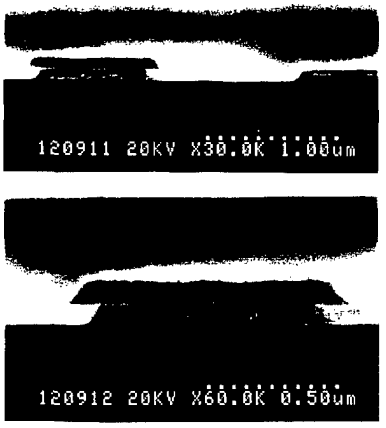


Fig. 9. New etching process using electro-chemical effect.



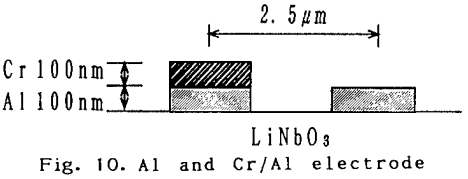


Fig. 10. Al and Cr/Al electrode pattern

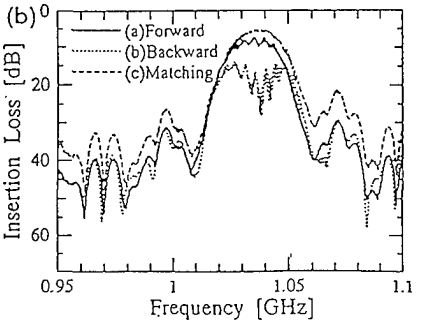


Fig. 11. Frequency response of UDT with different material electrode

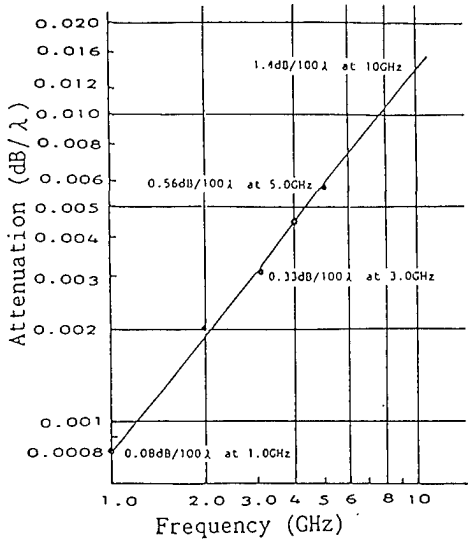


Fig. 12. GHz-range propagation attenuation of 128° Y-X $LiNbO_3$ at a room temperature

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

MEASUREMENT OF TEMPERATURE CHARACTERISTICS OF TWO-DIMENSIONAL SURFACE CHARGE PATTERNS OF SPURIOUS MODES IN RECTANGULAR QUARTZ PLATES

Yasuaki Watanabe, Hitoshi Sekimoto and Yoshimasa Oomura

Faculty of Technology, Tokyo Metropolitan University
Minami-osawa 1-1, Hachioji-city, Tokyo 192-03 Japan

Abstract

Temperature characteristics of surface charge patterns of spurious vibrations near the fundamental modes in rectangular quartz plates have been measured. The technique is based on the electrical probe method that uses piezoelectrically induced surface charge. To measure the temperature characteristics of surface charge patterns, a measurement fixture with a DUT mover is set in a computer-controlled temperature chamber and these instruments are incorporated with a network analyzer based measurement system. The measurement results using 1 and 4MHz rectangular AT-cut quartz resonators clearly indicate that the charge distributions of a spurious vibration and a fundamental one interchange their patterns according to temperature changes.

I. Introduction

A number of methods of plotting the vibrational distribution of quartz resonators have been developed[1]. The most well known method is Lang topography using X-rays[2]. Recent technological advances in the fabrication and processing of Laser and optical devices have been made the methods using interferometry possible to measure the spurious and fundamental vibration of high frequency resonators[3].

On the other hand, it is indispensable for designing and fabricating quartz resonators to avoid spurious vibrations near the fundamental resonant frequency.

The nearby spurious vibrations frequently couple with the fundamental resonance according to the resonator environment such as temperature changes, and exercise a bad influence upon the characteristics of the fundamental mode, even if those spurious resonances are electrically weak. It is easily understood that both of the modes will interchange their vibrational shapes when resonant frequency of the spurious vibration goes across the fundamental mode according to temperature changes. The above methods, however, may not be able to apply to measure this phenomenon because they need very large and complicated apparatus. Thus experimental papers on such a physically attractive phenomenon, to the authors' knowledge, have never been published. In contrast with the methods, the electrical probe method[4] developed by Koga and Fukuyo, consists of quite simple apparatus and it can be apply to measure the temperature characteristics of the mode shapes. This draws piezoelectrically excited surface charge distribution instead of the vibrational shape. We have already developed an automated measurement system incorporated with this method and shown that the charge patterns measured by the system have enough information to distinguish the resonant modes of spurious vibrations[5].

In this paper, it is described that the system has been applied to measure the temperature characteristics of mode shapes of spurious vibrations going across the fundamental thickness-shear mode in miniature rectangular AT-cut quartz plates. The principle of the electrical probe method and the system configuration are described in Section 2. The measurement software including data

0-7803-1945-1/94 $3.00 © 1994 IEEE

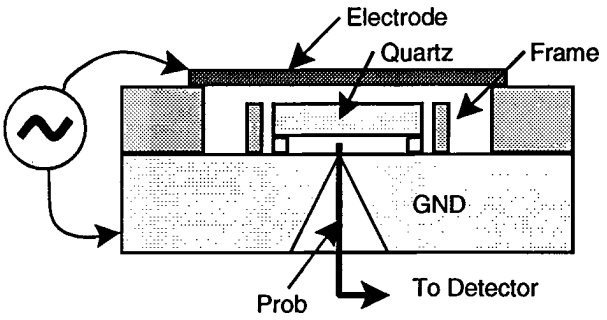


Fig.1 Schematic diagram of electrical probe method.

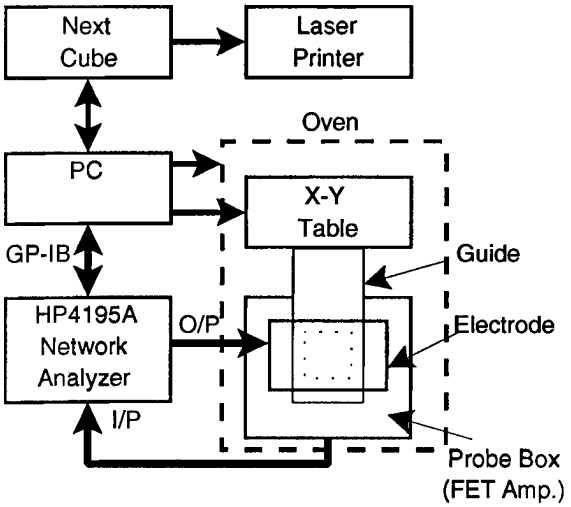


Fig.2 Surface charge distribution measurement setup.

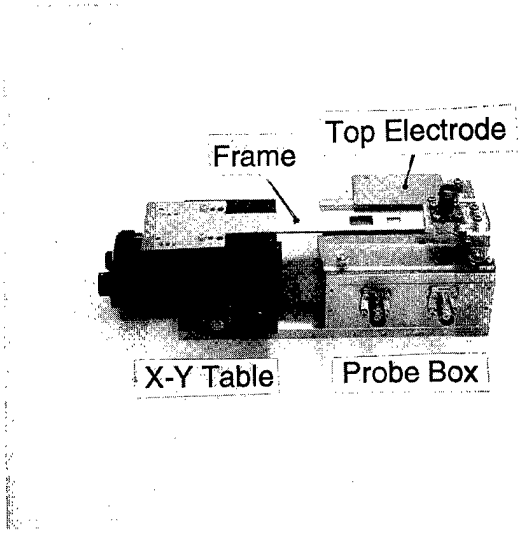


Fig.3 Probe box and D.U.T. Mover

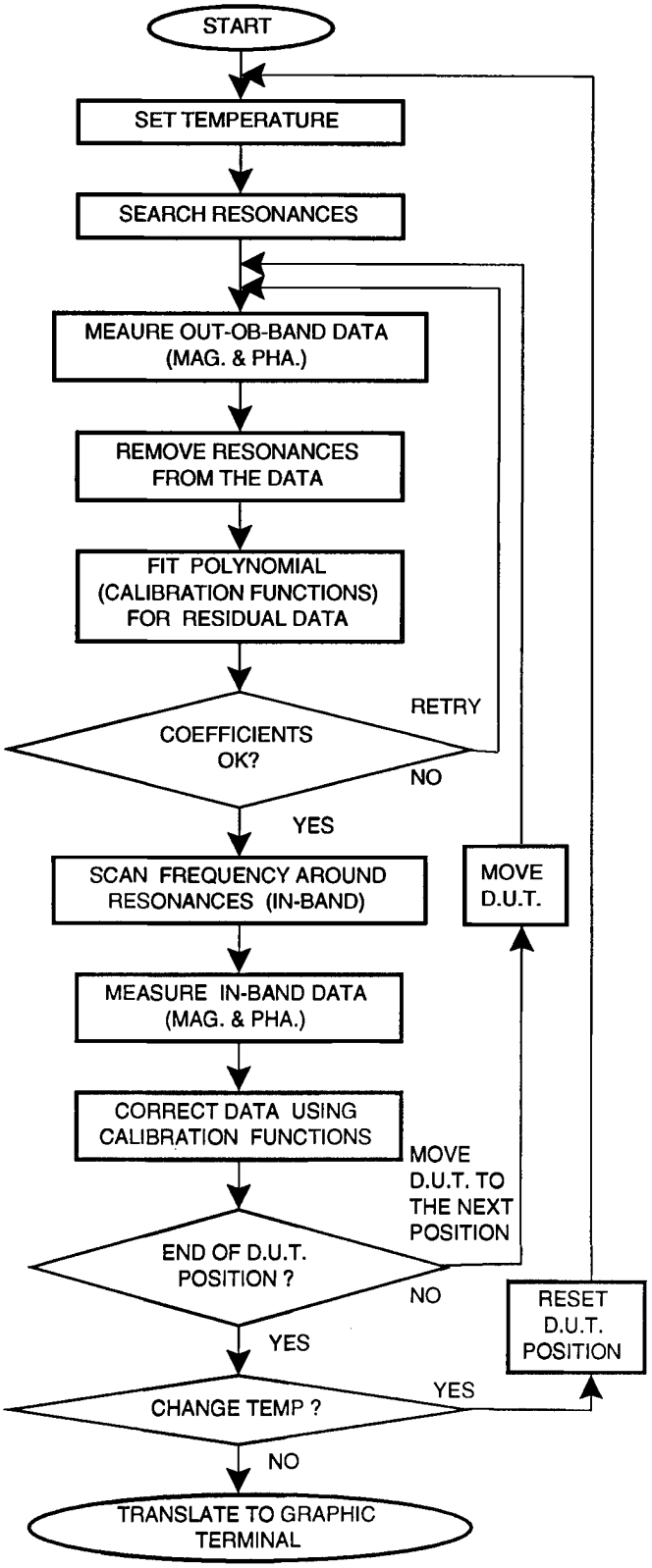


Fig.4 Flowchart for measurement and calibration.

correction and processing is described in section 3. In succeeding sections the experimental results using 1 and 4MHz rectangular AT-cut quartz plates are given.

II. Measurement system

the electrical probe method, originated with Koga and Fukuyo[4], in that a small electrical probe protrudes through an earth plane, both probe and plane being close to the surface of crystal. In this method the electrode opposite the probe is connected to a frequency source that drives the resonator, and the probe is connected to an RF voltmeter. The surface charge distribution can be obtained by mapping the detected voltage for each surface coordinate of the resonator.

Fig. 2 shows a block diagram of the measurement setup. The electrical prober mentioned above is incorporated in this system as a probe box. The needle has a diameter of 100 μ m. A computer-controlled X-Y table moves a frame containing a resonator under test. Fig. 3 shows a photograph of the electrical prober module that is taken off the top electrode. An FET amplifier is used to bring up the signal levels and to convert the impedance to 50 ohms. The prober module shown in Fig. 3 is placed in an oven. In this system, the HP4195A network analyzer was selected for actual measurement of amplitude and phase of surface charge.

III. Measurement software

oven to control temperature, an x-y stage to position a D.U.T., and a network analyzer to measure frequency characteristics of amplitude and phase of device surface charge. Following is a description of the measurement procedure.

A. Temperature Control

Temperature inside the oven is kept in $\pm$ 0.1 ℃ to directed temperature. After sufficient soaking time, at least one hour, following processes are carried out.

B. Device Positioning

First, a DUT is positioned by a frame joined to an X-Y table. The position of the DUT is adjusted within an accuracy of ± 1 μ m.

C. Out-of-Band Measurement

Device surface charge (voltage and phase) is roughly measured at wide frequency span including desired resonances. Data of fundamental and some spurious resonance regions are removed from the whole data. Correction functions expressed in polynomials are calculated using the residual data to offset measurement errors caused by the effects of resonator static capacitance and parasitic reactance, etc.

D. In-Band Measurement and Data Correction

Because influence of the DUT-movement must be avoided, it is necessary to scan the vicinity of each desired resonance. After scanning and finding a peak voltage, data corrections are carried out by deducting the above functions from raw voltage and phase data. In this way, residual "net" surface charge with its phase can be obtained. A flowchart of the measurement software is shown in Fig. 4.

E. Asymmetry Removing

In theoretical calculation of an AT-cut quartz plate, surface charge asymmetry is generally ignored because this does not give influence upon its resonant characteristics. For correspondence to the calculation, we remove the asymmetric components from the corrected data by superimposing the charge distribution map reversed with respect to each the X and Z' axis over the original pattern.

IV. Experimental Results

In this experiment, two rectangular AT-cut quartz plates are selected as DUTs. Table 1 shows the dimensions of plates. These dimensions are determined as a spurious mode having bad temperature characteristics comes close to the fundamental mode at room temperature because of the ease of measurement.

Fig. 5 shows the experimental results for the 1MHz plate. Dots mean resonant

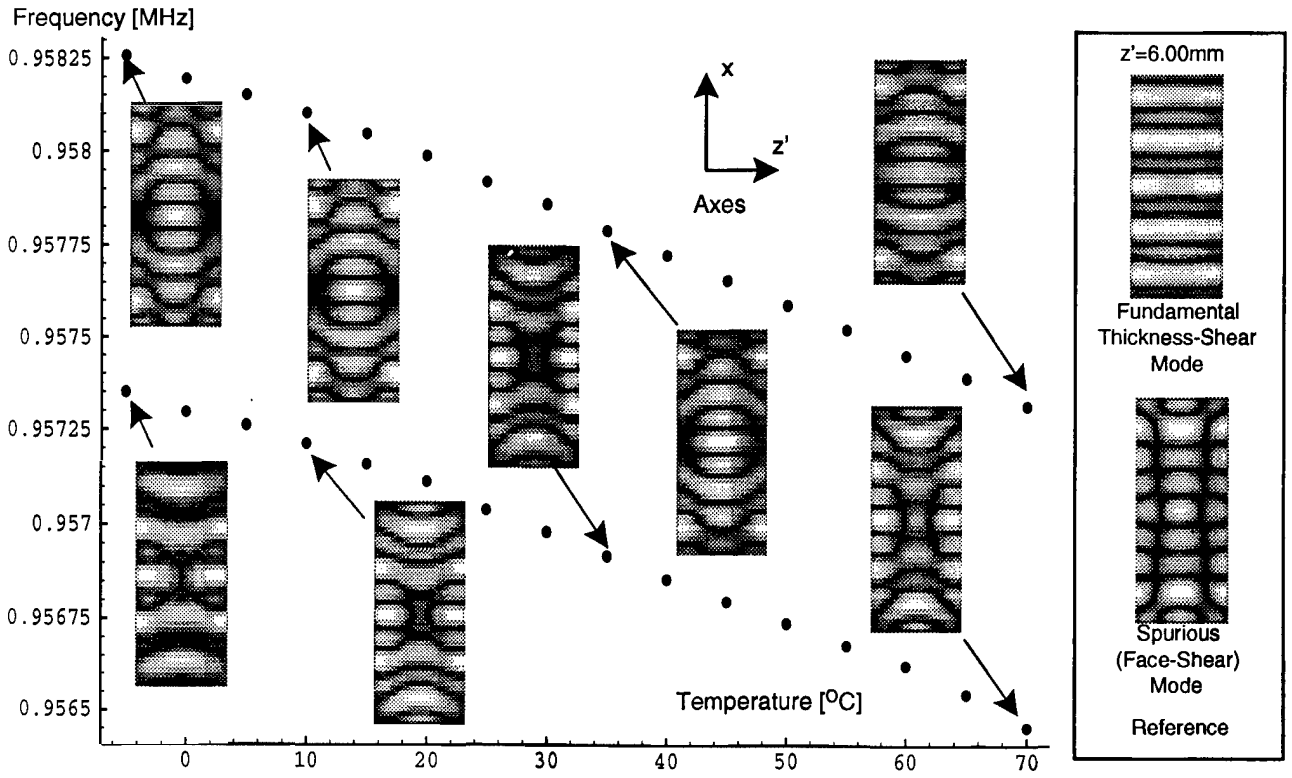


Fig.5 Experimental results for a 1MHz AT-cut quartz plate. (2a(x)=13.964mm, 2c(z')=5.92mm, 2b(y')=1.737mm)

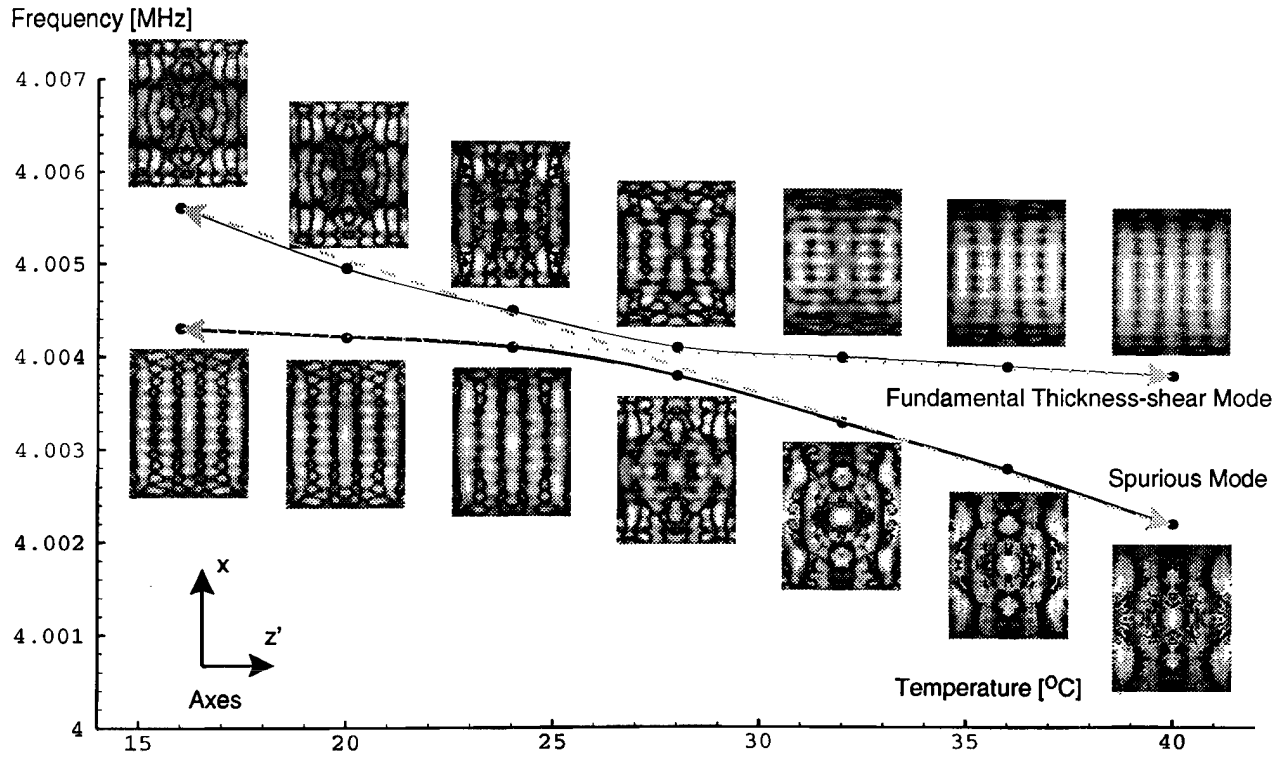


Fig.6 Experimental results for a 4MHz AT-cut quartz plate. (2a(x)=8.35mm, 2c(z')=7.00mm, 2b(y')=0.422mm)

frequencies of each mode. Surface charge patterns inset the figure were represented by absolute value. Measurement area including a DUT was divided into 50 x 50 portions. Diagrams put in the right side of this scheme, show results of the other DUT of the same dimension except Z' width (Z'=6mm), where the two modes are discrete. It is clearly understood from this figure that the fundamental thickness-shear mode and spurious (face shear) one gradually interchange their surface charge patterns according to temperature changes. At -5 ℃, a pattern of the lower frequency mode is similar to a fundamental mode pattern in the top right-hand corner of the figure. At 70 ℃, the lower pattern presents phase reversal along z' direction and is similar to a spurious mode in the right side of the figure. In this case, however, the coupling between the two modes is very strong and their patterns are not perfectly discrete in present temperature range.

Fig.6 shows the results for the 4MHz plate. Measurement area was divided into 80 x 80 portions. The figure also clearly indicates that the charge distributions of a spurious vibration and a fundamental one interchange their patterns at 28-32 ℃. Compared with previous case, the coupling between a fundamental mode and a spurious one is weak, therefore each charge pattern is perfectly discrete at low or high temperature regions.

V. Conclusions

In this paper, it is described that an automated surface charge measurement system has been applied to measure the temperature characteristics of mode shapes of spurious vibrations going across the fundamental thickness-shear mode in rectangular AT-cut quartz plates.

The measurement results using 1 and 4MHz rectangular AT-cut quartz resonators visually demonstrate that the mode shapes of a fundamental vibration and a spurious one interchange their patterns according to temperature changes.

The results obtained here contain the data of couplings and temperature characteristics of spurious vibrations in rectangular quartz plates; therefore the data will give many suggestions to consider approaches to the future numerical analyses.

References

[1] R.J.Williamson, Proc. of 44th FCS, pp424-428, 1990
[2] W.J.Spencer, Ch3,Vol.V, Physical Acoust.,Academic Press 1968.
[3] e.g. Y.Tsuzuki et al., Proc. of 43rd FCS, p113, 1989
[4] I.Koga and H.Fukuyo, Trans. on IECE Japan, 36, 1953
[5] Y.Watanabe et al., Proc. of 1993 IEEE UFFC Symp. pp571-574, 1993

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ORIGIN AND MEASUREMENT OF QUARTZ RESONATOR MAGNETIC SENSITIVITY

R. Brendel, B. Cretin, A. El Habti

Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon - 32 avenue de l'Observatoire - 25000 Besançon - France

Abstract

The magnetic sensitivity of quartz crystal resonators is a consequence of the ferromagnetic properties of the metal used as support for the vibrating plate. Various magneto-mechanic interactions can contribute to the overall sensitivity, the most important of them is shown to be magnetostriction: when submitted to a magnetic field, most of the ferromagnetic materials undergo an expansion or contraction wich induces a stress in the sustained plate and then a change in the crystal resonant frequency. The striking likeness between nickel magnetostriction curve and the frequency-vs-magnetic field curve of resonators is at the origin of this assumption. Other experiments have been performed to check on this assumption, in particular an experimental setup using a heterodyne interferometric laser probe with enhanced sensitivity has been used to investigate *in situ* the magneto-striction of the plate supports. Also, a comparison between magnetic behaviour of identical resonators mounted with different materials definitely proves the responsibility of the supports in the magnetic sensitivity of resonators and gives interesting information on its reduction.

1. Introduction

As the oscillator performance increases, new applications such as precision orbit determination need more and more accurate devices. It results that some perturbation sources ignored to date as having in most cases negligible influence, now emerge and need attention. Thus, for a few years, it has been observed that starting or stopping on-board magnetic torquers may lead to slight frequency shifts in an oscillator closed by it.

The order of magnitude of this effect (10^{-11} / gauss) is unacceptable in high precision orbitographic measurement (DORIS mission). Various papers [1-3], have shown that almost all components of the oscillator circuitry may have a more or less important effect on the magnetic sensitivity. At the same time some solutions to reduce this sensitivity have been proposed, for example by changing some component technologies [4] or by shielding the oscillator with high magnetic permeability material [5].

Nevertheless these improvements are either not sufficient to meet the required specifications or not satisfactory because of the resulting volume and weight increase. A large part of the oscillator magnetic sensitivity finds its origin in the resonator itself. Previous works have demonstrated that it is not due to the quartz material itself but rather to the magnetic properties of the metallic springs used to support the plate [1], [6]. Then, a reduction of the magnetic sensitivity goes through a better understanding of the magneto-mechanic interaction between springs and plate and a better knowledge of the magnetic properties of the material.

2. Resonator magnetic sensitivity measurement setup

The experimental setup used in magnetic sensitivity measurement described a few years ago [6] has been slightly modified so as to meet the recommendations of the IEEE Standard [7, 8]. The resonator is now kept in a fixed direction so as to avoid errors due to the earth's magnetic field or to gravitational effects. Only the magnetic field strength used is larger than the recommended value so as to compensate the reduction of the signal to noise ratio due to the experimental setup. In fact, the oscillation sustaining circuit is kept far from the resonator to insure that it is the only component submitted to the magnetic field. The other experimental measurement procedures and signal processing techniques are described in ref. [6].

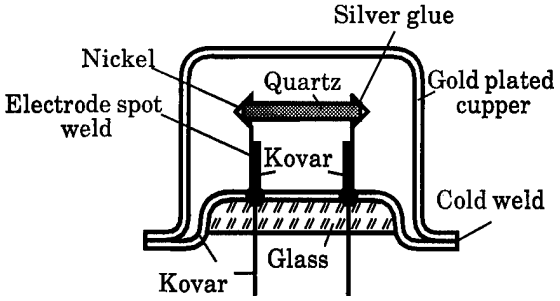


Fig. 1 : *Internal assembly of a resonator.*

0-7803-1945-1/94 $3.00 © 1994 IEEE

Most of the resonators used in the experiments are regular units as shown in Fig. 1. The quartz disc is cramped by two nickel springs soldered on two kovar supports. This latter material is used because it has the same thermal expansion coefficient as the insulating glass through which pass the supports. The electrical connection between quartz electrodes and springs is provided with a conducting silver glue. The resonator is maintained under vacuum by a gold plated copper cap cold welded on a kovar seat.

3. Possible magneto elastic interactions [9]

Various mechanisms can be considered to explain how a magnetic field induces a mechanical action in a ferromagnetic material.

3.1. Compass effect

The first effect is explained by the fact that ferromagnetic materials are constituted by small magnetic domains initially randomly oriented. When submitted to a magnetic field these domains tend to align along the same direction corresponding to the easiest magnetization direction which is usually the rolling direction of the material. As the domains reach the saturation state, their magnetization vectors oriented along the same direction are subject to a torque tending to align them with the external magnetic field just like a compass needle. This situation is sketched in Fig. 2.

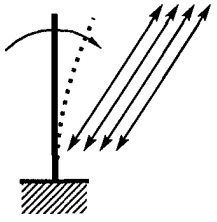


Fig. 2 : *Compass effect.*

Since the supports of the quartz plate are clamped on the seat, the assembly could bend as shown in figure 3 and presents a maximum effect when the magnetic field lies in the plane of the supports.

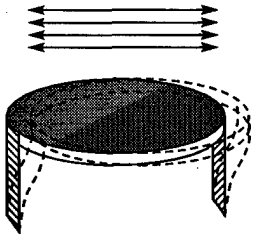


Fig. 3 : *Bending of the assembly.*

It should be noted that a change in the magnetic field sense also changes the sense of the magnetization vector so that the resulting torque doesn't change.

3.2. Mutual repulsion

Two parallel ferromagnetic supports close together submitted to a magnetic field act as two small magnets of same polarity. As a consequence they tend to repel each other and to induce in the quartz plate they support a diametrical tension (see Fig. 4).

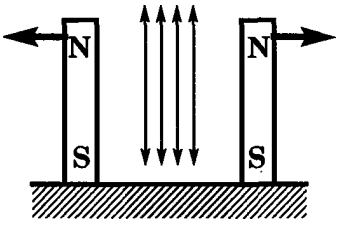


Fig. 4 : *Repulsion forces.*

As in the previous case, a change in the applied magnetic field sense also changes both magnet polarities so that the resulting effect is always a mutual repulsion.

3.3. Change of elastic modulus

This effect also called *ΔE effect*, consists of a change in the Young's modulus of ferromagnetic materials when submitted to a magnetic field. As a result, the pressure exerted by a ferromagnetic spring on a sustained disc is modified by an external magnetic field (see Fig. 5).

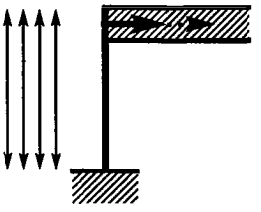


Fig. 5 : *ΔE effect.*

Fig. 6 shows the variation of Young's modulus for various ferromagnetic material [9]. The sign of this variation doesn't depend on the sense of the magnetic field.

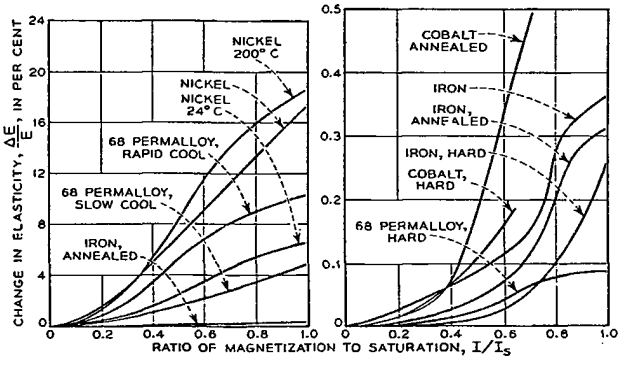


Fig. 6 : *Change in Young's modulus with magnetization (after [9]).*

3.4. Magnetostriction

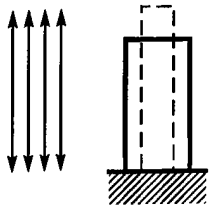


Fig. 7 : *Magnetostriction.*

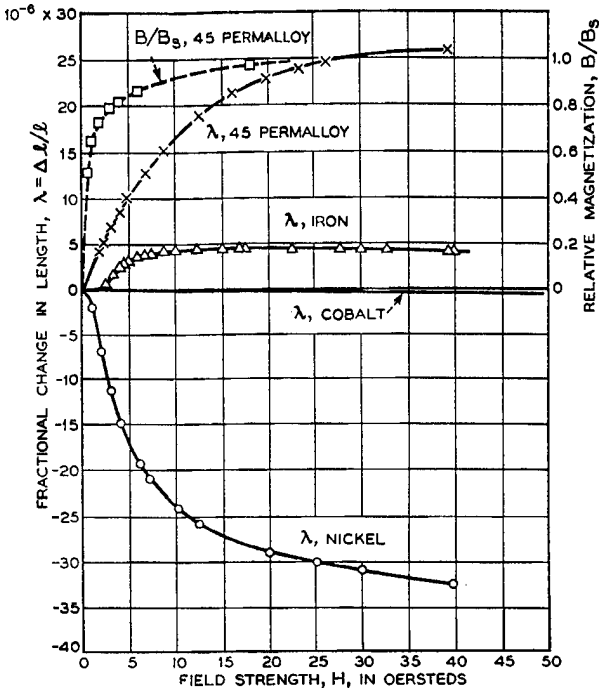


Fig. 8 : *Magnetostriction of some ferromagnetic materials [9].*

The last, and probably the most important, magneto-mechanic effect is the magnetostriction. That is a change in the dimensions of a ferromagnetic material submitted to a magnetic field. For example, as shown in Fig. 7, a rod of nickel submitted to a magnetic field undergoes a contraction in length and an expansion in the other directions so that the volume is kept approximately constant. In many cases, ferro-magnetic materials may present an expansion in length rather than a contraction. In all cases, the deformation (contraction or expansion) only depends on the kind of material but not on the sense of the applied magnetic field. Figure 8 shows the magneto-striction coefficient of some ferromagnetic materials [9].

The kovar used in various parts of the resonator packaging (see Fig. 1) is also a ferromagnetic material the magnetostriction of which has been measured at the *Laboratoire de Magnétisme Louis Neel* in Grenoble. The result is shown in Fig. 9 [10].

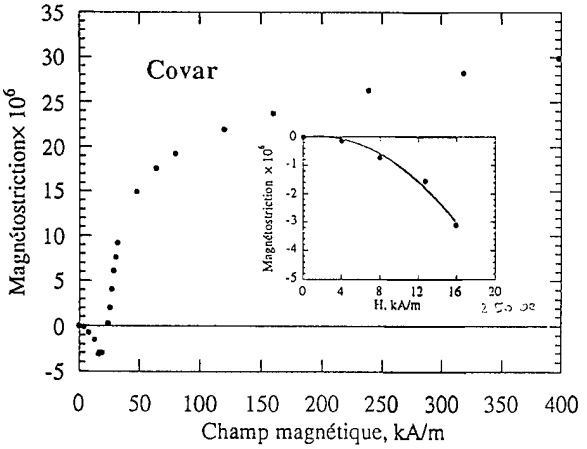


Fig. 9 : *Magnetostriction of kovar [10].*

It is easy to understand that the magnetostriction of the springs and/or the supports can induce stress in the disc they sustain.

3.5. Hysteresis

Most of the magnetic effects described above don't depend on the sense of the applied magnetic field and present a maximal effect when the field is directed along the largest dimension of the material. Also, most of these magnetic effects present a more or less pronounced hysteresis depending on the kind of material. For example, Fig. 10 shows the hysteresis of magnetostriction in nickel [9].

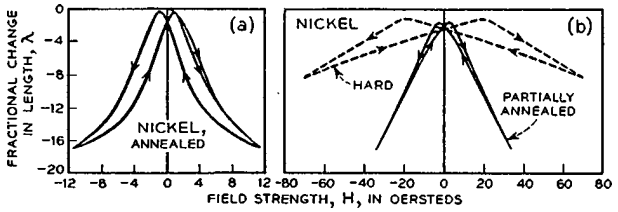


Fig. 10 : *Hysteresis of the magnetostriction in nickel (after [9]).*

Except for the ΔE effect, an external magnetic field gives rise to a mechanical displacement of the ferromagnetic springs and supports which, in turn, induces stress in the quartz plate and then a change in the resonant frequency.

4. Measurements of magneto-mechanic displacements

In almost all magnetic sensitivity measurements performed on resonators, the maximal effect is observed when the magnetic field is directed along the axis of the resonator (see Fig. 11, ref. [6] and Figs 16-19 thereafter).

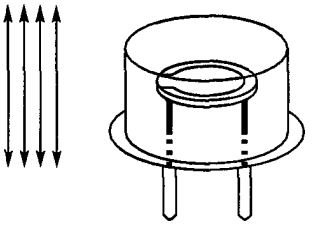


Fig. 11 : *Most sensitive direction of the magnetic effect.*

This fact implies that the "*compass effect*" is probably not responsible for the observed sensitivity. So as to tell one of the remaining effect from the others, a set of experiments have been performed using a heterodyne interferometric laser probe with enhanced sensitivity enabling dynamic displacement measurements as low as a few 10^{-5} Å·Hz$^{-½}$ [11]. Fig. 12 shows a simplified bloc diagram of the experimental setup.

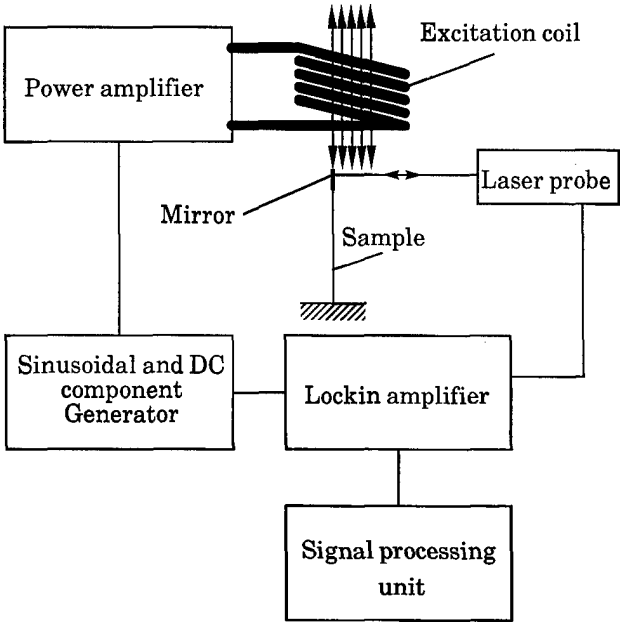


Fig. 12 : *Experimental setup for the measurement of magneto-mechanic displacements.*

A ferromagnetic sample of the same material as used for the resonator springs is clamped at one end on a fixed support, a small mirror glued at the free end is used to reflect the laser beam of an interferometric probe. The sample is submitted to a 1 kHz sinusoidal magnetic field produced with an excitation coil fed by a frequency generator through a power amplifier. In addition to this dynamic magnetic field of amplitude 10 gauss a DC magnetic component in the range ± 30 gauss can be superposed.

This arrangement enables to observe the small displacements of the free end of the sample by using a lock-in amplifier and a signal processing unit. The resolution of the displacement measurement increases with the excitation frequency ,due to optical noise, while the dynamic field magnitude decreases. A 1 kHz excitation frequency has proved to be a good compromise. Figs. 13 and 14 show the dynamical displacement at the end of the ferromagnetic sample in different conditions as a function of the DC magnetic component.

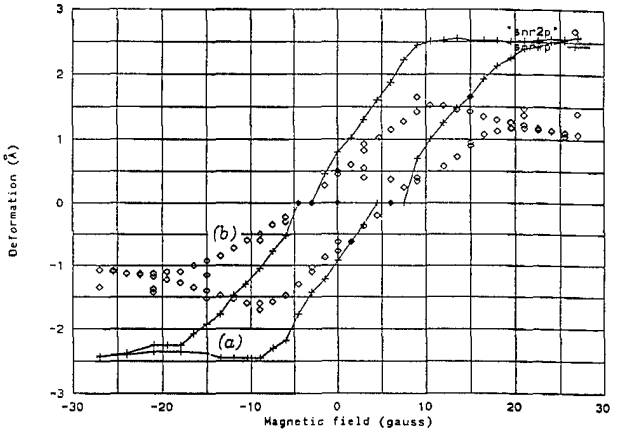


Fig. 13 : *Dynamical displacement of the free end of the sample vs. DC magnetic field component with (a) or without (b) the other support present.*

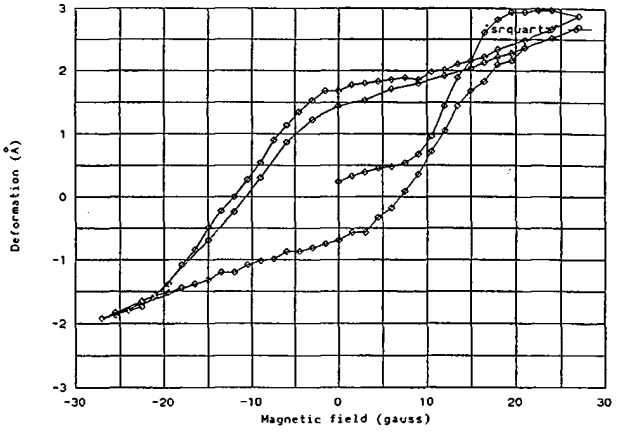


Fig. 14 : *Dynamical displacement of the end of the sample vs.DC magnetic field component when the quartz disc is in place.*

Because of the particular detection system, the curves obtained represent the deformation due to the small dynamical magnetic field around a DC biasing field so these curves are not directly comparable with the magnetostriction curves shown in figure 10, rather they represent their derivative.

Figure 13 shows two results, curve (a) is obtained by measuring the dynamical displacement of one spring in the presence of the other one when the quartz plate has been removed (see Fig. 15 a). Curve (b) is obtained in the same conditions but after the other spring has been removed (Fig. 15 b). Figure 14 shows the result obtained when the whole assembly - springs and quartz - is submitted to the magnetic field (Fig. 15 c).

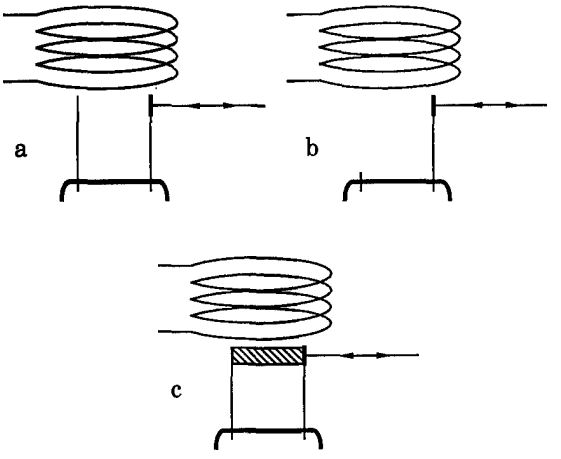


Fig. 15 : *Experimental conditions for Figs. 13 and 14.*

From these experiments it can be deduced that the frequency shift of the quartz plate is due to the displacement induced by the magnetic field on the springs and supports, the most probable origin being their magnetostriction (ΔE effect doesn't produce any displacement) nevertheless, the difference between curves (a) and (b) in Fig. 13 shows that the mutual repulsion effect cannot be neglected. We will show in the next section that other experimental results also incline towards the assumption of magnetostriction.

5. Magnetic sensitivity of resonators, results

A large number of experiments have been performed on resonators of various cuts, technologies and mounting. Some results have already been published [6] and only the last results will be briefly presented here.

5.1. Influence of the resonator technology

Several lots of resonators have been measured so as to investigate for the general characters and possible dispersion of their magnetic sensitivity. All results presented here have been obtained with 10 MHz third overtone SC cut quartz resonators using nickel springs in T2111 box enclosure submitted to a 1000 s-period sinusoidal magnetic field as described in ref. [6] and sec. 2. Although experiments have been performed with different magnetic field orientations, only the results obtained in the most sensitive direction (along the resonator axis) will be reported here.

Figs. 16 a and b show the results obtained with two QHS resonators from the same batch, Fig. 17 a and b are obtained with two other resonators of same technology from another batch. The most surprising fact when looking at these results is the great variety of "*magnetic signature*" while the overall amplitude keeps approximately the same order of magnitude.

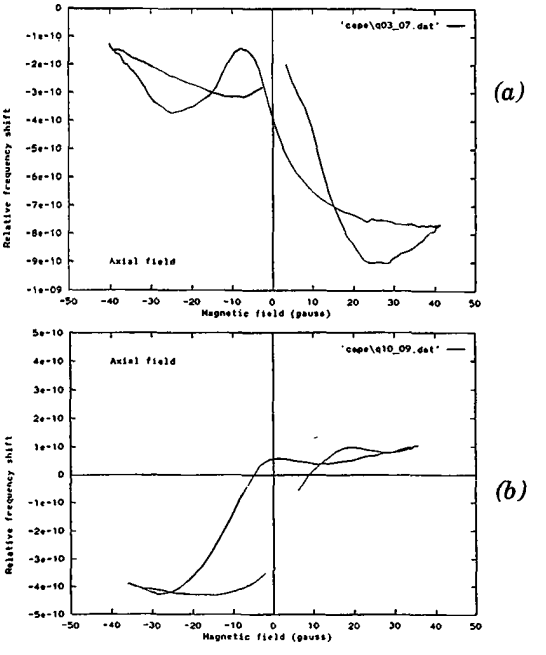


Fig. 16 : *10 MHz SC cut resonator QHS technology (same batch).*

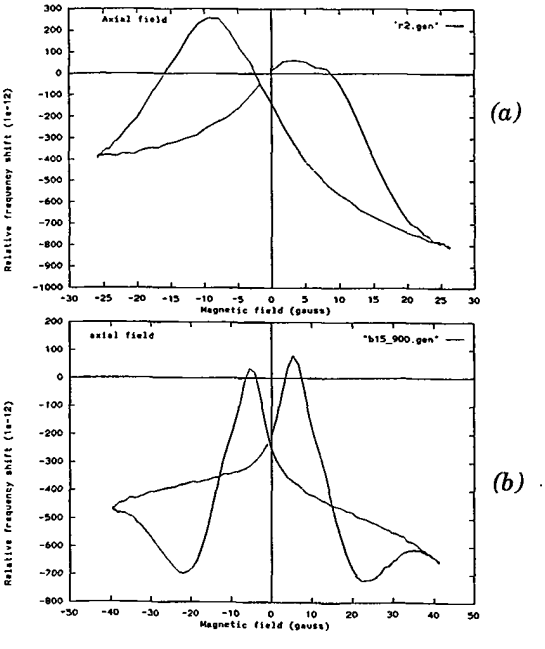


Fig. 17 : *10 MHz SC cut resonators, different batch but same technology as in Fig. 16.*

Fig 18 a and b are obtained with two QAS from the same batch. Their magnetic signatures are fairly similar but much larger than those of QHS (the scale in Fig. 18 is four times larger than the scale of Figs. 16 and 17). The similarity between the magneto-striction expansion coefficient (Fig. 10) and the curves shown in Fig. 18 is one of the most convincing features on account for explaining the magnetic sensitivity of resonators by the magnetostriction of the springs.

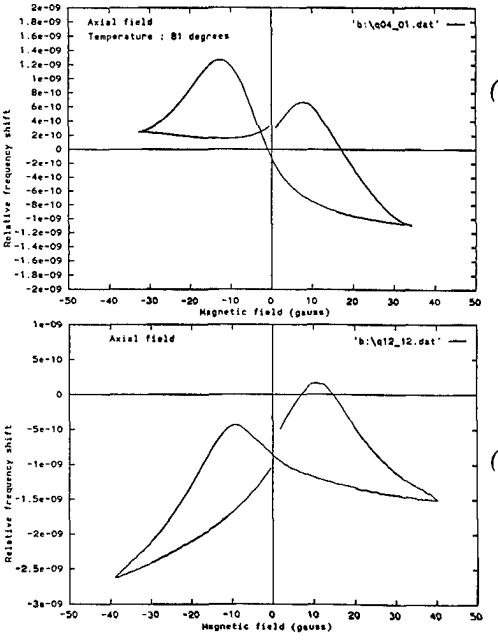


Fig. 18 : *10 MHz SC cut resonators QAS technology (same batch).*

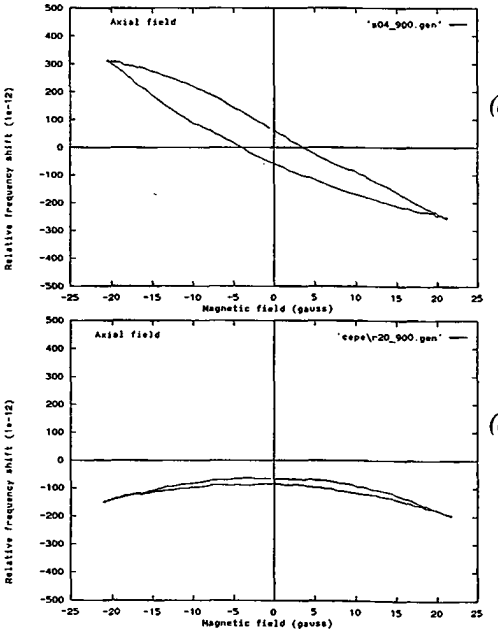
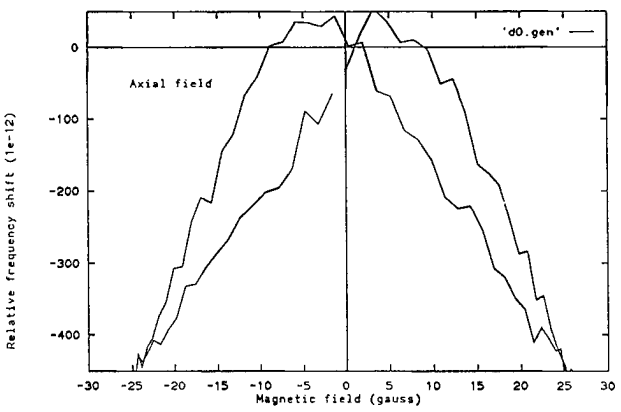


Fig. 19 : *10 MHz SC cut resonators QAS technology:*
(a) without intermediate plate.
(b) with intermediate plate.

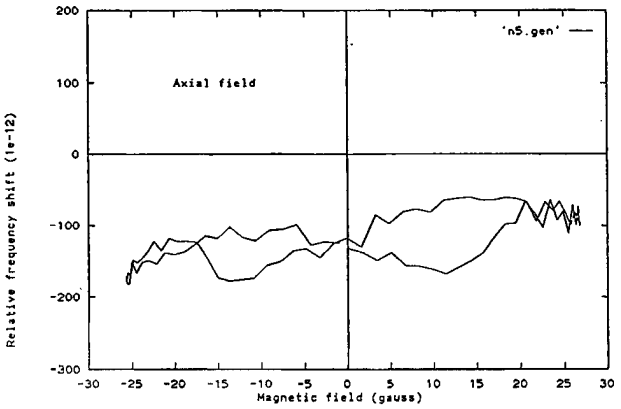
Another experiment has been performed to compare the results obtained with two QAS resonators. One of them (Fig. 19 a) is a regular one, in the other one the resonator is mounted on an intermediate plate (Fig. 19 b). The barometric sensitivity of the first resonator has been measured to be 25 times larger than the second, while its overall magnetic sensitivity is 6 times larger. Also these figures need some comments: if the magneto-elastic mechanism responsible for the magnetic sensitivity of resonator is indeed the magnetostriction of the springs, the frequency shift-vs-magnetic field curves would present the same symmetry as the magnetostriction curve (Fig. 10), so the magnetic sensitivity curve should be an even function of the magnetic field variable. This is approximately true in a number of figures presented in this section but this is not always the case (see for example Figs. 16 b and 19 a). It is highly probable that the dissymmetry of the magnetic sensitivity curves reveals in fact a dissymmetry in the mechanical mounting of the quartz plate. As a consequence, the stress induced in the plate by the magnetostriction of the springs is no longer symmetrical with respect to the plate axis and the resulting frequency shift losses the initial symmetry of the phenomenon. Hence the high barometric sensitivity of the resonator used for Fig. 19 a could be explained by a mounting dissymmetry revealed by the strong dissymmetry of the magnetic sensitivity curve, indeed it is well known that any dissymmetry in the mechanical assembly of a resonator drastically increases its sensitivity to environmental perturbations (acceleration, pressure, etc.) [4].

5.2. Comparison between spring materials.

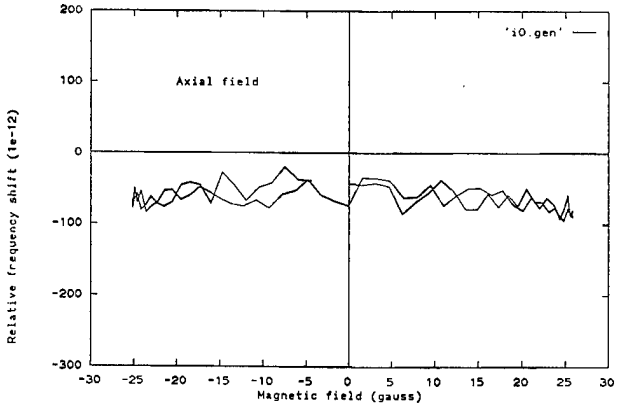
Many efforts have been spent to reduce the magnetic sensitivity of the resonators, among them materials different from the nickel usually used have been tested. A set of five resonators have been investigated, all of them are identical except for the springs made in different materials. The resonators are all 6.144 MHz AT cut QHS units. Because of the particular frequency of these resonators the experimental conditions have been slightly modified, the curves obtained, shown in Fig. 20, are then a bit more noisy, nevertheless they remain worth reading. They are presented from the most to the less magnetic sensitivity. The kovar appears to be much more sensitive than the nickel while stainless steel is noticeably less sensitive. It is not surprising that the resonator made with copper-alloy springs doesn't exhibit any perceptible magnetic sensitivity. This last experiment definitely proves the responsibility of the magnetic properties of the springs in the magnetic sensitivity of resonators. Of course, a reduced magnetic sensitivity is not the only goal to reach, the resonators still have to keep their performance in terms of mechanical behaviour, accelerometric and barometric sensitivity, ageing, etc.



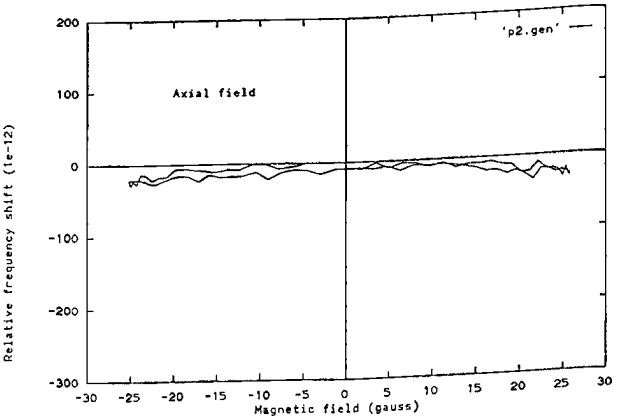
(a) Kovar springs.



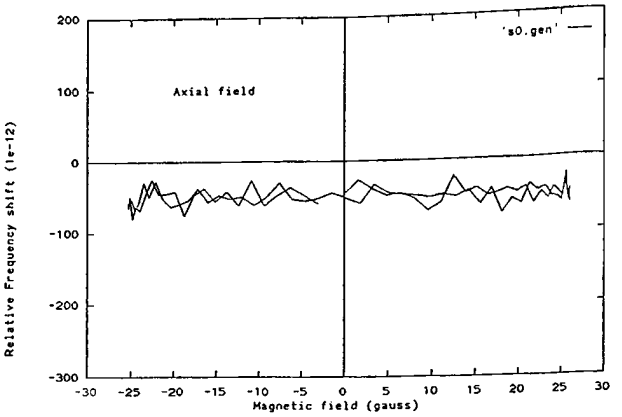
(b) Nickel springs.



(c) Stainless steel springs.



(d) Other kind of stainless steel springs.



(e) Copper-alloy springs.

Fig. 20 : *6.144 MHz AT cut QHS resonators.*

6. Conclusion

The magnetic sensitivity of resonators comes undoubtedly from the ferro-magnetic properties of the springs used to hold the quartz plate up. This sensitivity can be drastically reduced by using copper alloy springs while kovar and nickel should be avoided. Although the quartz plate itself is not responsible for the magnetic sensitivity, the magnitude of the phenomenon can be strongly increased by the geometrical characteristics of the disc, it is obvious that the thicker the plate is the lower the sensitivity appears. Even in a same lot of resonators, a great variety of magnetic signatures is often observed, some of them being abnormally dissymmetrical in view of the magnetic action and geometric assembly symmetries. It is highly probable that these anomalies betray a geometric dissymmetry in the plate mounting.

In most of the experiments, the higher sensitivity is observed when the magnetic field is directed along the resonator axis that is along the largest dimension of the springs. This fact, in addition with the similarity between magnetostriction curve and magnetic sensitivity curves of resonators, demonstrates that the magnetostriction is the most probable mechanism responsible for the observed phenomenon. It is highly probable that the complexity of the magnetic signature has to be attributed to the spring magnetostriction hysteresis rather than to a complicated stress and strain interaction between plate and springs. This interaction might be as simple as a diametrical tension or compression for which the induced frequency shift is well characterized. Then, the magnetic sensitivity of resonators could constitute an extremely sensitive tool for the experimental investigations on the mechanical actions exerted on the resonator plate. In fact, by knowing the magnetic behaviour of the spring it is possible to determine the kind of the actions the plate is submitted to and the device thus obtained can be easily used for example to experimentally investigate for the lowest sensitivity fastening points on the plate or to study new shapes of springs and supports.

These applications still need theoretical modelling of the magneto-elastic behaviour of the various materials and it is in these directions that the efforts are currently being made.

Acknowledgements

This work was supported by Thomson-CEPE.
The authors wish to thank Dr. G. Marianneau for his help.

References

[1] R. Brendel, C. El Hassani, M. Brunet, E. Robert, *"Influence of magnetic field on quartz crystal oscillators"*, Proc. 43rd Ann. Symp. on Freq. Control, pp. 268-274 (1989).

[2] R. Brendel, C. El Hassani, P. Krux, E. Robert, *"Magnetic sensitivity of oscillator components"*, Proc. 5th Europ. Freq. and Time Forum, pp. 390-395 (1991).

[3] F. Deyzac, *"Oscillateur à quartz de bord en environnement sévère. Sensibilité magnétique"*, ONERA. Technical Report # 37/7132 PY (1988).

[4] F. Deyzac, *"Oscillateur à quartz ultrastables à finalités militaire et spatiale. Amélioration du comportement mécanique"*, ONERA. Technical Report # 8/3945 PY (1993).

[5] J.-P. Granger, *"Blindage magnétique de l'OUS de DORIS"*, CNES Technical Report # S2/NT/10 000/3800/MT (1989).

[6] R. Brendel, E. Robert, *"Weak Magnetic Field sensitivity of Quartz Crystal Oscillators"*, Proc. 6th Europ. Freq. and Time Forum, pp. 99-104 (1992).

[7] R. L. Sydnor, R. Brendel, W. Wiedemann, *"IEEE Standards Project P1193, Electric and Magnetic Fields and Effects"*, Proc. 46th Ann. Symp. on Freq. Control, pp. 794-797 (1992).

[8] IEEE-SCC-27 Working Group, *"Guide for Measurement of Environmental Sensitivities of Standard Frequency generators"*, Proposition PAR-P1193 to be published (1994).

[9] R. M. Bozorth, *"Ferromagnetism"*, Van Nostrand, 8th edition, (1964).

[10] A. Barlet, E. de Lacheisserie, *"Analyse de l'aimantation et de la magnétostriction de quelques composants d'un résonateur à quartz"*, Technical Report, unpublished (1993).

[11] B. Cretin, W. X. Xie, S. Wang and D. Hauden, *"Heterodyne Interferometers"*, Optics Communications, **vol 65**, n° 3, (Feb. 1988).

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

MEASUREMENT OF SPURIOUS RESONANCES OF CRYSTAL UNITS USING NETWORK ANALYSERS WITH ERROR CORRECTION

Bernd W. Neubig
TELE QUARZ GROUP
D-74924 Neckarbischofsheim
Germany

Abstract

1. Introduction

The hybrid bridge technique for the measurement of spurios responses as standardized by IEC 283 (1968) [3] has significant limitations in accuracy and reproducibility. Significant progress has been made in the last years applying modern network analysers in connection with error correction methods to the measurement of crystal units [6].

The paper presents a systematic approach to adapt the new techniques to the measurement of spurious responses. The goal is to use the standard test set-up (such as the π-network or a s-parameter test fixture) as it is used for the determination of the crystal parameters of the main mode [5]. In connection with the published method for the measuerement of load resonance parameters[7], all crystal parameters can then be measured with one unique set-up configuration.

2. Classification of spurious responses

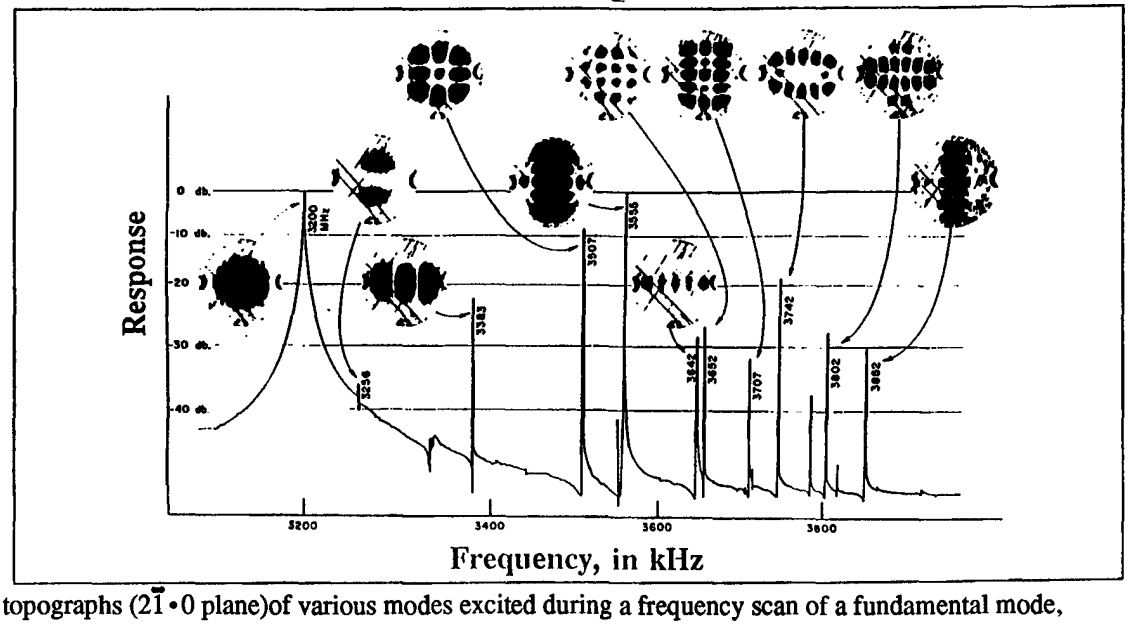
Unwanted or spurious modes have different origin and appearance. They can be

(1) anharmonic resonances of the same vibration mode (usually shear mode), whose frequency is in general above the relevant main mode (see Fig.1 [8]). These resonances are determined by the equation

$$f_{mnp} = \frac{1}{2}\sqrt{\frac{1}{\rho}}\sqrt{\frac{c_{65}\,n^2}{t^2} + \frac{c_{11}\,m^2}{l_x^2} + \frac{c_{56}\,p^2}{l_y^2}}$$

(2) unwanted modes (and overtones thereof) of other vibration modes, which may lie below or above the desired mode.

Fig. 1



X-ray topographs ($2\bar{1}\cdot0$ plane)of various modes excited during a frequency scan of a fundamental mode, circular, AT-cut resonator. The first peak, at 3.2 MHz, is the main mode; all others are unwanted modes. Dark areas correspond to high amplitudes of displacement.

0-7803-1945-1/94 $3.00 © 1994 IEEE

Spurious modes can be

(i) isolated from the main mode, or can be

(ii) mechanically coupled to it, causing non-linear effects which depend on drive level (power dissipation) etc.

Spurious modes can lie

(I) very close to the main mode (a few kHz), such that the peak of the spur is on the slope of the main response.

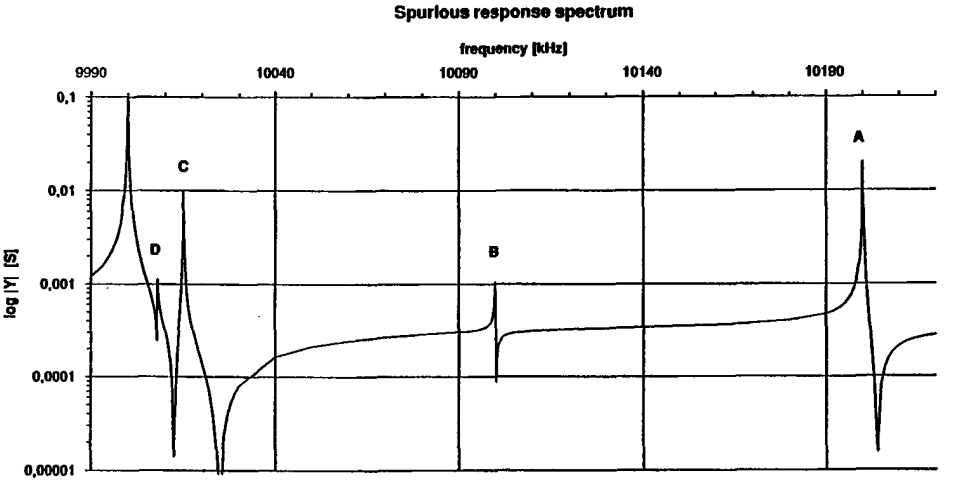
(II) well apart from the main mode, and therefore can be considered as a separate peak.

Unwanted modes can cause problems

(a) in filters by distortion of the transfer characteristic in the pass band or the stop band: attenuation "dips" and phase discontiuities in the pass band, detoriation of the attenuation in the stop band

(b) in oscillators, the working frequency can "jump" to a strong unwanted mode or the frequency response can show discontinuities ("activity dips") at particular operation conditions (temperature, load capacitance/pulling voltage, drive level). VCXOs can show "dips" ins the pulling characteristic (frequency vs. pulling voltage) or in the response of the deviation vs. modulation frequency. TCXOs can show "dips" in the frequency vs. temperature characteristics.

Fig. 2



In the following we will distinguish phenomenologically four different cases:

case A: strong spurs well "isolated" from the main mode,

case B: weak spurs well "isolated" from the main mode,

case C: strong spurs in the vicinity of the main modes,

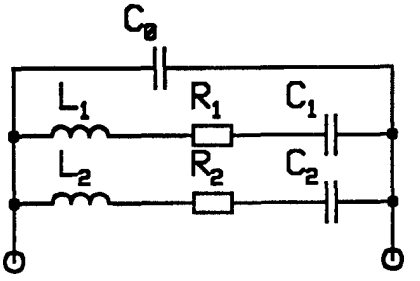
case D: weak spurs in the vicinity of the main mode.

These cases are depicted in Fig.2.

3. Equivalent electrical circuits

Non-coupled resonances can be represented by seperate series resonance circuits all connected in parallel to the static capacitance C_0. In the following we use the symbols C_0 for static capacitance, L_1,C_1,R_1 for the motional parameters of the first (main) mode, and L_i, C_i, R_i with i = 2, 3, 4,... for the i-th (spurious) mode (see Fig.3).

Fig. 3



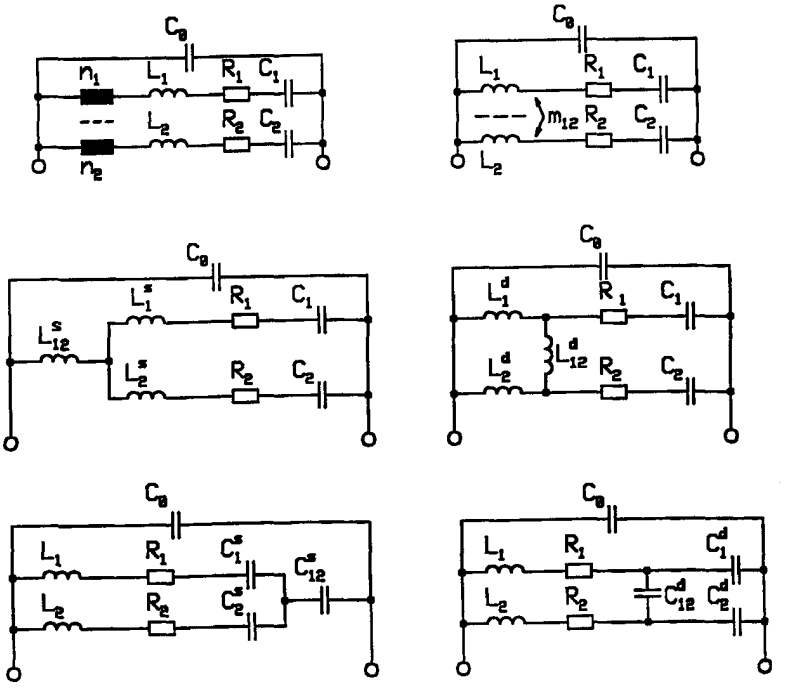
Resonances, which are mechanically coupled, must be represented electrically by a coupling circuit such as a transformer or an equivalent coupler e.g. three capacitors or inductors in delta- or star configuration, where one or two of them may have negative element values. Possible circuits are summarized in Fig.4.

4. Classical measurement methods

The classical measurement technique for spurious modes is the hybrid-coil method, described by Horton and Smythe [1] and Priebe [2], which was standardized in IEC 283 [3].

It uses a differential transformer with a variable capacitor C_C in the second branch to "compensate" the static capacitance C_0 (see Fig.5).

Fig. 4



Due to its simplicity in practical application (once, the hybrid transformer is made ...), this method is very popular. However it has significant disadvantages:

(1) Even if C_0 is compensated (C_C=C_0), it has not disappeared: It appears with twice its value in parallel to the source and load termination resistors. Therefore the frequency of minimum attenuation is not exactly equal to the series resonance frequency, and the resonance resistance computed by a formula neglecting C_0 - which is the usual practice - is wrong.

(2) Spurious resonances close to the main response can not be characterized properly. The "dB" or "differential dB" values cannot be directly converted to resistance values or to the motional parameters.

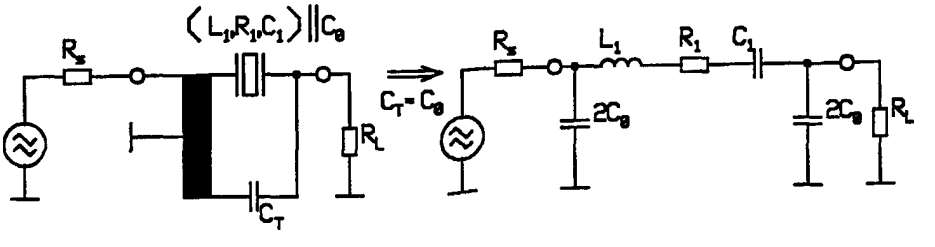
(3) Several problems are related to the hybrid transformer: It must have a flat frequency response, low winding capacitance, low stray inductance, perfect coupling and symmetry over a wide frequency range. Because this critical part is not commercially available, it suffers from poor reproducibility.

(4) The manual measurement is very time-consuming, as it is necessary to sweep in the vicinity of each resonance very slowly to achieve sufficient accuracy.

The disadvantages in item (3) can be overcome, if the hybrid transformer is replaced by a commercially available 180° hybrid coupler embedded in a 50 Ω environment, as proposed in [4]. This however does not eliminate the above mentioned problems (1),(2) and (4).

Stronger responses can also be measured by the conventional π-network method, however the measurement error increases, the weaker the spurs are and the closer to the main mode they are.

Fig. 5



5. Alternative measurement methods

The measurement technique to be proposed is based on a test set-up with network analyser or equivalent, the π-network according to IEC 444 and the error correction technique described in [5] and standardized in IEC 444-5. The same method can also be used in conjunction with a s-parameter test system and adequate crystal test fixture.

Step 1: Calibration with short, open and reference resistor in the frequency range of interest

Step 2: Measurement of the static capacitance according to IEC 444-5

Step 3: Determination of the resonance frequency f_r (resp. series resonance frequency f_s, or load resonance frequency f_L) and the equivalent parameters R_1, C_1(and L_1, Q_1) of the main mode with one of the error-corrected methods described in IEC 444-5. If spurious modes are very close to the main mode (i.e. closer than $f_s/1000$), the values derived herewith may include non-negletable errors, but can be used as starting values for the iterative procedure described later.

Step 4: Fast sweep (approximately 1 sec) through the frequency range of interest upwards (increasing frequency) and downwards. A minimum of 500 data points is recommended. Substract the error terms, compute $Y_{up}(f)$ and $Y_{down}(f)$. Substract the admittance $j\omega C_0$ from both sets.

Step 5: Inspect the data and select computationally the peaks with a given minimum (trigger) value Y_{min}. Average the related peak frequencies of both sweeps. This yields an excellent guess for the resonance frequencies of the spurs. The sensitivity of the peak selection can be improved by mathematical differentiation of the Y(f) values. Alternatively the network analyser's internal "peak search" function can be used if implemented.

Step 6: Zoom each spurious response individually and measure Y(f) with higher frequency resolution.

The following steps depend on the type of spur to be considered (see classification in clause 2).

case A (strong spur well "isolated" from the main mode)

Zoomed data points can be selected as in IEC 444-5. Fastest approach is the iterative procedure as desribed in [5]. From a minimum of two or three test frequencies the i-th spurious resonance can be characterized by f_i, R_i, C_i (and L_i,Q_i).

case B (weak spur well "isolated" from the main mode)

Method as in case A. To improve sensitivity, the following modifications could be considered:

* averaging of several sweeps
* reduced RF bandwidth and video bandwidth (slower speed !)
* use of frequency selective, low noise network analyser. The technique using direct measurement of amplidue and phase is in this respect advantageous over the s-parameter systems.
* increase impedance R_T "seen" by the crystal in the test fixture. The s-parameter fixture with its $R_T = 100\Omega$ is advantageous over the IEC 444 π-network ($R_T = 25\Omega$).

case C (strong spur in the vicinity of the main mode)

Data are zoomed with high frequency resolution in an interval covering the 3dB-bandwidth of the main mode and of the considered spur (its bandwidth is assumed to be approximately equal to that of the main mode - as is its Q in most cases).The parameters of both modes are then extracted by a parameter-fitting algorithm.

For uncoupled modes the equivalent circuit diagram of Fig.3 is used, i.e. three variables per resonance have to be determined (L_i, C_i, R_i or f_i, C_i, R_i). For coupled modes one circuit of Fig.4 has to be used. For N resonances $3 + 4*(N-1)$ elements have to be determined. Rather good starting values for the elements can be used which speeds up the parameter search and reduces the risk of ill-conditioned iterations to wrong local minima. These are:

f_1, R_1,C_1 (and therefore L_1 and Q_1) of the main mode from step 3

f_2 from the location of the peak maximum of Y(f) (after removal of $j\omega C_0$)

R_2 from the amplitude of the peak maximum, derived from

$$R_2 \approx \frac{1}{|Y(f_{peak})|}$$

C_2 from the assumption that the Q_2-value of the spur is approximately the same as Q_1 of the main mode

$$C_2 \approx \frac{1}{\omega_2 Q_2 R_2}$$

In the case of coupled modes, the starting value of element which determines the coupling can be arbitrarily set to zero.

The iteration procedure minimizes the error function

$$E = \sum w_i |Y_i - Y_i^M|^2 = \sum w_i ((G_i - G_i^M)^2 + (B_i - B_i^M)^2)$$

in which $Y_i = G_i + j\,B_i$ denotes the theoretical values from circuit analysis computed at the measurement frequency f_i, and $Y_i^M = G_i^M + j\,B_i^M$ represents the values derived from the measurement at f_i. The weighting factors w_i can be selected such, that the data points close to the resonances get the greatest significance, e.g. $w_i = |Y_i^M|$.
Several searching algorithms for minimization of E are applicable. One of them (Newton method) is described e.g in [9]. Another approach is the method of steepest descend, e.g. the Simplex algorithm. The evaluation of the different methods with respect to simplicity and stability is currently under progress.

case D (weak spur in the vicinity of the main mode)
The method is identical to case C, except that the starting value of R_2 has to be derived differently, because the amplitude of the spurious peak is too erraneous, as it is located on the slope of the main resonance curve.
A more reliable starting value for R_2 is gained as follows:
The approximate admittance of the main mode

$$Y_m(f_i) = \frac{1}{R_1 + j\frac{1}{\omega C_1}\left(\frac{\omega^2}{\omega_1^2} - 1\right)}$$

is computed at the measurement frequencies f_i by using the starting values of f_1, C_1, R_1 . It is then substracted from the measured admittance values Y_i^M in the vicinity of the main mode and the weak spur. From the peak of this residual Y_{spur} the starting value of R2 can be computed as

$$R_2 = \frac{1}{|\max(Y_{spur})|}$$

It should be noted, that the achievable accuracy for the parameters of very weak spurs is reduced. However, the method proposed herein is superior to the conventional methods used so far.

6. Conclusions

The paper describes a concept for the measurement of spurious resonances, which allows the use of network analysers in conjunction with error correction techniques. The test fixture can be the π-network or the s-parameter test fixture as described in IEC 444-5. The proposal includes the determination and characterization of weak spurs,which may lie close to a strong mode. It considers also the mechanical coupling of modes. The next steps are the optimization of the parameter fitting algorithm. The experimental verification and comporative measurements have to follow. These results will be reported later.

7. Acknowledgements

The author thanks Friedrich Krug (University of Erlangen), Eberhard Seydel (KVG), and Dr. Frieder Sohre (Philips Doetinchem) for the valuable discussions during the preparation of this work.

8. References

[1] Horton,W.H. and Smythe,R.C.: The hybrid-coil bridge method of measuring unwanted modes of vibration in quartz crystals; Proc. 17th Annual Symposium on Frequency Control (ASFC) 1963, pp.316-324

[2] Priebe,F.K.: Attenuation and resistance measurements of unwanted modes of quartz crystals; Proc. 19th ASFC (1965), pp.487-508

[3] IEC publication 283 (1968): Methods for the measurement of frequency and equivalent resistance of unwanted resonances of filter crystal units

[4] IEC document 49(Germany)109 (1985): Draft - Proposal for extension of the π-network method for the measurement of main and unwanted resonances of crystal units up to 500MHz.

[5] Neubig, Bernd W.: Measurement of Quartz Crystal Units up to 500 MHz and Above by the Use of a Pi-Network with Error Correction; Proc. 11th Quartz Devices Conference, Kansas City (1989), Vol.1, p.60ff.

[6] IEC publication 444-5 (1993): Methods for the measurement of quartz crystal devices for the determination of equivalent electrical parameters using automatic network analyser techniques and error correction

[7] Neubig, Bernd W.: PI-Network Measurement of the Load Resonance Frequency without Load Capacitor;Proc. 4th European Frequency and Time Forum (EFTF) (1990), pp.481-486

[8] Vig, John R.: Quartz crystal resonators and oscillators - a tutorial; U.S. Army Electronics Technology and Devices Laboratory (1990)

[9] Peach,R.C., Morris,S.E.: A system for precision parameter measurements on quartz crystal resonators and bipoles; Proc. 39th ASFC (1985), pp.527-534

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

TEST OSCILLATOR FOR STUDY OF DRIVE LEVEL DEPENDENCE OF QUARTZ CRYSTALS

Lennart Lindberg and Guy Portnoff

Quartz Pro Instrument AB, Box 6062, S-175 06 Järfälla, Sweden

Abstract

Test oscillators, in contrast to passive measuring methods, are able to span the whole range of drive levels from noise to normal operation. This is important for simulation of start-up conditions for practical oscillators, as well as for checking filter crystals for operation in a wide range of signal levels. A wide-band, negative-resistance Butler oscillator is modified with a ramp generator, controlling the negative resistance. This is zero when the crystal is inserted; any transient electrical or mechanical shocks due to insertion cannot affect start-up of oscillations. Analog output signals represent crystal current and oscillator negative resistance. Amplification of hf-signals makes possible observation of low level oscillation ('second level of drive'). Repeated sweeps can demonstrate change of resistance between sweeps. The oscillator can serve as part of an automatic testing facility under computer control.

Introduction

Three methods for study and measurement of drive level dependence are today in use:

1. Test oscillators *e.g.* as described by Yerna [1,2]
2. Passive measurements using IEC-444 pi-network with vector voltmeter and phase-locked signal generator, complemented with variable attenuators and preamplifier.
3. Passive network analyser measurements at fixed frequency and swept output level.

The passive methods 2) and 3) require complex and expensive equipment, which allows *resistance measurement* down to about a nanowatt, but not down to the noise level.

The network analyser method when used with fixed frequency and swept output level may give questionable results if the resonance frequency depends on the drive level.

Advantages of test oscillator methods:

It can span the whole range of levels from noise to normal operation. This is important for two reasons: 1) It simulates start-up conditions of any practical oscillator. 2) It allows checking of filter crystals intended for operation at low signal levels.

Test oscillators are simple and inexpensive, and can be included as part of an automatic testing facility under computer control, for screening of crystals to specified acceptance limits.

Main features of new design:

The negative resistance is controlled by a ramp voltage.

The negative resistance is zero when the crystal is inserted, so that any possible transient electrical or mechanical shocks due to insertion of the crystal cannot affect start-up.

Amplification of hf-signals before detection makes possible observation of low level oscillation ("second level of drive").

Repeated sweeps can demonstrate reversible effects for the motional resistance from sweep to sweep during operation.

Analog output signals represent the crystal current and oscillator negative resistance. A hf-signal output is available for monitoring on an oscilloscope and for frequency measurement.

The drive level is related to the *crystal current*, which is the physically most relevant quantity, since it is directly related to the *vibrational amplitude* in the crystal [3,4], while the power is rather a measure of the *heat* developed in the crystal.

The Oscillator Circuit

Variation of the negative resistance

Usually the wide-band, Butler oscillator is operated with self-limitation due to nonlinearities. This results in

0-7803-1945-1/94 $3.00 © 1994 IEEE

waveforms far from sinusoidal because of the conflicting requirements of *purely sinusoidal current in the crystal* (in the mechanical branch of the equivalent circuit), and *pulsed currents in the associated transistors*; the instantaneous difference between the two currents will charge the stray capacitances and the crystal C_0, which results in sharp peaks in the voltage waveforms.

Figure 1 shows our modified Butler oscillator. We have preferred linear, class-A operation in the whole range of levels of interest. This facilitates and makes possible accurate measurements of crystal current and equivalent resistance. The oscillator input transistor T_1 is shunted by another transistor, T_2 and these have a common emitter resistor, keeping the sum of their emitter currents, $I_{E12} = I_{E1} + I_{E2}$ almost constant. Linear operation is achieved by setting the dc-currents in the oscillator transistors T_1+T_2 as well as in T_4 well above the peak maximum crystal current requested. This has also the advantage that the dynamic emitter impedances, r_{e12} of T_1+T_2, and r_{e4} of T_4 are always low and well defined, of the order V_T/I_E ($\approx$5 ohm, $V_T \approx kT/e$), and that all currents in the circuit are closely to sinusoidal. An *ALC*-circuit is used for final limitation of the current.

We assume that the oscillator transistors have high f_T (2 - 5 Ghz), and will in the following analysis neglect all phase shift effects. In practice the small, but unavoidable phase shift will cause the crystal to oscillate at a frequency slightly different from its true series resonance frequency, but this will have a negligible influence on the measured equivalent resistance.

The fraction $p = I_{E1}/I_{E12}$ of the emitter current I_{E12} flowing into T_1 is controlled by the base current supplied to T_1 through the base resistor R_{B1}, which is connected to the positive going ramp generator. The crystal current, i_x flows into the emitters, making $i_{E12} = -i_x$.

The collector current in T_1, $i_{C1} \approx p\alpha i_x$ develops the collector voltage $v_{C1} = p\alpha R_{C1} i_x$, of which a fraction, A_{v3} is supplied to the base of T_4. A_{v3} is the voltage amplification of the source follower, T_3, and is slightly smaller than 1.

Closing the loop through the crystal and applying Kirchhoff's voltage law from B_4 to B_1 we obtain Barkhausen's condition for stationary oscillations with constant amplitude:

$$p\alpha A_{v3} R_{C1} i_x - r_{e4} i_x - R_1 i_x - r_{e12} i_x = 0 \qquad (1)$$

This can also be written in the form:

$$R_1 + R_- = 0 \qquad (2)$$

where R_1 is the crystal equivalent resistance, and R_- is the negative resistance of the oscillator,

$$R_- = -p\alpha A_{v3} R_{C1} + r_{e4} + r_{e12} \qquad (3)$$

According to eq. (2) $|R_-|$ *is a measure of* R_1

Setting for example $\alpha A_{v3} \approx 0.9$, $R_{C1} = 220$ ohm and $r_{e4} + r_{e12} = 10$ ohm, we find $R_- = 0$ for $p = 0.05$. The current in T_1 must consequently not be cutoff to make $R_- = 0$. This is an advantage; we need not to operate at very small currents in T_1, where transistor parameters may be different. In case a lower maximum $|R_-|$ is sufficient, a lower R_{C1} could be used.

Negative resistance measurement

The negative resistance R_- shown by the oscillator to the crystal is determined by the control voltage V_R, and *this is valid whether* V_R *is controlled by the ramp voltage, or by the ALC-signal.*

An output signal suitable for recording of $|R_-|$ can be derived in either of two ways: 1) From V_R, which is linearly related to the collector current, I_{C1} , provided the transistor β of T_1 is independent of the collector current. 2) From the dc-voltage drop across the collector resistor R_{C1}, which is a direct measure of I_{C1}, independent of transistor β. Either of these voltages is amplified and corrected for zero point offset, creating an output signal $V(R_-)$, representing $|R_-|$.

Crystal current measurement

The crystal current, i_x flows through T_4 and T_5, which form a cascode, and finally through R_{C5}, reduced by a factor slightly smaller than 1 because of the base currents in T_4 and T_5. The hf-voltage across R_{C5} is proportional to the crystal current. It is amplified, rectified and amplified in a dc-amplifier to provide an output signal, $V(I_x)$, representing the crystal current level.

This signal is also compared with a preset voltage V_{REF}, to produce an *ALC*-signal limiting the crystal current. When the prescribed current level is reached, the *ALC*-signal takes over the control of V_R, reducing the negative resistance exactly to the value required to maintain constant amplitude. The hf-attenuator, 20, 10, 0 dB is used to set the maximum crystal current to 1, 0.3, or 0.1 mA resp. The level can also be varied by changing the reference voltage, V_{REF}.

Observation of very small crystal currents

The hf-signal, representing the crystal current, is also fed to another hf-amplifier, with amplitude-dependent amplification, enhancing weak signals. This allows the weak signals, e.g. representing a 'second level of drive' to be amplified sufficiently to be seen in the output signal,

without overloading the amplifier. This amplifier also produces the output signal for frequency measurement.

Operation

Negative resistance sweep

When the crystal is inserted the ramp signal is at ground; the circuit shows no negative resistance. After a short delay the ramp voltage is started, and the current in T_1 increases, which results in an almost linearly increasing negative resistance. This goes on until oscillations have started and reached such a level that the *ALC*-circuit, which initially has a high output voltage, is activated and takes over control, preventing further rise of the control voltage, V_R.

The ramp can be operated as single sweep or repeatedly. In both cases the initial delay is important to be sure that any insertion transient has decayed, resp. that the crystal oscillations due to the previous sweep have had time to decay well below the noise level. Compare the superregenerative detector, which is a periodically re-starting oscillator, capable of detecting signals in noise !

Growth and decay of oscillations

As well known, a series resonance circuit, R_1, C_1, L_1, started at $t = t_0$, will perform damped free oscillations:

$$i = i_0 e^{-\beta t} \sin(\omega t + \varphi) \qquad (4)$$

where

$$\beta = R/2L = \pi f/Q \qquad (5)$$

If the circuit also contains a negative resistance R_- in series, R_1 should be replaced by $R_1 + R_-$. If this is < 0 the amplitude will grow exponentially; if > 0 it decays, but more slowly than when $R_- = 0$. In both cases the (instantaneous) *e*-folding time constant is, when $T = 1/f$:

$$\tau = 1/\beta = (QT/\pi)(R_1/|R_1 + R_-|) \qquad (6)$$

For *e.g.* a 10 MHz crystal with $Q \approx 100\,000$, $\tau_0 = QT/\pi \approx$ 3 msec

If $|R_-| = 2R_1$ the growth from noise level, say 10^{-9} A to 1 mA, requires $\ln(10^6) \approx 14$ *e*-folding times. If $|R_-| = 1.1\ R_1$, $\tau = 10\tau_0$, and the growth time will be still a factor of 10 longer, *i.e* of the order several tenths of a second.

The decay time in our oscillator is, if the current in transistor T_1 is cut off, determined by the resistance $R_1 + r_{e12} + r_{e4}$, thus the *e*-folding time is smaller than τ_0.

Recording

Figure 2a shows a schematic resistance characteristic for a drive level dependent (DLD) crystal as a function of the crystal current. The negative resistance of the oscillator, $|R_-|$ is represented by a horisontal line, which begins moving upwards when the sweep starts.

When the noise level start resistance is exceeded, at ***A***, oscillations begin to build up exponentially, or even faster, until the 'second level of drive', ***B***, is reached.

Then the point of operation follows the crystal characteristic until the peak, ***C***, whereafter the amplitude again increases rapidly till the *ALC*-circuit takes over the control, ***D*** , and the normal operating resistance is obtained.

Figure 2b shows schematically the corresponding recorded signals representing $|R_-|$ and I_x as functions of time.

When the rising ramp exceeds the initial R_1, oscillations grow, and the operating point moves from ***A*** to ***B***. The 'second level of drive' is in the region ***B*** - ***C***. In this case, with the high peak at ***C***, there is a big step down from ***C*** to ***D***, showing the large reduction of crystal resistance.

Even with good crystals there will be a small step down when the ALC takes over. This is caused by the long time it takes to build up oscillations from the noise level, as discussed in the text. During this time $|R_-|$ has increased a little.

Examples of recordings

We have tested the oscillator with a few crystals from an old collection of 'bad' crystals, plus one 'good' crystal.

In all the following recordings the sweep time is about 9 seconds. The zero level for $|R_-|$ is marked "1→ " at left. The scale for $|R_-|$ is 20 ohm/div, and for the crystal current, I_x, 0.2 mA/div. The linear $V(I_x)$-signal is utilized in all recordings except in Figure 6b.

Figure **3a, b, c** shows three recordings taken in sequence on a crystal having a 'second level of drive', which changes from sweep to sweep.

Figure **4a, b, c** shows three records in sequence on a crystal that at first shows a mysterious start resistance behavior, and finally a normal behaviour – it becomes a 'cured' crystal! In all the three records the oscillations start at the same low $|R_-|$-value.

In the first two the oscillation amplitude stops growing at a certain level, less than the *ALC*-level, and thereafter *decays*, more or less steeply, to a lower level, a level which a little earlier was passed under rapid growth, in spite of the much lower $|R_-|$-value at that time.

We conclude that in this case *the resistance - drive level dependence cannot be described by a single-valued function* like that in figure 2a, because the resistance

obviously *changes during the measurement*, depending on the treatment the crystal has been subject to. This conclusion is also confirmed by the final removal of the DLD in fig 4c! In this experiment we have in fact observed *both a temporal increase of the resistance, as well as a final decrease.*

Figure **5a** shows records of $V(|R_1|)$ and $V(I_x)$ for *a* 'good' crystal. The down-step when the *ALC*-circuit takes over control is very small and likely due to the time it takes to build up oscillations.

Figure **6a.** This crystal begins oscillating at relatively low $|R_1|$, at a 'second level of drive', at its fundamental frequency. Then, at a much higher level, it jumps to its third overtone. This is possible because the bandwidth of the oscillator is very wide, and allows operation on fundamental or on overtone, depending on which has the lowest resistance.

Passive measurements of the resistance of the same crystal, shown in figure 8, confirm that it has a high resistance peak of about 70 ohm at 1 μW at its fundamental, but not on its 3rd overtone; where R_1 is varying from about 40 ohm at low level to 30 ohm at normal level. This shows that *the resistance peak is frequency dependent.*

Figure **6b** shows one more record from a sequence of sweeps on the same crystal as in 6a. Here the output signal of $V(I_x)$ with enhanced low level response is utilized.

Figure 7a shows a record of a crystal with a similar behaviour as in figures 3 and 4. In this case the 'second level of drive' is almost as high as the chosen ALC-level.

Figure **7b** shows records for a crystal with high start resistance at very low current level, too low to be recorded. This is the same crystal as in 7a, after having been operated at 3 mA.

Figure **8** shows passive measurements in a IEC π-network of the resistance at fundamental and 3rd overtone for the same crystals used in figures 5a and 6a.

Discussion and Conclusions

The test oscillator

The oscillator is useful in research, in particular for investigation of time-dependent phenomena that change with the pre-history of the crystal, – cases when a reproducible resistance versus drive level characteristic does not exist. Combinations with other measurements, such as *thermal hysteresis, phase noise and aging* may shed light on possible correlations.

Further development of the circuit is highly desirable, so that resistance versus drive level characteristics can be recorded, both under rising and falling drive level.

The oscillator is useful in production; for production checking, and for screening, under computer control.

New results

From the few tests on 'bad' crystals we have done, we think we have got the following indications:

1. DLD-dependence can be quite different on the fundamental and third overtone modes of operation for the same crystal.
2. DLD cannot always be described as a single-valued function of resistance versus drive level, because the resistance (at a given level) may change during the measurement.
3. The effects on the crystal by oscillation may result either in an increase or a decrease of resistance.

Physics behind DLD-phenomena

Probably different physical mechanims are acting behind different manifestations of DLD. The importance of small particles attached to the surface by nonlinear forces is extensively discussed and well verified experimentally in the literature [3, 5 - 8]. Recently Kanazawa [9] discussed the effect of viscous films.

We think that the 'sleeping sickness' phenomenon could be due to a temporary partial breaking of binding forces between particles and surface at high vibration amplitude, bindings that later close again, maybe a process resembling the tixotropy phenomenon in certain liquids.

Suggestions for further improvements

In its present state of development the oscillator cannot produce a record of resistance as a function of crystal current, but it might be possible to obtain that function:

As soon as the *ALC*-circuit comes into action, the level is controlled by the reference voltage V_{REF}. At that moment another ramp voltage could be initiated and serve as V_{REF}. It could have both a rising and a falling ramp, thus allowing real *measurement of crystal resistance as a function of crystal current in* a range of maybe two orders of magnitude in current.

References

[1] Yerna, Y. C., 1984: Resistance measurements at very low drive levels . *Proc. 38th Ann. Freq. Contr. Symp.*, 499.

[2] IEC-Documenent: 49(Netherlands)81, 1989.

[3] Bottom, V.E., 1983: The behavior of quartz resonators at low drive levels. *Proc. 5th Quartz Crystal Conf.*, August 1983.

[4] Bottom, V.E., 1982: *Introduction to Quartz Crystal Unit Design.* van Nostrand Reinhold Publ. Co. NY.

[5] Bernstein, M. 1967: Increased crystal unit resistance at oscillator noise levels. *Proc. 21st Ann. Freq. Contr. Symp.*, 244

[6] Knowles, J.E., 1975: On the origin of the 'second level of drive' effect in quartz oscillators. *Proc.29th Ann. Freq. Contr. Symp.*, 230

[7] Dworsky, L., and Kinsman R.G. 1994: A Simple single model for quartz crystal resonator low level drive sensitivity and monolithic filter intermodulation. *IEEE Trans Ultrasonics, Ferroelectrics, and Frequency Control.*, **41**, 261

[8] Cross, J.A., 1979: Electrostatic effects in the adhesion of powder layers. In Mittal, K.L.(ed.) Surface Contamination Vol. , 89. Plenum Publ. Co. N.Y.

[9] Kanazawa, K.K. and Gordon II, J.G., 1985: The oscillation frequency of a quartz resonator in contact with a liquid. *Anal. Chim. Acta*, **175**, 99.

[10 Kanazawa, K.K. and Gordon II, J.G., 1985: Frequency of a quartz crystal microbalance in contact with a liquid.. *Anal. Chem.* **57**, 1770.

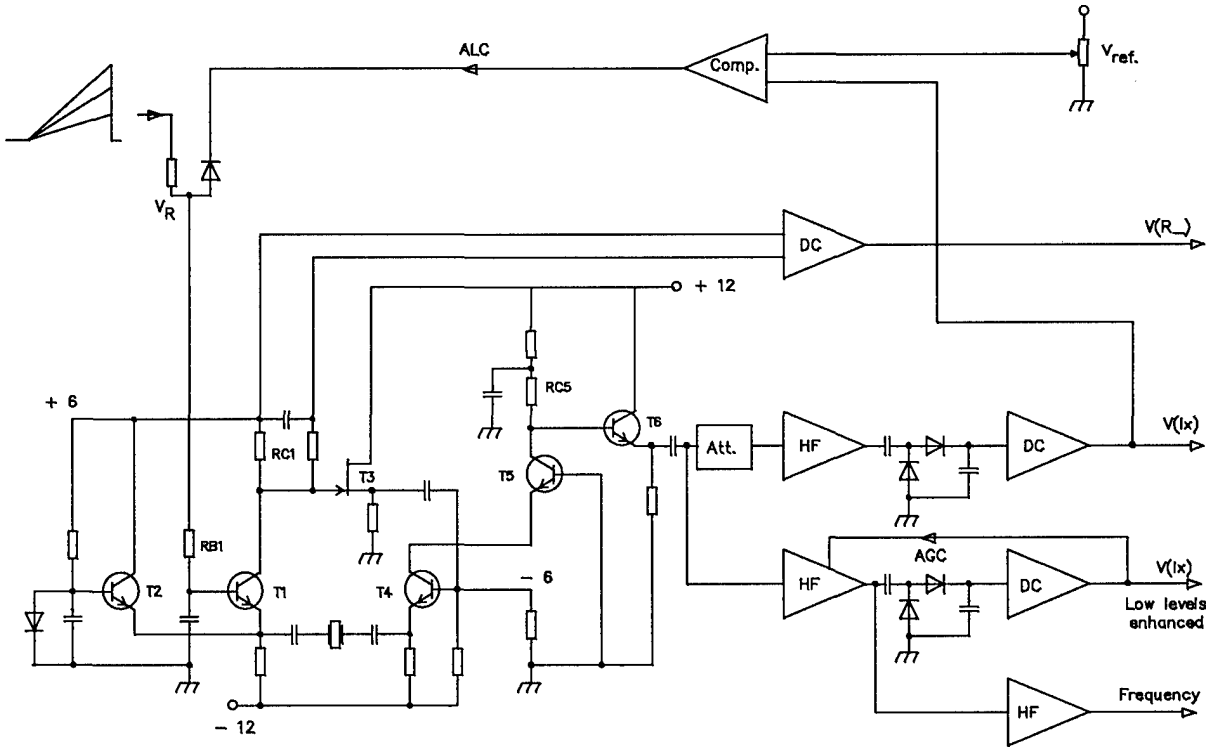


Fig.1 Schematics of the modified Butler–oscillator used for DLD characterisation of crystal resonators.

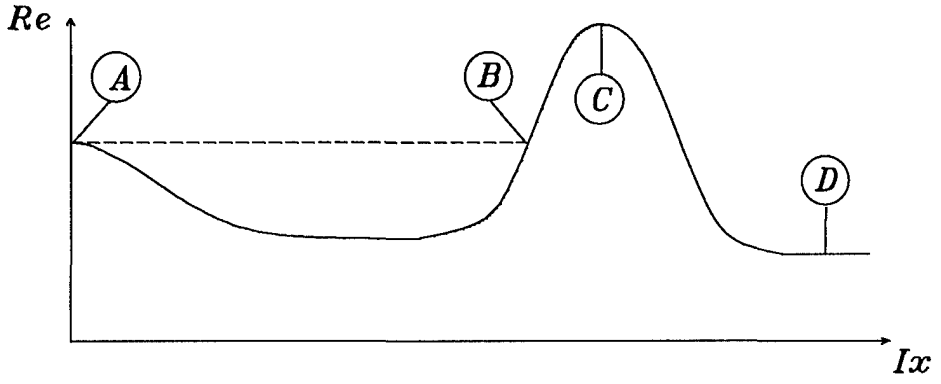


Fig. 2a Fictitious DLD–behaviour during start–up showing equivalent resistance of the crystal as a function of crystal current.

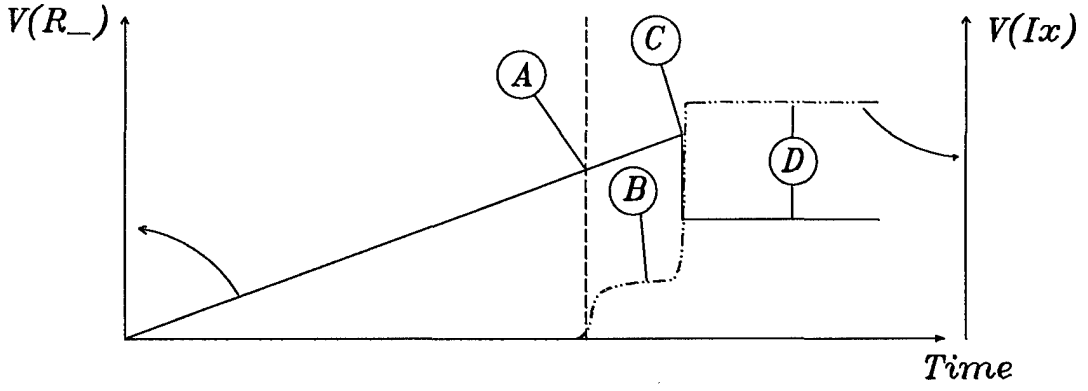


Fig. 2b Typical recording showing the negative resistance and crystal current versus time.

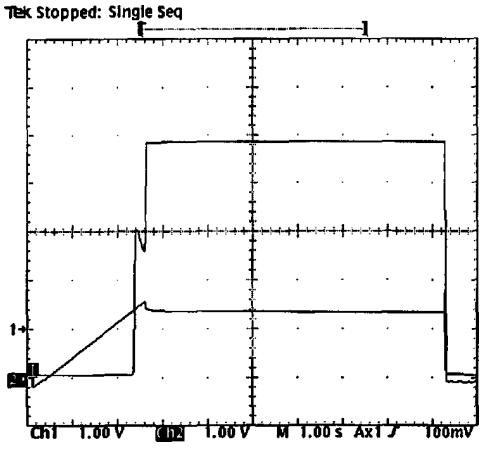


Fig. 3a First sweep showing a first start attempt rising to 3 div. and thereafter fall back somewhat before starting finally.

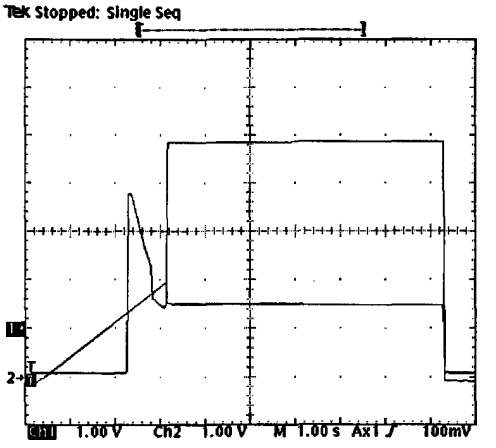


Fig. 4a First sweep of a crystal showing a pronounced first start attempt.

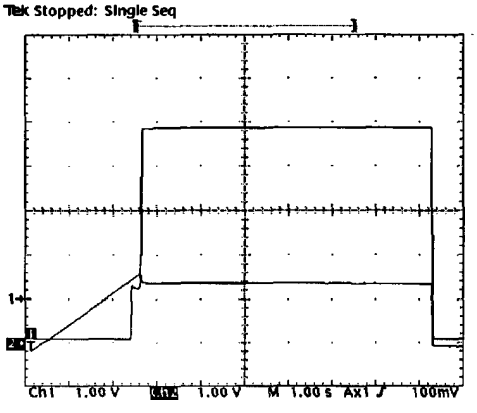


Fig.3b Second sweep of the same crystal as above, now with a slightly different appearance, rising about one div. followed by a smaller fall.

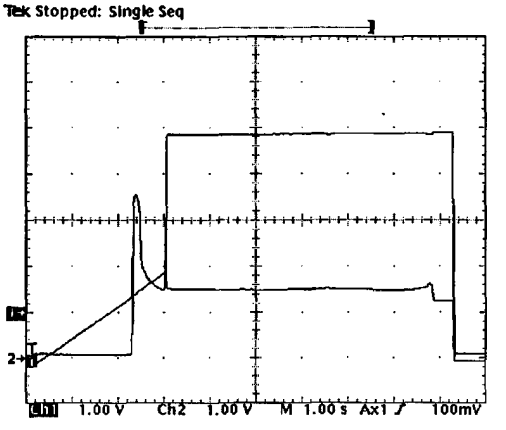


Fig. 4b Second sweep shows a slight difference in the start-up behaviour.

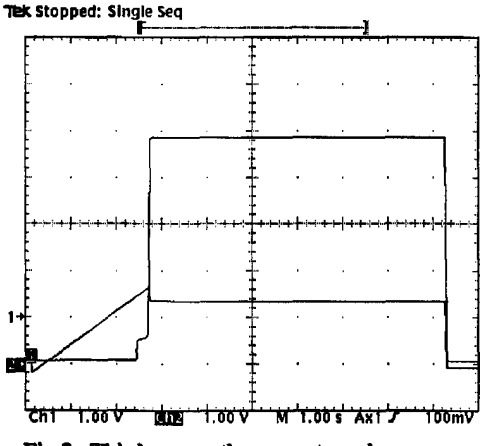


Fig.3c Third sweep, the current reaches only 1/2 div. and does not fall at all before the final start.

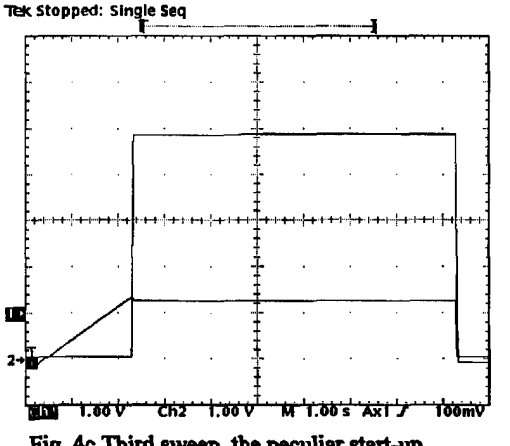


Fig. 4c Third sweep, the peculiar start-up behaviour has completly disappeared resulting in a "cured" crystal.

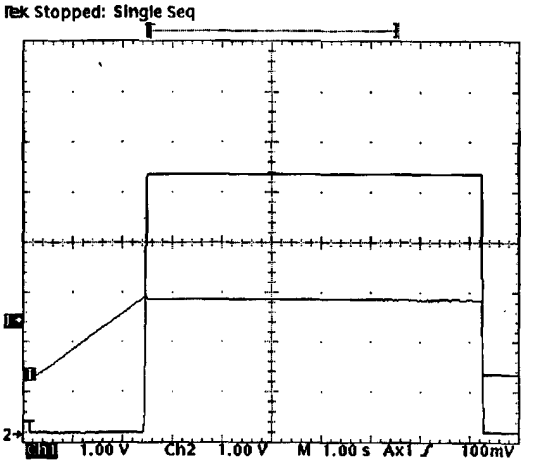


Fig. 5a A good crystal without any DLD effect at all.

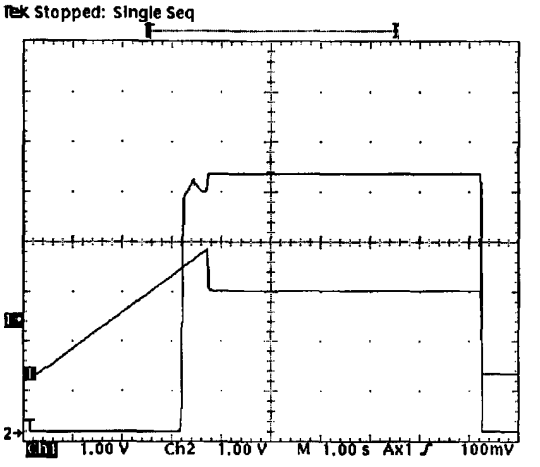


Fig. 7a Similar type of start-up behaviour as in fig 3 and 4.

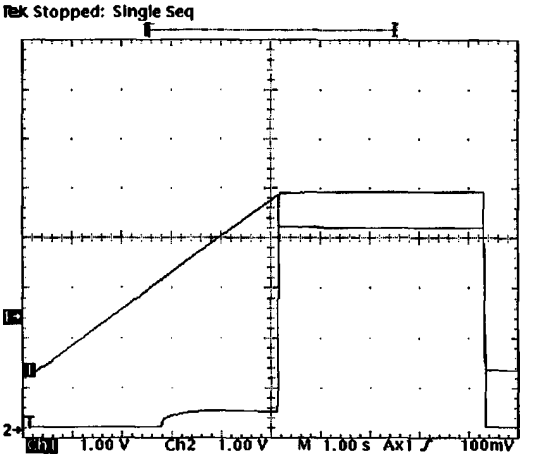


Fig. 6a Starting resistance of about 16 ohms then stuck at a "second level of drive" reaching 50 ohms thereafter dropping to 36 ohms.

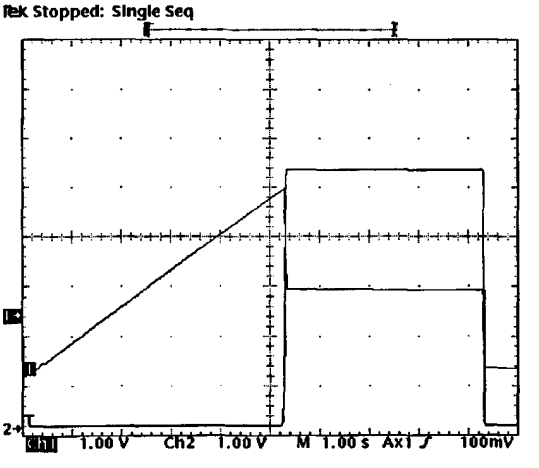


Fig. 7b Same crystal as above after a 3 mA sweep, resulting in two large changes, a higher starting resistance and a higher peak resistance.

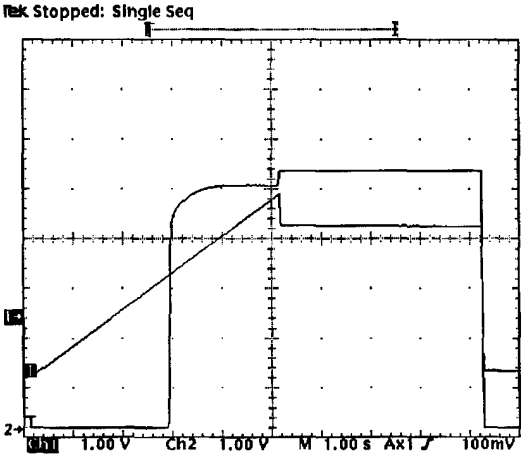


Fig. 6b Same crystal as above but utilizing the enhancement of low levels.

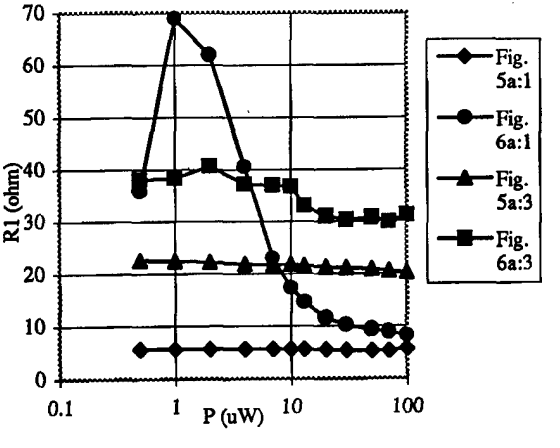


Fig. 8 Passive DLD-test in a pi-network.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

METHOD AND DEVICES FOR MEASURING DRIVING POWER OF QUARTZ CRYSTAL UNITS

M.E. Karaulnik, A.E. Karaulnik

Individual Private Enterprise Firm "Ligas", 107077, Moscow, 11th Parkovaya str. 30, Russia

Abstract

A brief review of the state of the problem of measuring the power dissipated across crystal units is presented. Simplifier formulas are deduced for the computation of the drive level of crystal units operating in oscillatory circuits. A new method of the power measurement and devices for its practical use are described.

1. Introduction

The problem of estimating the drive level of crystal units arises at once while designing crystal units and then at the stages of their technological adjustment, of the analysis of the influence of different destabilizing factors (temperature changes, component "aging", radiation and mechanical effects, etc.) upon oscillator and filter characteristics in operating conditions. Just for these reasons the problems above mentioned constantly attract the attenuation of designers and users of quartz crystal stabilizing devices. As result of various investigations a meaningful dependence was ascertained for crystal unit parameters (frequency, motional resistance, quality factor, spectral characteristics, etc.) upon their drive level. There are at least two qualitatively distinguished drive levels: a normal one (when at other equal conditions crystal unit parameters are the most stable and the drive level is commensurable in its value to the noise level of the measurement devices) and a high one (when unreversible changes of crystal unit parameters take place and reversible changes of these parameters in the form of a "hysteresis loop" as well, which are analogous to the characteristics of materials being magnetized). Reversible changes are changes, for example, of frequency and resistance reproducible in the limits of measurement errors after multiple measurement cycles. Unreversible changes of crystal unit parameters appearing after the high drive level effect are qualified in the IEC Technical Committee 49 Standard (august 1989) as DLD effect. While defining the drive level of crystal units, one should consider the drive power as the most spread and universal measure, though in some cases it is much more convenient to take into account the amplitude or the effective value of piezocurrent, the voltage across the crystal unit or even the amplitude of mechanical vibrations of any geometrical points of the piezoelement. Therefore, taking the power dissipated in a crystal unit as a general measure of the drive level we admit if it is necessary the usage of other possible measures. Below we give the deduction of simplified formulas for computing the power in crystal units applied in oscillatory circuits (parallel resonance circuits) and describe the supposed new method and its application.

2. Definition of power dissipated in crystal units operating as a component of a parallel resonance oscillator circuit

For oscillator circuits operating at the series resonance of a quartz vibrator the techniques of measuring the above mentioned power is simple for it is defined by the motional resistance of the crystal unit and the piezocurrent and slightly depends on the reactive parameters of the crystal unit (motional and static capacitances and motional inductance). In the case of oscillator circuits operating within the resonance spacing of the crystal unit, the power dissipated in the crystal plate substantially depends upon the reactive parameters, the motional resistance of the crystal unit and its frequency. Formulas for the power computation well-known from [1] for example are convenient enough for practical computation. In [2] it is recommended to compute the power value taking into account the circuit capacitance value. As it seems to the authors there may be applied a more accurate method.

Depending on the circuit type of the quartz crystal oscillator one may measure the current across the crystal unit or the voltage arriving across the crystal unit. Therefore it is necessary to deduce a formula for computing the power dissipated in the crystal unit using the current and voltage values.

With this aim in view we may use the usual interpretation of a crystal unit in the form of a equivalent circuit (Fig.1) [3].

0-7803-1945-1/94 $3.00 © 1994 IEEE

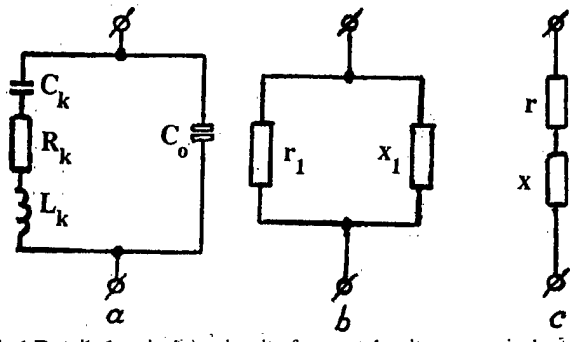


Fig.1 Detailedequivalent circuit of a crystal unit: a — equivalent circuit in detail; b — parallel connection of equivalent active and reactive resistances; c — series connection of the resistances (active and reactive).

To define the power dissipated across the crystal unit using the voltage across it we may apply the expression (known from the electrotechnique):

$$P = \frac{U_к^2 \cos^2 \varphi}{R_к}, \qquad (1)$$

where U_k - effective voltage value across the crystal unit;

$\varphi = \operatorname{arctg} \dfrac{\omega L_к - \dfrac{1}{\omega C_к}}{R_к}$ - phase angle of the motional arm of the crystal unit.

By using trigonometrical equality

$$\cos^2(\operatorname{arc\,tg}\,\theta) = \frac{1}{1+\theta^2},$$

we get

$$P_к = U_к^2 \frac{R_к}{R_к^2 + \left(\omega L_к - \dfrac{1}{\omega C_к}\right)^2} \qquad (2)$$

It is seen from (2) that for the computation of the dissipated power it is necessary to know the motional parameters of the crystal unit, the voltage across the quartz plate and the generated frequency. Formula (2) may be expressed by using the value of deviation of the oscillator frequency from the crystal unit series resonance frequency. After some simple transformations we get the following

$$P_к = \frac{U_к^2 R_к}{R_к^2 + \left(\omega L_к - \dfrac{1}{\omega C_к}\right)^2} =$$

$$= \frac{U_к^2 R_к}{\omega^2 L_к^2 \left(1 - \dfrac{1}{\omega^2 L_к C_к}\right)^2 + R_к^2} =$$

$$= \frac{U_к^2 R_к}{\omega^2 L_к^2 \left(\dfrac{\omega^2 - \omega_1^2}{\omega^2}\right)^2 + R_к^2}$$

Taking into account that $\omega_1 = \dfrac{1}{\sqrt{L_к C_к}}$ and

$$\frac{\omega^2 - \omega_1^2}{\omega^2} \cong \frac{2(\omega - \omega_1)}{\omega} = \frac{2\Delta\omega}{\omega},$$

where $\Delta\omega$ - deviation of the oscillator frequency from the series resonance frequency of the crystal unit;

ω_1 - series resonance frequency of the crystal unit;

ω - quartz crystal oscillator frequency,

we receive the expression

$$P_к = \frac{U_к^2 R_к}{16\pi^2 (L_к \Delta f)^2 + R_к^2} \qquad (3)$$

Really, if $\Delta f = 0$ (i.e. if the oscillator operates at the series resonance frequency of the crystal unit), then the expression defining the power takes form

$$P_к = \frac{U_к^2}{R_к} \qquad (4)$$

For the oscillator operating at antiresonance frequency of the crystal unit the power dissipated in the last is defined by the expression

$$P_к = U_к^2 R_к (\omega C_0)^2,$$

for which it was taken into account that

$$\frac{\omega_2 - \omega_1}{\omega} \cong \frac{1}{2} \frac{C_к}{C_0},$$

$$\frac{R_к^2}{\omega^2 L_к^2} = \left(\frac{1}{Q_к^2}\right) \ll \frac{C_к}{C_0},$$

where Q_k - quality factor of the crystal unit.

This formula, often recommended for the computation of the dissipated power across the crystal unit, operating in the parallel resonance circuit, in practice leads to a significant error. For defining the dissipated power using the value of the current across the piezoelement we may apply the formula

$$P_к = I_к^2 r, \qquad (5)$$

where I_k - the effective value of current across the crystal unit which equals to the sum of the piezocurrent and the current across the static

capacitance of the crystal unit C_o. The term „r" is included into the equality (Fig.1, c)

$$Z = r + jx, \qquad (6)$$

where $$r = R_к \frac{(1/\omega C_0)^2}{R_к^2 + \left(\omega L_к - \frac{C_к + C_0}{\omega C_к C_0}\right)^2}; \qquad (7)$$

$$x = -\frac{1}{\omega C_0} X \qquad (8)$$

$$X \frac{\left[R_к^2 + \left(\omega L_к - \frac{1}{\omega C_к}\right)^2\right] - \frac{1}{\omega C_0}\left[\omega L_к - \frac{1}{\omega C_к}\right]}{R_к^2 + \left(\omega L_к - \frac{C_к + C_0}{\omega C_к C_0}\right)^2}$$

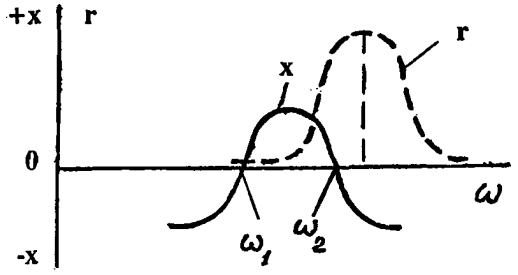


Fig.2. Active and reactive resistances of the crystal unit versus the frequency (ω_1 - series resonance frequency, ω_2 -parallel resonance frequency)

In Fig.2 curves are given for the dependance of r and x values upon the frequency. The substitution of (7) into (5) gives the expression

$$P_к = I_к^2 R_к \frac{(1/\omega C_0)^2}{R_к^2 + \left(\omega L_к - \frac{C_к + C_0}{\omega C_к C_0}\right)^2} \qquad (9)$$

If we take into account (as for the deduction of (3)) that

$$\frac{\omega^2 - \omega_1^2}{\omega^2} \approx \frac{2\Delta\omega}{\omega}$$

and - $\frac{C_к}{C_0} \ll 1$,

we receive the expression

$$\left(\omega L_к - \frac{C_к + C_0}{\omega C_к C_0}\right) \approx \frac{1}{\omega C_к} \frac{2\Delta f}{f} \qquad (10)$$

The substitution of (10) into (9) leads to an expression which allows to compute the power dissipated across the crystal unit by using its current value

$$P_к = \frac{I_к^2 R_к}{(R_к \omega C_0)^2 + \left(\frac{2C_0}{C_к} \cdot \frac{\Delta f}{f}\right)^2}, \qquad (11)$$

where $\Delta f = f - f_1$ — the deviation of the oscillator frequency from the series resonance frequency of the crystal unit.

So, to define the power dissipated across the crystal unit by using the value of the current across the crystal unit, we need the following values: R_k, C_o, C_k, f_1, f and I_k, where f_1 - the series resonance frequency of the crystal unit, f - the oscillator frequency, I_k - effective value of the current across the crystal unit.

There exist several methods of measuring motional parameters and static capacitance of crystal units. Therefore only a method of measuring current across the crystal unit (or voltage across it) will be considered below. Direct connection of the measuring apparatus with the oscillator circuit lines for the definition of the voltage or the current at its points is accompanied by meaningful errors in the measured values for standard HF-meters have low values of the input impedance. Besides that the power dissipated across the crystal unit strongly depends on the operating frequency, i.e. on the value of the deviation of the generated frequency from the series resonance frequency of the crystal unit. In fig.3 the dependance is given [4] of the coefficients P_{KU}, P_{KI}, which characterize the power dissipated across the crystal unit for given current I_k or the voltage Vk upon the mistuning η;

$$\eta = \frac{\omega_2 - \omega}{\omega_2 - \omega_1},$$

where ω_1 - angular frequency of the crystal unit series resonance; ω - crystal oscillator frequency.

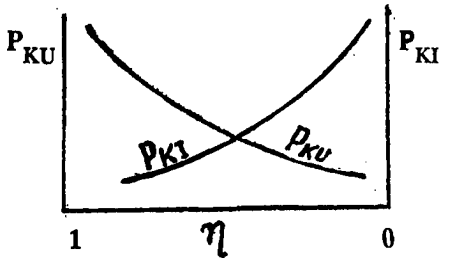


Fig.3 The dependance of coefficients defining the power dissipated across the crystal unit at U_k = const or I_k = const upon the frequency.

It follows from Fig.3 that the measurement of voltages across the oscillator components with the aim in view of defining current (or voltage) by connecting meters to any circuit points (Fig.4) must not be accompanied by significant changes of the oscillator frequency or its operating conditions (regime)

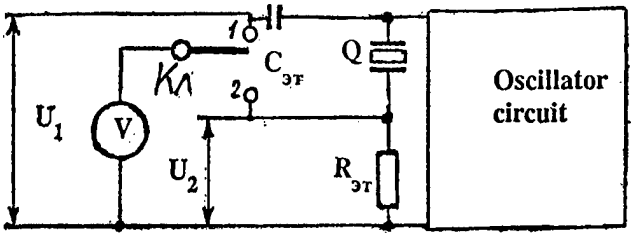


Fig.4 Circuit for measuring the voltage and current of crystal units.

In Fig.4 there is given a scheme of connecting meters for defining the current or voltage, where $R_{эт}$ — the known resistance having a value much (10÷100 times) less than the motional resistance of the crystal unit; $C_{эт}$ — the know capacitance (1pF); V — HF-millivoltmeter. The voltage across the crystal unit according to the designations in Fig.4 may be expressed as

$$U_к = \frac{U_1}{x_{c_{пр}}}(x_{c_{эт}} + x_{c_{пр}}), \qquad (12)$$

where U_1 - millivoltmeter reading in the position 1 of the selector "$K_Л$"

$x_{Cэт}$ — the resistance of the capacitor $C_{эт}$

$x_{Cпр}$ — the resistance of the input capacitance of the meter.

The current value is defined with the selector "$K_Л$" in position 2. Taking into account that $R_{эт} << |x_{Cпр}|$, we get

$$I_к = \frac{U_2}{R_{эт}} \qquad (13)$$

The values of U_k and I_k from (12) and (13) may be applied for the computations on the basis of (3) and (11). The method above described allows to measure the voltage or the current of the crystal unit only by using such oscillatory circuits in which one the crystal unit leads may be grounded.

In the case when the crystal unit is used in a capacitance or inductance three-point circuit and has no lead directly grounded, formula (3) should be used. Then the voltage U_k is defined as the sum of voltages across the crystal unit electrodes relative to the oscillator case, and this fact does not practically influence upon the error of the definition of the voltage dissipated across the crystal unit. Here the definition error for U_k achieves not more than 10% if the mistuning value of the crystal unit relative to the series resonance $\Delta f/f$ is not less than $1x10^{-6}$.

Simultaneously the accuracy of the power estimation from (3) and (11) at 10 kHz, 100 kHz and 4 Mc was analyzed. In all the cases the results showed a discrepancy not exceeding 20%. It follows from the above mentioned, that formulas (3) and (11) may be recommended for being practically applied to estimate voltage dissipated across crystal units operating in oscillatory circuits.

3. Universal means of measuring crystal unit drive level and device for its practical application

Up to now the estimation of the crystal unit drive level by using the power value is realized through the computation methods applying the results of measuring voltages, currents and resistances of passive and oscillator circuits for exciting crystal units. IEC - standards contain general methods for the corresponding measurements. Yet these standards use the term "power" implying traditionally and synonymously the power of heat losses in crystal units. But the investigations carried out by the authors showed that piezoelement vibrations are accompanied by direct and indirect HF piezoelectric effects. When the outer electromagnetic field "plays" closely to the piezoelement, then the last in its turn "excites" a new orthogonal electromagnetic field beside. This one leads to a new piezoeffect, which one in its turn excites one more electromagnetic field. And this process continues up to the full attenuation of the direct and indirect effects. The attenuation may take place during one vibration period, the second one, the third one, etc. Mathematically it may expressed by applying Maxwell equations [5]. The fact that the nature of this phenomena is directly connected with the properties of the electromagnetic field is confirmed by the well known effect of crystal unit luminescence if they are excited by a strong piezocurrent [6]. The electromagnetic nature of the light eradiation has been proved.

By admitting the supposition that for various crystal unit types the relations between thermal losses and losses due to electromagnetic eradiation of the excited piezoelement follow the strict laws, the authors carried on investigation with the aim in view of designing a new method of measuring crystal unit drive level and accordingly a device for its practical application. Crystal units under investigation were excited in standards oscillators insuring stepped or smooth control of the excitement power. Then calibration of the excitement power of the crystal unit in the scheme of the technological (standard) oscillator was carried out by using the reading of HF-field indicator received by applying special pickups. After that, applying the calibration table, the dependence of

piezoelectric crystal unit parameters on their drive level was analyzed. As it is clear from the above described, by using the method under discussion the authors practically in whole succeeded in decising the problem of measuring the power dissipated across the crystal unit without the "interference" of meters into the oscillators circuit.

As standard oscillators the authors used specially designed for this aim universal control-measuring oscillators UKIG-1A, UKIG-1B, UKIG-2.

General technical characteristics of these oscillators are given below (Table).

4. Conclusion

The method of measuring the power dissipated across the excited crystal units and the devices for its practical application which were designed by the authors in the result of the investigation and experiments, allowed significantly improve the accuracy of controlling the power (from 20 up to 5÷7%). The application of the method and the device while optimizing crystal units gave good results in:

- designing crystal units with frequency pulling[7];
- designing crystal units with lens piezoelectric elements at 1 and 5 Mc [8];
- improving the perspective means of producing piezoelectric crystal units having a piezoelement of BVA configuration [9];
- designing subminiature and other types of crystal units [10].

References (Russian interpretation)

1. Jabotinskiy M.E., Zilberman P.E. Comments on the frequency dependence of crystal oscillators upon the power dissipated across the crystal unit. "Radiotechnika i elektronika", 1958, t.3, №2, s.275-277.
2. Quartz pour Oscillareurs. I.E.C. recommendations. Publications 122-1 and 122-2.
3. Groshkovskiy Y. Excitement of HF-vibrations and frequency control. Moscow: Izd-vo Inostrannay literatury, 1953 (transl. from Polish).
4. Magazanik A.A. VXO and exciters for the frequency telegraphing. Moscow: Svyaz-izdat., 1955.
5. John R. Pierce. Electrons, waves and messages. Honovere House, Garden City, New York, 1956. Moscow: Gosudarstvennoe izdatelstvo matematicheskoy literatury, 1961, s.108-132.
6. Smagin A.G. Piezoelectric crystal units and their application. Izd-vo Komiteta standartov mer i izmeritelnyh priborov pri SM SSSR. 1967.
7. Karaulnik A.E., Karaulnik M.E., et.al. Avtorskoe svidetelstvo SSSR № 349078 s prioritetom ot 25 yanvarya 1971.
8. Karaulnik A.E., Karaulnik M.E. Dependence of quartz piezoelement quality factor upon the material and the form of the exciting electrodes. "Elektronnaya technika", Ser.5: Radiodetaly i komponenty. Vyp.4, 1974, s.44-51.
9. Karaulnik A.E., Karaulnik M.E., et. al. Avtorskoe svidetelstvo SSSR № 1542379 s prioritetom ot 8 oktyabrya 1989.
10. Karaulnik A.E., Grouzinenko V.B. Researches aimed at designing a miniature ruggedized configuration of BVA-type crystal unit. Proc. 47th ASFC, p.p.548-559, 1993.

Table

Operating Characteristics	UKIG-1	UKIG-2
Operating frequency range	5÷100 kHz (UKIG-1A) 100÷500 kHz (UKIG-1B)	0,5÷70 Mc
Reproducibility of the excitement frequency (for crystal unit quality factor ≥50000):		
approximate indicator	$\pm 1 \times 10^{-6}$	$\pm 1 \times 10^{-6}$
accurate indicator	$\pm 0.1 \times 10^{-6}$	$\pm 0.5 \times 10^{-6}$
Frequency shift relative to the series resonance frequency	$\pm 2 \times 10^{-6}$	$\pm 1.5 \times 10^{-6}$
Dissipated power	10; 20; 25; 40; 50; 100 200; 250; 400; 500; 1000 μWt (stepped control)	0.01÷1.0 mWt (smooth control)
Measurement error for equivalent motional resistance	≤10%	≤10%
Measurement error for other equivalent motional parameters of crystal units	≤10%	≤20%

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

IMPLEMENTING ISO9000 IN A SMALL COMPANY

Errol P. EerNisse and Clifford D. Mercer

Quartzdyne, Inc.
1020 Atherton Drive Bldg. C
Salt Lake City, UT 84123

Abstract

Companies in the frequency control field, particularly those involved in international sales, are under increasing pressure to certify to ISO9000 quality system standards. Since these standards only define a general guideline for quality systems, they allow some latitude in satisfying the requirements. This latitude allows small companies as well as large companies to obtain certification to ISO9000. We report here our efforts as a small company in moving towards ISO9000 certification and some interesting side benefits obtained from the exercise.

Introduction

Quartzdyne, Inc. changed corporate culture several years ago and moved over a two year period from strictly R&D to production and sales. The R&D was focused from the company's inception in 1979 on sensor technology based on bulk quartz resonators. In 1991 production and sales began on a line of precision pressure transducers for down-hole applications in the oil and gas industry. As sales grew, the broad-scope R&D activities diminished as the demand for manpower in production diverted efforts from R&D to production.

After initial production startup, it became quickly apparent that a quality control system for operation of the company was needed in order to improve the reliability and durability of our products. The down-hole environment is very harsh; temperatures can exceed 175^{o} C, shock and vibration exposure is high, and corrosive liquids and gases abound. Furthermore, the cost of using a transducer down-hole is large because of the expense of shutting down drilling or production activities while inserting or removing the measurement tools containing the transducer. Reliability is paramount because of these costs: typically the measurements involve no real-time data acquisition (memory tools powered by batteries are common) so a transducer failure is not detected until after the complete expense of the test has been incurred.

An additional factor driving the installation of a quality control system was the rapid growth of the sales (doubling several years in a row). Yield needed to improve to avoid large increases in personnel. Also, delivery times are important to the customers since new wells and new oil fields require additional measurement tools on a sporadic basis and long-range, planned delivery schedules for transducers is not always practical. Thus, backlog times needed to reduce.

Finally, 85% of the sales of transducers typically are outside the U.S. The European portion of sales, where ISO9000 certification is becoming prevelant faster than in the U.S., is about 60%.

All of the above motivations led us to establish a production control system with ISO9000 in mind as a future goal. Study of ISO9000 documents led to the conclusion that the overall concept of an ISO9000 quality system is an efficient and professional way to run a company. Since there is lattitude in the ISO9000 guidelines, a small company such as Quartzdyne, Inc. could aspire to obain ISO9000 certification.

A few words about ISO9000. Registration does not certify a specific product or guarantee quality. It does certify the quality system and ensure that the quality system provides for continuous quality improvement. Revised standards are under review at this time and are expected to be approved world-wide mid-1994.[1] As an example of the growth of the number of companies obtaining certification in the U.S. , the first quarters of 1992, 1993, and 1994 saw the number of U.S. companies obtaining certification grow from 279 to 1259 to 3165, respectively.[1] There is a fast growing library of information for help in obtaining certification.[1-5] In addition, seminars and courses abound; many current ones are listed in reference [1].

0-7803-1945-1/94 $3.00 © 1994 IEEE

Basic Philosopy

Perhaps the first and most important decision to make is a firm commitment from the top of the management structure to implement a quality system that can be certified to ISO9000. This is because sufficient resources must be allocated not only to implement the system but to maintain the system. ISO9000 certification is on-going, internal audits as well as an external audit by the certifying registrar must occur annually. A plan for implementing a quality control system must be developed. Every employee in the company must become involved.

The quality system must measure quality and monitor improvement. The intent is to create an atmosphere of problem prevention, not problem fixing. This is exemplified in the proposed 1994 ISO9000 Standards where the subject of Preventive Actions is added to complement the subject of Corrective Actions.

Our First Steps

Our first task was to document what was already being done. The lattitude allowed in ISO9000 permits the company to define what is necessary and appropriate for its products in all categories of the standard. Review what is already being done to see if it makes sense, or is superfluous. Then plan to add only what makes sense for your specific operations in each category of the standard. Do not plan to implement anything that you will not actually do. A key part of an ISO9000 audit is to check if you are doing what your documents say you are doing.

In our case, by the end of 1992, a part numbering system had been established, bills of material were developed, and control drawings and manufacturing procedures had been formalized. A document control system was in place to insure that the correct information was always provided to both in-house operations and outside vendors, even during product improvement phases. This included engineering change orders to document design changes. The calibration facility for verifying the performance of our products was documented to be traceable to the National Institute of Standards and Technology.

A computerized Management Information System (MIS) was purchased early in 1993. The first tasks involved incorporating what had been done earlier by hand or with computer databases into the MIS. This involved the part numbers, bill of materials for all parts, control drawings and manufacturing procedures. An item master data base was transferred to the MIS and a vendor data base was created. These measures established a system of configuration control and lot tracking of components and sub-assemblies used in individual products i.e., traceability. All materials, subassemblies, procedures, and inspection steps are recorded for each serial-numbered product shipped.

Procedures were established to separate non-conforming material from the production flow. Non-conforming material reports from the production flow as well as from customer returns were implemented.

After review of the ISO9000 Standards and a variety of related literature, a Quality Manual and Operating Procedures were written to define the existing quality system.

What's Left To Do Before Auditing

During 1993, sales doubled and the pressure of producing product with a reasonable backlog slowed our progress in implementing the quality system. In early 1994, personnel were added to ensure that the quality system implementation proceded at a faster pace. Review of what had been done was provided by a QC engineer experienced in ISO9000 auditing. Several of the categories already established needed improvement and a number of additional categories required implementation.

The vendor assessment and purchase control documents are being improved by formalizing in-house workmanship and inspection standards for incoming vendor-supplied materials, components, and sub-assemblies and by broader reference to national and international standards. The non-conforming material reports are being incorporated into a database for routine sorting to facilitate identification of patterns that can be addressed by preventive actions. Additional instrument calibration procedures and documentation are scheduled. Documentation of employee training for in-house procedures is the last category scheduled.

After all the elements are in place, the existing Quality Manual and Operating Procedures will be revised to clearly define what we are doing.

Summary

Using the ISO9000 quality standards is simply a better way to run a company. Because of the latitude allowed, small companies can minimize the overhead of implementation and maintenance with carefully thought out controls and procedures. A clear benefit for a small company is that the company now has an operating system that decentralizes control, which minimizes the impact due to the loss of a key individual.

References

1. Quality Systems Update, Vol. 4, Number 5, May 1994 (published by CEEM Information Services, 10521 Braddock Road, Fairfax, VA).

2. American National Standard ANSI/ASQC Q90-1987 to Q94-1987 and related documents, American Society for Quality Control, 611 East Wisconsin Avenue, Milwaukee, Wisconsin 53202.

3. Questions and Answers on Quality, the ISO9000 Standard Series, Quality System Registraion, and Related Issues, NISTIR 4721, U.S. Department of Commerce, National Institute of Standards and Technology, Standards Code and Information Program, Office of Standards Services, Gaithersburg, MD 20899.

4.ISO9000 NEWS, ISSN 1018-6638, ISO Central Secretariat, 1 rue de Varembé, Case postale 56, CH-1211 GENÈVE 20, Switzerland , or contact American National Standards Institute, 11 West 42nd Street, New York, NY 10036.

5. Jack Kanholm, ISO EXPLAINED--AN INTERPRETATION GUIDE, AQA Co., 334 Crane Blvd., Los Angeles, CA 90065.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SAW DEVICES ON LITHIUM TETRABORATE ($Li_2B_4O_7$) (INVITED)

Hidenori Abe, Masashi Ohmura and Hisatoshi Saitou

JAPAN ENERGY CORPORATION
O&E Components Division
3-17-35 NiizoMinami, Toda-shi, Saitama 335, Japan

Abstract

A 45° rotated X cut, Z propagation $Li_2B_4O_7$ substrate has attracted attention as a piezoelectric substrate for SAW devices. The substrate is advantageous not for SAW transversal filter but for SAW resonator and resonator filter applications. It is especially suitable for SAW resonator filters which have the relative bandwidth of 0.3% to 0.6% considering coupling coefficient and temperature coefficient of delay.

Also, the chip size of SAW resonator filters can be reduced because of the large reflection coefficient for aluminum strips on $Li_2B_4O_7$ substrates.

SAW resonator filters which have low insertion loss, high rejection level and small size are reviewed and described for pager, CT1, GSM and cordless telephone receiver applications.

Moreover, the fabrication process for $Li_2B_4O_7$ substrates is also mentioned.

Introduction

Lithium tetraborate ($Li_2B_4O_7$: abbreviated as LBO) has been discovered by R.W. Whatmore et al. It is an attractive piezoelectric material for SAW devices, which has zero temperature coefficient of delay, TCD, and fairly large coupling coefficient, k^2 [1-3]. Since then, many intensive efforts on single crystal growth [4-7], SAW and bulk wave properties [5,8-21], piezoelectric properties [5,20,22], device applications [23-26] and fabrication processes [19,24,25] of LBO substrates have been made.

Rayleigh wave propagation property on LBO substrates has been investigated by Y. Ebata et al., Fukuta et al., T. Shiosaki et al., M. Adachi et al. and W.S. Ishak et al.

Leaky SAW propagation property in LBO substrates has been reported by M. Adachi et al. and M. Murota et al.

BGS wave that is shear horizontal wave propagating below the surface of LBO substrates has been investigated by W. Soluch, and T. Sato et al.

Parameters of electrical equivalent networks for SAW metallic gratings, such as the reflection coefficient for aluminum strips, on LBO substrates have been estimated by Y. Ebata et al., M. Takeuchi et al., M. Koshiba et al. and H. Abe et al.

The fabrication process of SAW devices on LBO substrates is reported by W.S. Ishak et al., S. Matsumura et al. and H. Abe et al.

The SAW resonator and resonator filter on LBO substrates have been fabricated by Y. Ebata et al. and M. Ohmura et al.

This paper reviews and describes SAW properties, the fabrication process and SAW device applications of LBO substrates.

SAW Properties of LBO Substrates

Substrate cut angle and propagation direction

The SAW velocity, Vf, k^2 and TCD are major factors that determine the SAW device performance.

R.W. Whatmore et al. have reported an X cut, Z propagation LBO substrate that has zero TCD and k^2 of 1.2%. The curve of the change in specific time delay takes a parabolic form. The second order temperature coefficient is $310 \times 10^{-9}/°C^2$ [1-3].

Y. Ebata et al. have described that Vf, k^2 and TCD are accomplished over the full cuts and propagation direction using material constants and temperature coefficients reported by N.M. Shorrocks et al. and using aluminum(Al) constants by P.M. Sutton. While rotating the cutting and propagating angles, Vf varies from 3056m/s to 4320m/s, k^2 from 0% to 1.37% and TCD in a free surface from -9ppm/°C to 200ppm/°C. The (90°+18°,90°,90°), (90°,90°+8°,90°) and (90°,90°,90°+8°) orientations in Eulerian angle have zero TCD on a free surface. They also have k^2 of 1%, 1.2% and 1%, respectively. However, TCD for the (90°+18°,90°,90°) orientation LBO substrate changes to -15ppm/°C when the surface of LBO substrates is covered with an infinitely thin Al film. As the rotation angle increases, the turnover temperature becomes low. Also, when the surface of LBO substrates is covered with Al strips of the metalized ratio of 1.0 whose thickness normalized by wavelength is 0.01, the (90°+45°,90°,90°) orientation LBO substrate has zero TCD of the resonance frequency at room temperature. It is called as the 45° rotated X cut, Z propagation LBO (abbreviated as 45° X-Z LBO) substrate [8,24].

T. Shiosaki et al. and M. Adachi et al. have reported that all dielectric, elastic and piezoelectric constants, and their first

0-7803-1945-1/94 $3.00 © 1994 IEEE

and second temperature coefficient, are determined by measuring resonant and antiresonant frequency. Therefore, using their values, Vf, k^2 and TCD of Rayleigh and Leaky waves are calculated by computer over the full substrate cut angles and propagation directions. For Rayleigh wave, Vf varies from 3100m/s to 4300m/s, k^2 from 0% to 1.6% and TCD from -20ppm/°C to 110ppm/°C. Around 45° of λ in the (90°+λ,90°,90°) orientation, the TCD tends to zero, and values of k^2 and TCD have little λ angular dependence where λ is an Eulerian angle [20-22].

Reflection coefficient for Al strips

Y. Ebata et al., M. Koshiba et al., M. Takeuchi et al. and H. Abe et al. have estimated the reflection coefficient for Al strips on a 45° X-Z LBO substrate. Fig.1 shows the SAW reflectivity dependency of Al film thickness reported by Y. Ebata et al. where the broken line represents the measured value. The reflectivity on 45° X-Z LBO substrates is at least several times greater than that for ST quartz. The reflection coefficient is remarkably larger than that of any other reflector constructions previously reported. Therefore, the chip size of SAW resonator and resonator filter can be reduced because of large reflectivity.

Y. Ebata et al. have also investigated the reason for its high reflectivity, using the coupled mode theory and the perturbation theory. It is predicted that high reflectivity for Al strips is due to the large displacement component parallel to SAW propagation direction on LBO substrates [23].

Consequently, a 45° X-Z LBO substrate is suitable to fabricating small size SAW resonator filters with the fractional bandwidth of 0.3% to 0.6%, considering the coupling coefficient of 0.8%, zero temperature coefficient of delay and large reflectivity.

Fabrication Process for LBO Substrates

LBO substrates are well known to dissolve in water and acid solution that is usually used in SAW device fabrication process. Therefore, Al electrodes on LBO substrates cannot be fabricated by the conventional Al etching process that requires acid etchants. Fig.2 shows the dependence of the dissolution rate of LBO substrates on pH in a solution. The dissolution rate on LBO substrates is inversely proportional to pH in a solution. Also, LBO substrates do not dissolve in organic solvents [5,19,25].

The assembly process is similar to that of high frequency SAW devices, using $LiTaO_3$ or quartz substrates, but has been slightly altered in consideration of LBO crystal's weakness against moisture.

Two-layer lift-off process

In two-layer lift-off process, the upper and lower resists are exposed by UV and Deep UV light, respectively. Therefore, LBO substrates come in contact with only organic solvent and do not dissolve during two-layer lift-off process.

The resonant frequency deviation of LBO substrates is evaluated using one-port SAW resonator operating at 90MHz that is fabricated by two-layer lift-off process. In a 45° X-Z

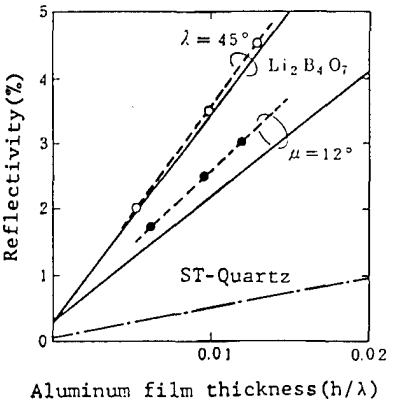


Fig. 1. Reflectivity dependency on aluminum film thickness. (From Y. Ebata et al. [23])

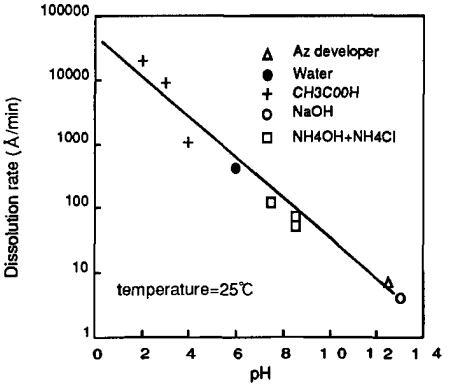


Fig. 2. Dependence of the dissolution rate of $Li_2B_4O_7$ substrates on PH.

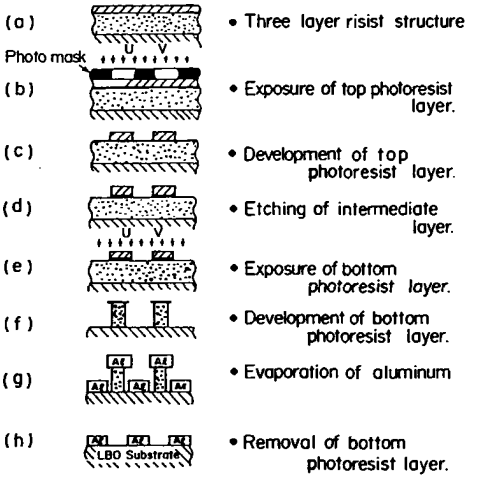


Fig. 3. Structures and fabrication procedures of the three layer lift-off process.
(From S. Matsumura et al. [24])

LBO substrate, the resonant frequency deviation is evaluated at ±310ppm. The result demonstrates the stability of the fabrication process and highly reproducible crystal growth [25].

Three-layer lift-off process

There are several problems, such as the accuracy of linewidths and reproducibility, on the conventional lift-off process. Therefore, S. Matsumura et al. have reported that the three-layer lift-off process using an Al intermediate layer has been established for patterning Al electrodes on LBO substrates. The structure and fabrication procedures of the three-layer lift-off process are shown in Fig.3. SAW devices, such as a front end SAW filter for a paging receiver, as described later, have been produced with high yield and high reliability by the three-layer lift-off process and improved assembly process.

The fractional frequency and minimum insertion loss changes lie within ±280ppm and ±0.3dB, respectively, after exposure to high humidity of 90% for over 1000 hours at 60°C. Other reliability tests, such as high temperature exposure test and DC operating test under temperature cycles show similar results [24].

Alkaline etching process using single layer resist

It is well known that Al metal dissolves in alkaline solution. However, alkaline etchants, such as a solution of NaOH or NH_4OH, break down the profile of the resist pattern. It causes the destruction of Al electrodes with fine width. Therefore, new alkaline etchant of pH=12 has been developed, which keeps the profile of the resist and is able to etch Al metal effectively. The etching rate of new alkaline etchant is 300 Å/min for Al metal and is below 3Å/min for LBO substrates as shown in Fig.2. Therefore, an alkaline etching process using single layer resist has been established for patterning Al electrodes on LBO substrates. Fig.4 shows the structure and fabrication procedures.

SAW Devices on LBO Substrates

One-port SAW resonator

Y. Ebata et al. have pointed out that conventional configuration resonators, made of Al strips on LBO substrates, have revealed poor unloaded Q. The reason is considered to be the bulk mode conversion loss at the grating reflector edge. Because radiated bulk waves cancel each other for infinite perturbation of a half-wavelength periodicity, intrinsic radiation bulk wave loss is confined to the immediate vicinity of the grating edges. The conventional configuration requires a distance between the IDT and reflectors to be $(N/2\pm1/8)\lambda$ where λ is wavelength. The mode conversion loss of the resonator concentrates around the boundaries, which are out of a half-wavelength periodicity.

To improve the resonator Q, the distance between the IDT and reflectors is set to be a half of the wavelength for suppressing the radiation bulk wave and the periodicity of the IDT electrodes is changed a little for tuning to the resonance condition of the grating reflector cavity. The structure is called QARP (quasi-constant acoustic reflector periodicity) where the periodicity of IDT electrodes and reflecting strips including the distance between them coincides with almost a half wavelength all over the resonator. Y. Ebata et al. have reported that the unloaded Q's are successfully improved to more than 15000 for 60MHz and 310MHz resonators, and are comparable to those for quartz and $LiTaO_3$ resonators [23].

SAW resonator filter consisting of two IDTs and two reflectors

S. Matsumura et al. have described a SAW resonator filter with QARP configuration. Cascading the two-port resonator can easily form a multi-pole resonator filter, called an IDT coupled SAW resonator filter. Increasing the number of IDT

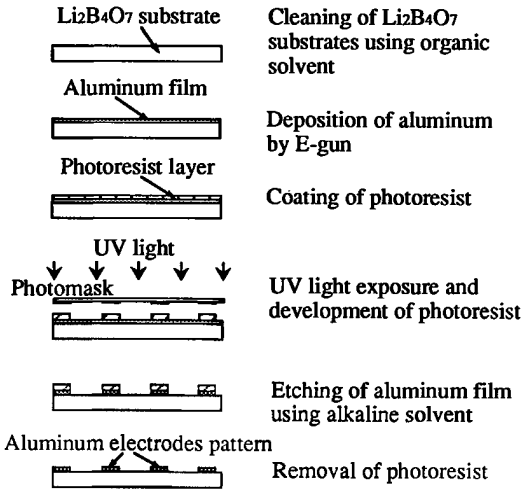


Fig. 4. Fabrication procedures of the alkaline etching process for $Li_2B_4O_7$ substrates.

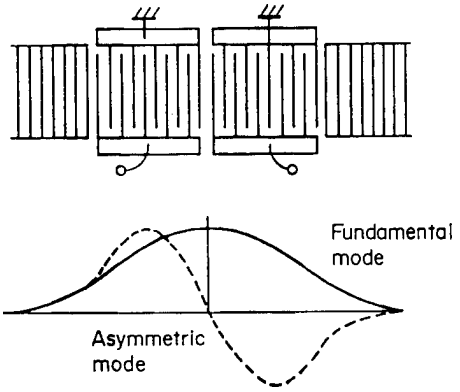


Fig. 5. Arrangement of IDT for the conventional 2-port resonator and its charge distributions of fundamental and spurious asymmetric modes. (From S. Matsumura et al. [24])

electrodes, however, causes higher longitudinal adjacent resonance modes for the grating reflector cavity. Charge distributions in the resonator for each response are analyzed as shown in Fig.5. Optimum IDT pitch and separation on LBO substrates can couple the modes and form a wider passband of the SAW resonator filter by the effect of waves reflected from the electrodes of IDTs. Connecting three resonator filters with QARP configuration in cascade, a front end filter for paging receiver has been successfully constructed on LBO substrates. Fig.6 shows a frequency response of a fabricated SAW filter, whose center frequency, bandwidth and minimum insertion loss of the fabricated SAW filter are 281.65MHz, 900kHz and 2.5dB, respectively. The image suppression level is more than 50dB. The center frequency drifts only 50kHz in the parabolic curve as temperature changes from 0°C to 50°C [24].

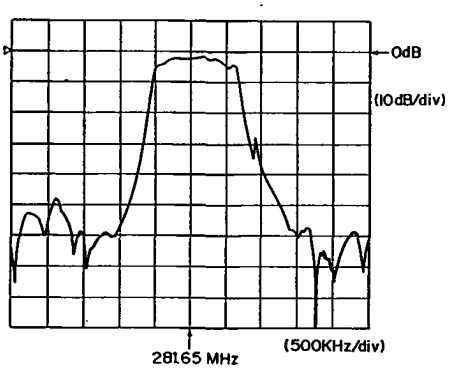


Fig. 6. Frequency response of the fabricated 280MHz SAW resonator filter for a paging receiver. (From S. Matsumura et al. [24])

SAW resonator filter consisting of three IDTs and two reflectors

Y. Ebata et al. have described SAW resonator filter consisting of three IDTs and two reflectors. Cascading the two-port resonator as shown in Fig.5 can easily form a multi-pole resonator filter, called an IDT coupled SAW resonator filter. However, with an increase in the number of IDT electrodes, higher mode responses appear, which are longitudinal adjacent resonance modes for the grating reflector cavity. Among them, the asymmetric mode is the closest to fundamental mode and causes spurious responses in the filter. In order to suppress the closest higher mode, a two-port resonator is constructed with completely symmetrical configuration. It consists of a center IDT and two outer IDTs connected in parallel. All transducers couple to the fundamental mode, but cancel out the charge induced by the asymmetric mode. A front end filter for a paging receiver is constructed with three 3-IDT resonator filters connected in series on LBO substrates, which forms the QARP configuration [23].

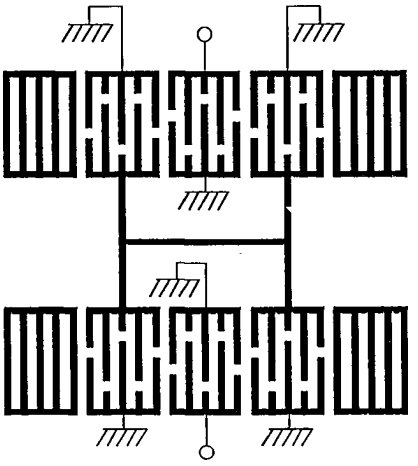


Fig. 7. Configuration of the SAW resonator filter in 900MHz on $Li_2B_4O_7$ substrate.

M. Ohmura et al. have demonstrated the SAW resonator filter on 45° X-Z LBO substrates for a front end filter. Fig.7 shows the configuration of the SAW resonator filter consisting of three IDTs and two reflectors, which has two SAW propagating tracks. It is important to reduce the insertion loss of SAW resonator filter. The electrode resistance is thought to be one reason of high insertion loss. Fig.8 shows the minimum insertion loss of SAW resonator filters which have Al thicknesses of 510, 740 and 950Å. The insertion loss is minimum at Al thickness of 740 Å when the number of IDT pairs is from 30.5 to 34.5 pairs and the terminal impedance is 50Ω. It may be interpreted in terms of the electrode resistance, the mode conversion loss and so on. Also, it becomes evident by simulation that insertion loss, ripple and 1dB bandwidth decrease with the number of center IDT pairs. Because 1dB bandwidth is required to be more than 3MHz in the specification of a front end filter, the number of center IDT pairs is determined.

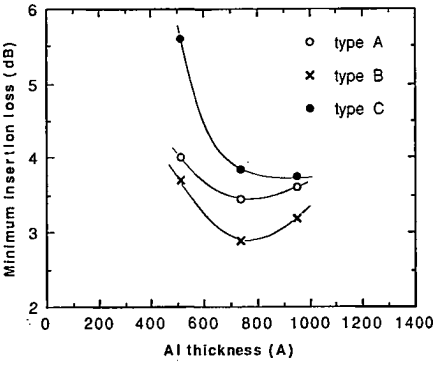


Fig. 8. Dependence of minimum insertion loss on the Al thickness in a 900MHz SAW resonator filter. The numbers of IDT pairs in type A, B, C are 30 or 34 pairs. The aperture is 550μm or 825μm. The number of reflector strips is 65.

Fig.9 shows the frequency response of the SAW resonator filter in the terminal impedance of 50Ω. The electrical performance of SAW filter has a center frequency of 914MHz, insertion loss of 3dB, bandwidth at 3dB down of 4.5MHz and rejection level of more than 60dB. The center

frequency shifts only 350ppm and the passband ripple is less than 1.5dB in the temperature range from -10°C to 55°C [26].

SAW filter consisting of three IDTs with multielectrodes

A GSM receiver requires an IF filter having a fairly broad bandwidth, high rejection, small group delay ripple, low insertion loss and small size. To obtain low insertion loss, a SAW filter consisting of three IDTs with multielectrodes in a SAW propagating track are used. Fig.10 shows the configuration of SAW filter consisting of three IDTs with multielectrodes in a SAW propagating track.

To satisfy an fractional bandwidth of more than 0.46%, 45° X-Z LBO substrates are chosen in consideration of k^2 and TCD in comparison with other substrates, such as $LiTaO_3$, quartz and $LiNbO_3$.

To suppress the spurious level caused by transversal mode, the spurious level is examined using SAW filters which have three aperture and five Al thicknesses. The spurious level increases as the decrease in temperature and the increase in Al thickness in addition to the increase in aperture. The values of Al thickness and aperture are chosen, for example, at an Al thickness normalized by wavelength of 1.67% and an aperture normalized by wavelength of 8.9.

To suppress the spurious level caused by the side lobe peak of IDT radiation conductance, the ratio of center IDT pairs to side IDT pairs is chosen at 1.35.

To attain the high rejection level, two SAW propagating tracks is used. The side IDTs connected in parallel in the upper track are connected in cascade with the center IDT in the lower track as an IDT coupling. It is temporarily called as the reversed connection. Fig.10 shows the configuration of the SAW filter in the reversed connection, which consists of three IDTs with multielectrodes in a SAW propagating track.

To obtain a small delay ripple and a fairly broad bandwidth, the dependence of the passband characteristics, such as delay ripple, fractional bandwidth and amplitude ripple, on the total number of IDT pairs is investigated theoretically and experimentally in the configuration of the reversed connection. Fig.11 shows the frequency response of the SAW filter obtained.

The electrical performance of the SAW IF filter on 45° X-Z LBO substrates has a minimum insertion loss of 5dB, fractional bandwidth of 0.46%, group delay ripple of 2μs, and relative rejection level of more than 50dB. The dimension of the SMD package is 13x6.5x2.1 mm^3. The center frequency shifts 800ppm in the temperature range from -20 °C to 75°C.

SAW resonator filter consisting of three IDTs, two internal reflectors and two reflectors

A SAW resonator filter consisting of three IDTs and two reflectors on 45° X-Z LBO substrates has the performance of low insertion loss, high rejection and moderate bandwidth as described in the previous section. However, the ground of the center IDT in the propagating track must be connected to the ground terminal by a fine wire. The wire must step over the propagating track because the ground pad of the center IDT is on the opposite side to the ground terminal.

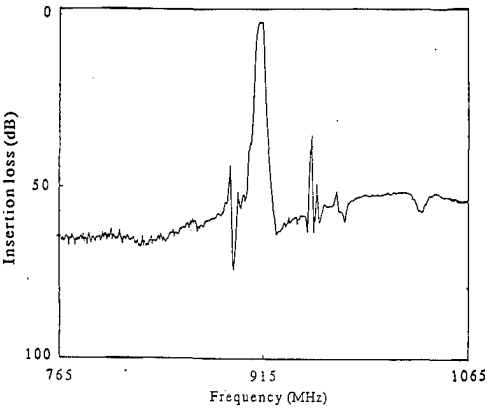


Fig. 9. Frequency response of a 900 MHz SAW resonator filter.

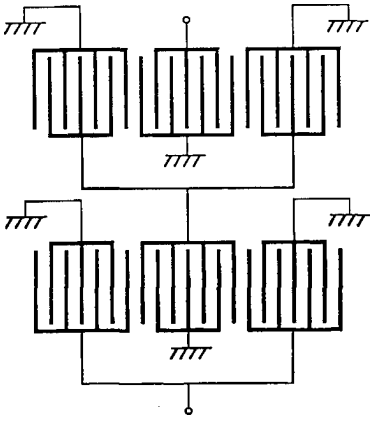


Fig. 10. Configuration of SAW filter consisting of three IDTs with multielectrodes in a propagating tracks. The electrical coupling between tracks is temporarily called as the reversed connection.

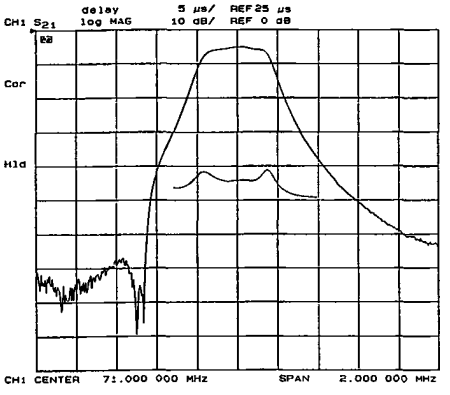


Fig. 11. Frequency response of the reversely connected SAW filter. The SAW filter has three IDTs with multielectrodes in a propagating track.

Therefore, two internal reflectors are introduced as the lead pattern for the ground pad of the center IDT to shorten the fine wire for the ground of the center IDT. Fig.12 shows the configuration of SAW resonator filter with two internal reflectors. It is anticipated that the charge distribution of SAW resonator filter with two internal reflectors is different from that of SAW resonator filter without the internal reflector. Therefore, the SAW filter is designed again to obtain the required bandpass performance. It appears that the optimum number of IDT pairs decreases with the normalized Al thickness and the number of strips of the internal reflector.

On the other hand, the insertion loss and rejection level changes by the electrode resistance of internal reflectors. Also, the bandwidth decreases with the number of strips of the internal reflector. The number of strips of the internal reflector and Al thickness are chosen in consideration of the required insertion loss and bandwidth for a front end filter of the cordless telephone. Fig.13 shows the frequency response of SAW filter obtained.

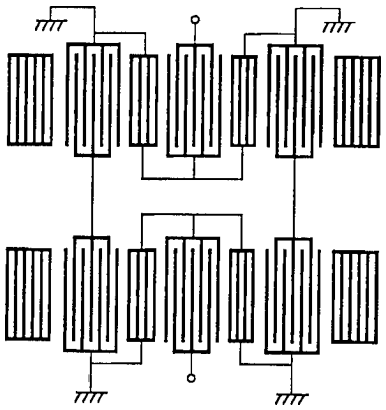


Fig. 12. Configuration of the SAW resonator filter consisting of three IDTs, two internal reflectors and two reflectors.

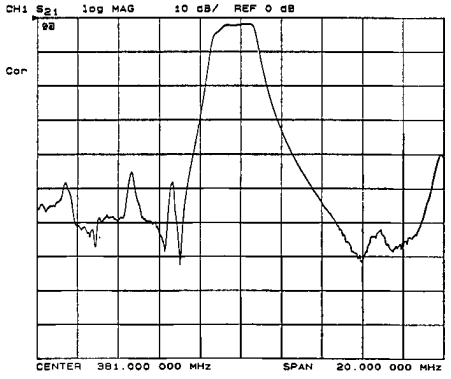


Fig. 13. Frequency response of the SAW filter consisting of three IDTs, two internal reflectors and two reflectors.

Acknowledgments

The authors would like to express hearty thanks to the late Dr. T. Ogawa for his encouragement. We also thank Mr. H. Honmo for helpful support with the preparation of the manuscript and presentation.

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] N.M. Shorrocks, R.W. Whatmore, F.W. Ainger and I.M. Young, "Lithium Tetraborate: A New Temperature Compensated Piezoelectric Substrate Material for Surface Acoustic Wave Device", Proc. IEEE Ultrasonics Symp., 1981, pp.337-340.

[3] B. Lewis, N.M. Shorrocks and R.W. Whatmore, "An Assessment of Lithium Tetraborate for SAW Applications", Proc. IEEE Ultrasonics Symp., 1982, pp.389-393.

[4] D.S. Robertson and I.M. Young, "The Growth and Growth Mechanism of Lithium Tetraborate", J. Material Sci., vol.17, 1983, pp.1729-1738.

[5] K. Fukuta, J. Ushizawa, H. Suzuki, Y. Ebata and S. Matsumura, "Growth and Properties of $Li_2B_4O_7$ Single Crystal for SAW Device Application", Jpn. J. Appl. Phys., vol.22, Suppl.22-2, 1983, pp.140-142.

[6] M. Adachi, T. Shiosaki and A. Kawabata, "Crystal Growth of Lithium Tetraborate ($Li_2B_4O_7$)", Jpn. J. Appl. Phys., vol.24, Suppl.24-3, 1985, pp.72-75.

[7] S. J. Fan, "Properties, Production and Application of New Piezoelectric Crystal Lithium Tetraborate $Li_2B_4O_7$", IEEE International Frequency Control Symp., 1993, pp.353-358.

[8] Y. Ebata, H. Suzuki, S. Matsumura and K. Fukuda, "SAW Propagation Characteristics on $Li_2B_4O_7$", Jpn. J. Appl. Phys., vol.22, Suppl.22-2, 1983, pp.160-162.

[9] C.D.J. Emin and J.F. Werner, "The Bulk Acoustic Wave Properties of Lithium Tetraborate", Proc. 37th Annual Frequency Control Symp., 1983, pp.136-143.

[10] Y. Fujiwara, M. Ono, M. Sakai and N. Wakatsuki, "Strip Type Resonator of Lithium Tetraborate", Proc. 39th Frequency Control Symp., 1985, pp.351-355.

[11] A. Ballato, J. Kosinski and T. Lukaszek and S. Mallikarjun, "Measured Properties of Doubly-rotated Lithium Tetraborate Resonators and Transducers", Proc. IEEE Ultrasonics Symp., 1983, pp.459-462.

[12] R.C. Peach, C.D.J. Emin, J.F. Werner and S.P. Doherty, " High Coupling Piezoelectric Resonators Using Lithium Tetraborate", Proc. IEEE Ultrasonics Symp., 1983, pp.521-526.

[13] M. Murota and Y. Shimizu, "Characteristics of Leaky Surface Waves Propagating on $Li_2B_4O_7$", Proc. IEEE Ultrasonics Symp., 1989, pp.291-294.

[14] M. Murota and Y. Shimizu, "Characteristics of Leaky Surface Waves Propagating on $Li_2B_4O_7$ Substrate with a Thin Aluminum Plate", Jpn. J. Appl. Phys., vol.29,

Suppl.29-1,1990, pp.119-121.
[15] M. Koshiba, S. Mitobe and M. Suzuki, "Equivalent Networks for Surface Acoustic Wave Metallic Gratings", Jpn. J. Appl. Phys., vol.25, Suppl.25-1, 1985, pp.133-135.
[16] M. Takeuchi and K. Yamanouchi, "SAW Reflectors and Resonators with a Positive and Negative Reflectivity", IEICE Technical Report, CPM85-19, 1985, pp.39-46 (in Japanese).
[17] W. Soluch, "Properties of Bleustein-Gulyaev Wave in Li2B4O7 Crystal", Proc. IEEE Ultrasonics Symp., 1993, pp.237-241.
[18] T. Sato and H. Abe, "BGS Wave on Li2B4O7", Proc. IEICE fall Conference, 1993, pp.1-251 (in Japanese).
[19] W.S. Ishak and C.A. Flory, "Acoustic Modes in Lithium Tetraborate", Proc. IEEE Ultrasonics Symp., 1987, pp.241-245.
[20] M. Adachi, T. Shiosaki, H. Kobayashi, O. Ohnishi and A. Kawabata, "Temperature Compensated Piezoelectric Lithium Tetraborate Crystal for High Frequency Surface Acoustic Wave and Bulk Wave Device Applications", Proc. IEEE Ultrasonics Symp., 1985, pp.228-232.
[21] T. Shiosaki and A. Kawabata, "Future Trends in Piezoelectric Materials and Applications", Ferroelectrics, vol.95, 1989, pp.9-14.
[22] T. Shiosaki, M. Adachi, H. Kobayashi, K. Araki and A. Kawabata, "Elastic, Piezoelectric, Acousto-optic and Electro-optic Properties of $Li_2B_4O_7$", Jpn. J. Appl. Phys., vol.24, Suppl.24-1, 1985, pp.25-27.
[23] 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.
[24] S. Matsumura, T. Omi, N. Yamaji and Y. Ebata, "A 45° X cut $Li_2B_4O_7$ Single Crystal substrate for SAW Resonators", Proc. IEEE Ultrasonics Symp., 1987, pp.247-250.
[25] H. Abe, H. Saitou, M. Ohmura, T. Yamada and K. Miwa, "Lithium Tetraborate ($Li_2B_4O_7$) SAW Resonators", Proc. IEEE Ultrasonics Symp., 1987, pp.91-94.
[26] M. Ohmura, H. Abe, K. Miwa and H. Saitou, "A 900MHz SAW Resonator Filter on $Li_2B_4O_7$ Crystal", Proc. IEEE Ultrasonics Symp., 1990, pp.135-138.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

LEAKY SAW PROPAGATION PROPERTIES ON $Li_2B_4O_7$ SUBSTRATES

MASATOSHI ADACHI, TADASHI SHIOSAKI * KAZUHIRO OHTSUKA*[1] and AKIRA KAWABATA

Department of Electronics & Informatics, Faculty of Engineering,
Toyama Prefectural University, Kosugimachi, Toyama 939-03, Japan
*Department of Electronics, Faculty of Engineering, Kyoto University, Kyoto 606, Japan

ABSTRACT

Theoretical and experimental results have shown the potential usefulness of $Li_2B_4O_7$ crystals for leaky SAW applications. The existences of orientations with zero TCD have been predicted by the computer analysis and confirmed experimentally for the leaky SAW. Both theoretical and experimental results have shown that the leaky SAW with the zero temperature coefficient of delay (ZTCD) possesses particularly useful cuts such as (λ, 90°, 70°), (0°, 75°, 42°) and (0, 75°, 75°). The maximum value of k^2 exists at θ of around 60° to 70° on the (λ, 90°, θ). k^2 of the Rayleigh wave is zero at (λ, 90°, 70°), which gives no spurious response for the leaky SAW. Temperature-compensated leaky SAW orientations have been located at both the (0, 75°, 75°) and (45°, 90°, 70°) with v_p of 4225 and 3680 m/s and k^2 of 1.8 and 2.3%, respectively. Further it has been confirmed theoretically and experimentally that a longitudinal leaky SAW with the fast velocity of 6825m/s also exists around (0°, 45°, 90°).

1. Introduction

Lithium tetraborate (Li_2B_4O7) has much attention as one of promising candidates for the SAW materials because of its zero temperature coefficient of delay (TCD) and fairly large coupling, among various SAW materials found to date[1-8]. $Li_2B_4O_7$ is a non-ferroelectric polar piezoelectric material and it melts congruently at 917°C. This crystal has been successfully grown by the Czochralski and Bridgman methods. Many intensive efforts have been made to grow good quality and sizable crystals and also to use this material for high frequency SAW devices[9-11].

Surface acoustic waves (SAW) are widely used for various devices such as filters, resonators, convolvers and optical deflectors. Among the several orientations for zero TCD, the (45°, 90°, 90°) orientation of $Li_2B_4O_7$ in Eulerian expression, that is, the (110) cut [001] propagation is very useful for Rayleigh SAW applications. On the other hand, the leaky SAW velocity is always faster than those of the slow shear wave and Rayleigh wave. Therefore, the leaky SAW is more suitable for applications in high frequency SAW devices than the Rayleigh wave, if its coupling is high, both temperature sensitivity and propagation loss are low.

In this paper, measured and calculated results on the leaky SAW properties related to the fast velocity, high coupling and low TCD are reported.

2. Experimental Procedure

Leaky SAW delay lines have been fabricated on substrates with a number of orientations. Two kinds of IDT's were used in this study. One was a normal IDT with 32 finger pairs and 60μm spatial period. The center to center propagation path length was 2.88mm. This IDT was used to measure the leaky SAW properties for (0°, 75°, θ). The other was a split IDT of the video intermediate frequency (VIF) filter with 58 μm spatial periods, in which the aluminum film was deposited on the propagation surface. This split IDT for VIF was used to measure the leaky SAW properties for (45°, 90°, θ) and (0°, μ, 90°) substrates. To avoid the use of acids for aluminum etching, the lift-off process was used with a negative mask. The phase velocity v_p was calculated from the center frequency in the four port transmission experiment using a network analyzer.

3. Leaky SAW Properties of $Li_2B_4O_7$

In order to calculate the leaky SAW, phase velocity v_p, coupling coefficient k^2, TCD and propagation loss P_L of the leaky SAW and the Rayleigh wave propagating on $Li_2B_4O_7$ substrates were calculated for all cuts and propagation directions from a similar theoretical method to that reported by Nakamura et al. [12], using material constants determined by the present authors[4]. Substrate cut angles and propagation directions are defined by three angles (λ, μ, θ,) on the basis of Euler's expression as shown in Fig. 1. In this expression, the (0°, 0°, 0°) represents the z-cut and x-propagation.

The following features on the leaky SAW properties were obtained from the calculation.

i) v_p varies from 3600 to 5200 m/s,

ii) k^2 changes from 0 to about 7%. The maximum value of k^2 exists at θ of around 60° to 70° on the (λ, 90°, θ) substrate,

iii) TCD varies from -20 to +120ppm/°C. The zero TCD for leaky SAW's exists at appropriate cuts such as (λ, 90°, 70°), (0°, 75°, 42°) and (0°, 75°, 75°), and

iv) k^2 of the Rayleigh wave is zero at (λ, 90°, 70°), which

*[1]Present address: R & D Center for Electronic Devices, Kyocera Corporation Shiga Plant,10-1 Gamo-cho, Shiga 529-15 Japan.

0-7803-1945-1/94 $3.00 © 1994 IEEE

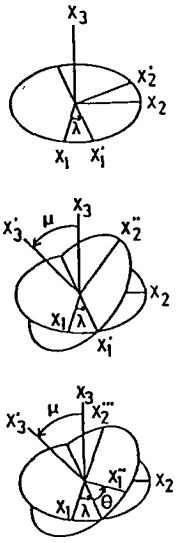


Fig. 1 Coordinate system used to define SAW propagation.

gives no spurious response for the leaky SAW.
v) v_p of the longitudinal leaky SAW on the (0°, μ, 90°) varies from 5310 to 7440 m/s.

These results suggest the potential usefulness of both (45°, 90°, θ) and (0°, 75°, θ) cuts for the leaky SAW applications. From a detailed computer calculation, it was predicted that temperature compensated leaky SAW orientations were located around(0°, 75°, 75°) with v_p of 4225 m/s and k^2 of 1.8%, and also(45°, 90°, 70°) with v_p of 3680 m/s and k^2 of 2.3%. Furthermore, k^2 of the Rayleigh wave was zero at this (45°, 90°, 70°)

Leaky SAW delay lines were fabricated on (0°, 75°, θ) substrates for θ = 68° and 75°, (45°, 90°, θ) substrates for θ = 0° to 90°, and (0°, μ, 90°) substrates for μ = 0° to 90°.

3.1 Leaky SAW properties on (0°, 75°, θ) substrates.

Figure 2 shows both experimental and theoretical results on the v_p, k^2 and TCD with substrate surface orientations and propagation directions at (0°, 75°, θ). Open circles are measured values and lines are calculated ones for the leaky and Rayleigh SAW's. In these substrates, both the Rayleigh wave and leaky SAW were excited at θ of 75°. The SSBW is observed at θ of 0°. However this wave is not suitable one, because of its large TCD and small coupling. From a calculation, it was predicted that temperature compensated leaky SAW orientation were located around (0°, 75°, 75°) with v_p of 4225 m/s and k^2 of 1.8%. Figure 3 shows the distribution of the mechanical displacements U_1, U_2, U_3 and potential φ in the x_3 direction for the (0°, 75°, 75°) substrate. It is shown that they are well confined in the surface layer within a depth of three wavelengths. The experimental values of the v_p, TCD, k^2 and propagation loss P_L with a substrate surface orientation and propagation direction at (0°, 75°, 75°) were measured as 4224 m/s, 10 ppm/°C, 1.7% and 0.36 dB/λ, respectively. There are good agreement between the observed and calculated results.

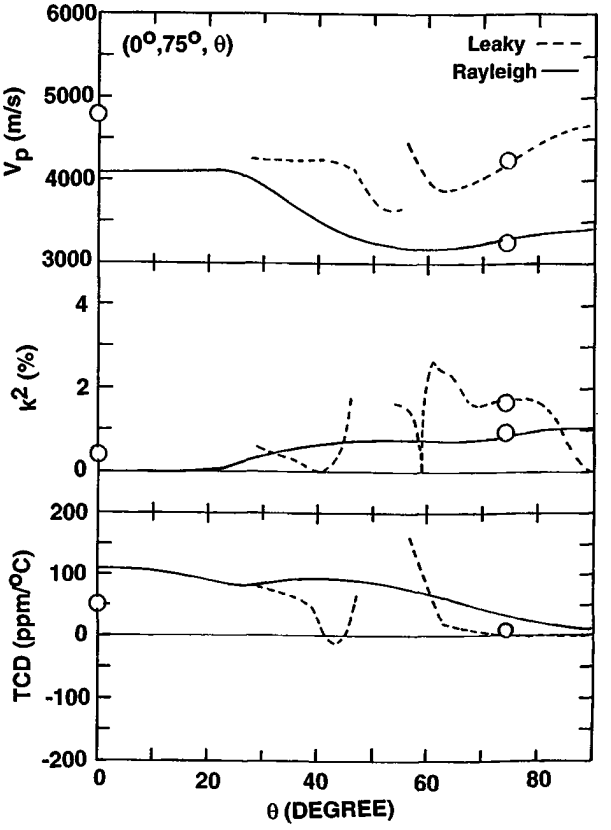


Fig. 2 v_p, k^2 and TCD as a function of angle θ.

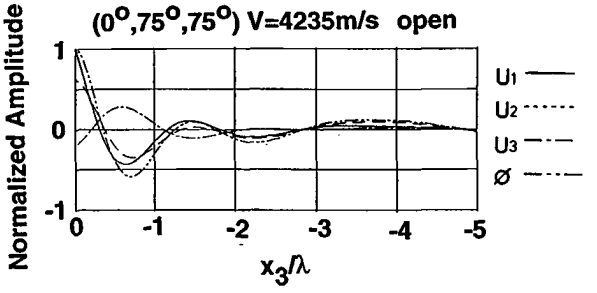


Fig. 3 Mechanical displacements and potential versus normalized distance into crystal for the shear leaky SAW at (0°, 75°, 75°).

3.2 Leaky SAW properties on (45°, 90°, θ) substrates

The SAW propagation surface is shorted electrically, because the IDT for the VIF devices was used in this experiment. Both experimental and theoretical results on the v_p, k^2 and TCD with substrate surface orientations and propagation directions at (45°, 90°, θ) are shown in Fig. 4.

Open circles are measured values and lines are calculated ones for the leaky, Rayleigh and shear waves, respectively. There are good agreement between the observed and calculated results. In the (45°, 90°, θ) substrates, both the Rayleigh wave and leaky SAW were excited at θ of 20° to 90°, except 70°. Figure 5 shows frequency responses for the Rayleigh and leaky waves. Only the leaky SAW appeared at θ of 70° as shown in Figs. 4 and 6. This cut gives a useful substrate orientation for the leaky SAW with small TCD, large coupling and no spurious response of the Rayleigh wave. The measured value of v_p was 3670 m/s, which is in good agreement with the calculated value, 3680 m/s. Figure 7 shows the distribution of the mechanical displacements U_1, U_2, U_3 and potential ϕ in the x_3 direction for the (45°, 90°, 70°) substrate. It is shown that they are well confined in the surface layer with a depth of four wavelengths when the propagation surface is electrically shorted. In the displacement profile, only the mechanical displacement U_3 component is dominant, indicating that this wave possesses a possibility of the shear leaky SAW. From a detailed computer calculation, the leaky SAW is divided into three wave branches at θ of 68°, the leaky SAW and the slow shear wave as partially shown in Fig. 4, although the unknown wave is also calculated, but it is not plotted in the figure. The

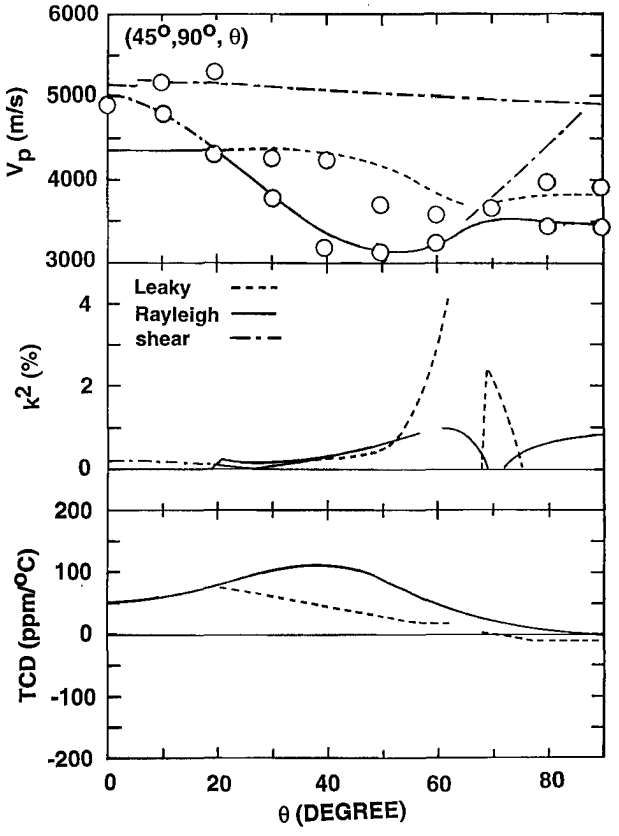


Fig. 4 v_p, k^2 and TCD as a function of angle θ.

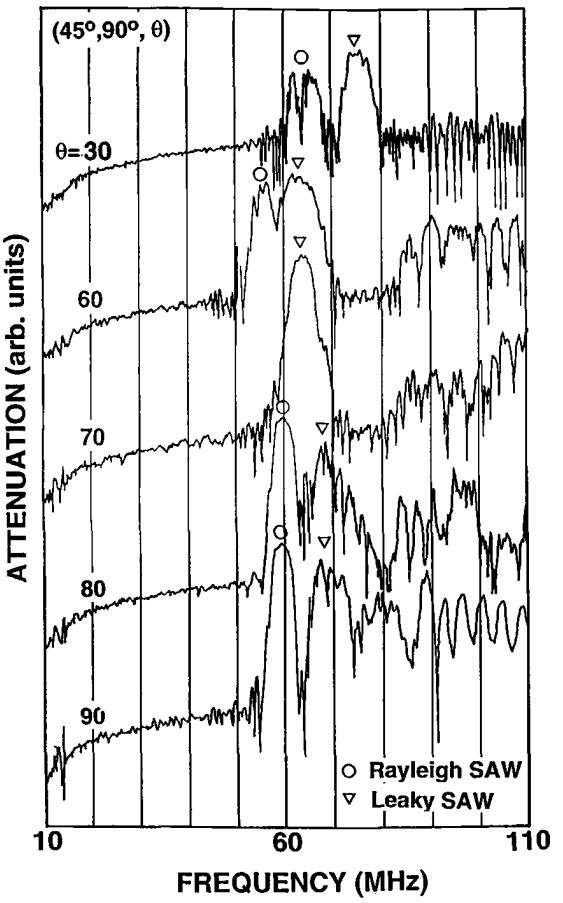


Fig. 5 Measured frequency responses of a delay line as a function of angle θ.

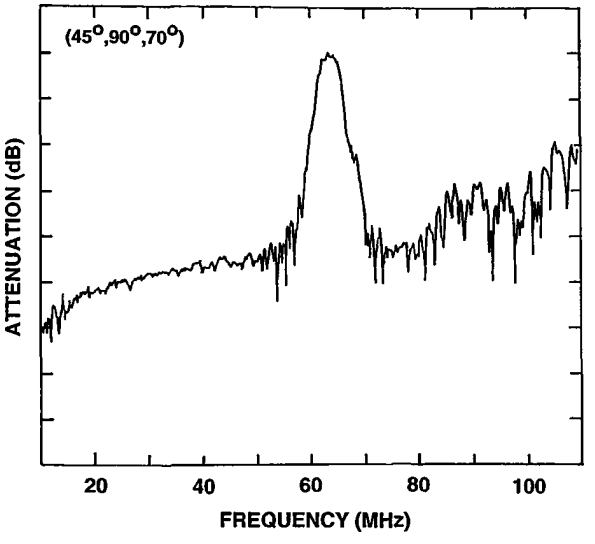


Fig. 6 Measured frequency response.

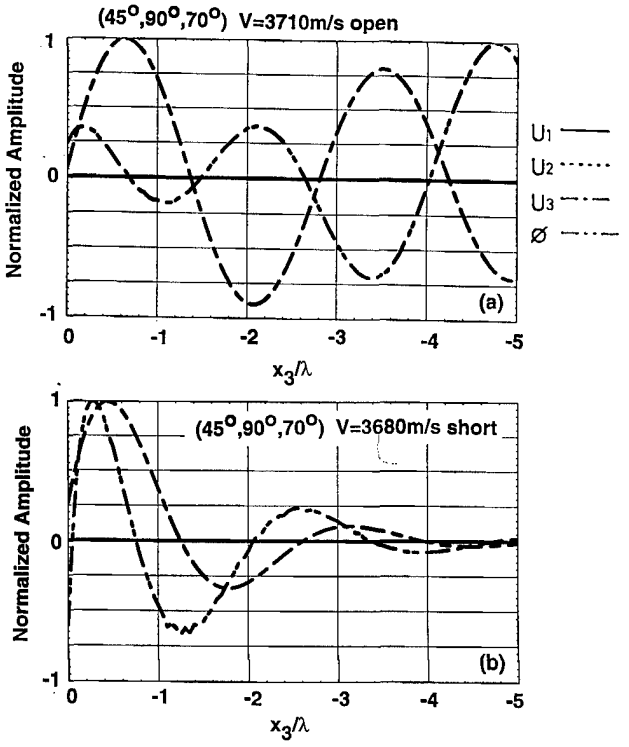


Fig. 7 Mechanical displacements and potential versus normalized distance into crystal for the leaky SAW at (45°, 90°, 70°).

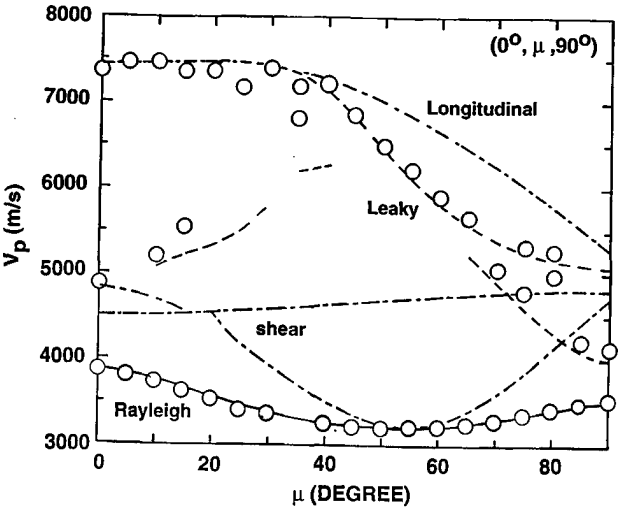


Fig. 8 v_p as a function of angle μ.

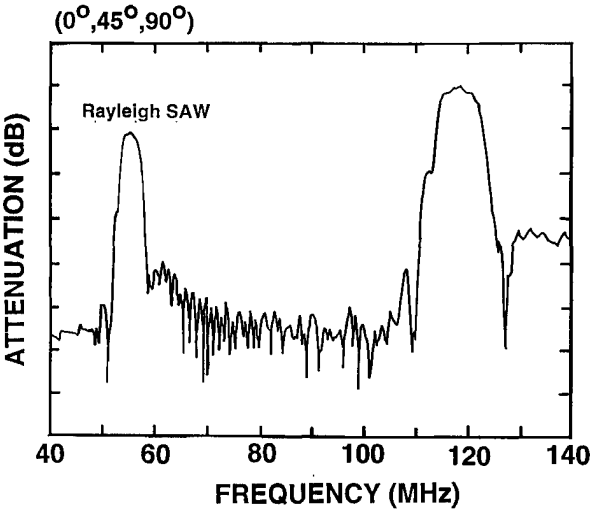


Fig. 9 Measured frequency response.

more detailed calculation and measurements are necessary for better understanding of the leaky SAW properties on the (45°, 90°, 70°) substrate, because this cut gives a useful substrate orientation for the leaky SAW with small TCD, large coupling and no spurious response.

3.3 Longitudinal Leaky SAW properties on (0°, μ, 90°) substrates.

Figure 8 shows the phase velocity v_p as a function of μ. So-called leaky SAW's do not exist at μ of 0° to 90°. The Rayleigh SAW, longitudinal leaky SAW and longitudinal bulk wave have been observed in the wide regions. The Rayleigh SAW possesses v_p of 3250 to 3800 m/s, k^2 of 1 to 1.6% and TCD of 10 to 60 ppm/°C from the calculation. At μ of 80°, the zero TCD is obtained for (0°, μ, 90°) substrate. On the other hand, the longitudinal leaky SAW's with fast velocities were observed at μ of 40° to 80°. The longitudinal leaky SAW velocities vary from 5310 to 7440 m/s. Among these longitudinal leaky SAW's, the longitudinal leaky SAW was excited strongly at μ of around 45°. Figure 9 shows the frequency response for both the longitudinal leaky SAW and the Rayleigh SAW on the (0°, 45°, 90°) substrate. Measured value of v_p for this longitudinal leaky SAW was determined as 6825 m/s. Figure 10 shows the distribution of displacements U_1, U_2, $U_3$and ϕ in the x_3 direction for the (0°, 45°, 90°) substrate. It is shown that they are well confined in the surface layer within four wavelengths when the

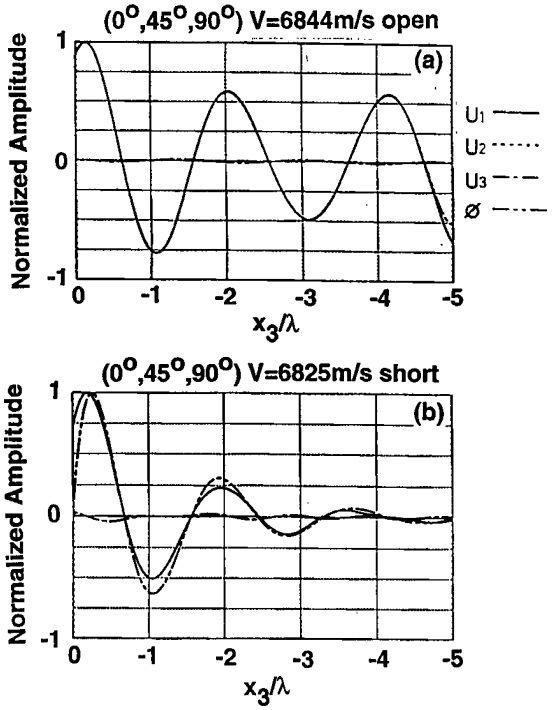


Fig. 10 Mechanical displacements and potential versus normalized distance into crystal for the longitudinal leaky SAW at (0°, 45°, 90°).

propagation surface is shorted electrically, although only the mechanical displacement U_1 component is dominant. Sato et al.[13] reported independently about these longitudinal leaky SAW's with the fast velocity in the 1994 spring meeting of the Institute of Electronics, Information and Communication Engineers. More detailed theoretical studies are necessary for better understanding of these longitudinal leaky SAW properties.

4. Conclusions

Theoretical and experimental results have shown the potential usefulness of $Li_2B_4O_7$ crystals for leaky SAW applications. The existences of orientations with zero TCD have been predicted by the computer analysis and confirmed experimentally for the leaky SAW. Both theoretical and experimental results have shown that the leaky SAW with the zero temperature coefficient of delay (ZTCD) possesses particularly useful cuts such as (45°, 90°, 70°) and (0°, 75°, 75°). The maximum value of k^2 exists at θ of around 60° to 70°. k^2 of the Rayleigh wave is zero at (45°, 90°, 70°), which gives no spurious response for the leaky SAW. Temperature-compensated leaky SAW orientations have been located at (0°, 75°, 75°) and (45°, 90°, 70°) with v_p of 4225 and 3680 m/s and $k^2$of 1.8 and 2.3%, respectively. Further it has been confirmed theoretically and experimentally that the longitudinal leaky SAW with the fast velocity of 6825 m/s also exists around (0°, 45°, 90°).

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 Materials," Electron. Letters, Vol. 17, pp. 11-12 (1981).

[2] N. W. Shorrocks, R. W. Whatmore, F. W. Ainger and I. M. Young,"Lithium Tetraborate - a New Temperature Compensated Piezoelectric Substrate Materials for Surface Acoustic Wave Devices," IEEE 1981 Ultrason. Symp. pp. 337-340 (1981).

[3] B. Lewis, N. W. Shorrocks and R. W. Whatmore,"An Assessment of Lithium Tetraborate for SAW Applications," IEEE 1982 Ultrason. Symp. pp. 389-393 (1982).

[4] M. Adachi, T. Shiosaki, H. Kobayashi, O. Ohnishi and A. Kawabata,"Temperature compensated Piezoelectric Lithium Tetraborate Crystal for High Frequency Surface Acoustic Wave and Bulk Wave Device Applications,' IEEE 1985 Ultrason. Symp. pp. 228-232 (1985).

[5] T. Shiosaki, M. Adachi and A. Kawabata," Growth and Properties of Piezoelectric Lithium Tertraborate Crystal for BAW and SAW Devices," Proc. the sixth International Symp. Applications of Ferroelectrics pp. 455-464 (1986).

[6] H. Abe, H. Saito, H. Ohmura, T. Yamada and K. Miwa,"Lithium Tetraborate ($Li_2B_4O_7$) SAW Resonators," IEEE 1987 Ultrason. Symp. 91-94 (1987).

[7] S. Matsumura, T. Omi, N. Yamaji and Y. Ebata," A 45° x Cut $Li_2B_4O_7$ Single Crystal Substrate for SAW Resonators, IEEE 1987 Ultrason. Symp. pp.247-250 (1987).

[8] Y. Fujiwara, M. Ono, M. Sasaki and N. Wakatsuki," Strip Type Resonator of Lithium Tetraborate," Proc. 39th Annual Frequency Control Symp. pp. 351-355 (1985).

[9] M. Murota and Y. Shimizu," Characteristics of Leaky Surface Wave Propagating on $Li_2B_4O_7$ Substrate," Jpn. J. Appl. Phys. Vol. 28 No.Suppl. 28-1, pp. 120-122 (1989).

[10] M. Adachi, S. Yamamichi, M. Ohira, T. Shiosaki and A. Kawabata," Crystal Growth of $Li_2B_4O_7$ and its Leaky SAW Properties," Jpn. J. Appl. Phys., Vol. 28, No. Suppl. 28-2, pp. 125-127 (1989).

[11] R. Komatsu, S. Uda and K. Hikita,"Mode of Occurrence and Cause of Cracking of $Li_2B_4O_7$ Single Crystals during Growth by the Czochralski Method," Jpn. J. Appl. Phys. Vol. 32 No. 9B, pp. 4364-4366 (1993).

[12] K. Nakamura, M. Kazumi and H. Shimizu, IEEE Ultrason. Symp. pp. 819-822 (1977).

[13] T. Sato and H. Abe,"Longitudinal Leaky SAW Properties on Lithium Tetraborate Substrates," 1994 Spring Meeting of Institute of Electronics, Information and Communication Engineers [in Japanese].

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

LITHIUM TETRABORATE AS A PROMISING MATERIAL FOR BAW FILTERS

Shestopalov K.V., Nefedov V.A., Zadneprovsky B.I.

All-Russian Research Institute for the Synthesis of Materials (VNIISIMS)
601600, Institutskaya, 1, Alexandrov, Vladimir Region, RUSSIA

Abstract: At present Li tetraborate, a new piezoelectric material, is created. This crystal displays high electromechanical coupling and has the cuts with zero temperature-frequency coefficient. These properties have attracted attention of scientists developing the bulk acoustic wave (BAW) filters. For using this crystal it is very important to know and to predict theoretically the effect of temperature on the piezoelement resonant frequency. We have calculated resonant frequencies of rotated Y cut thin plate with metal electrodes and have determined temperature-frequency dependencies for above cut under various rotation angles and various electrode dimensions. The obtained data are in a good agreement with theoretically calculated ones. The temperature- frequency curves of the studied resonators have parabolic shape with the turnover point appearing in the room temperature region. The second order temperature coefficient is found to be $0.30 \cdot 10^{-6}\ ^{\circ}C^{-2}$. Q values range from 2000 to 9000, the dynamic resistance is in range of 25 to 100 Ω. Using acoustically coupled resonators as the base, we have manufactured monolithic filters with the bandwidth of 0.2 to 1.5 %.

Introduction

To elaborate cheap piezoelectric filters having bandwidth of 1-2% as well as resonators with the large resonance spacing it is necessary to look for a new material having larger electromechanical coupling factor (K). AT-cut quartz traditionally used with this purpose has K = 8.9 % that is not enough for such devices. At present Li tetraborate is a new piezomaterial that is used in parallel with the traditional ones (such as Li tantalate, Li niobate, berlinite, langasite). Li tetraborate crystal belongs to tetragonal system, point group 4 mm, it has 6 independent elastic constants (C_{ik}), 3 independent piezoelectric constants (e_{ik}), and 2 independent dielectric constants (ε_{ik}).

Li tetraborate has been analysed elsewhere to determine cuts with zero temperature-frequency coefficient. It has been stated [1] that a +51° rotated Y cut displays a zero temperature first order coefficient of frequency for the thickness shear mode slow wave, a high electro-mechanical coupling (26 %) and a rate of acoustic wave propagation 3240 m/s. The electro-mechanical coupling coefficient of the fast thickness shear mode is zero for above orientation and for the thickness expansion it is - 20.5 % for a value 6880 m/s rate of acoustic wave.

At the same time cuts with the zero temperature-frequency coefficient for both electrically excited modes have been calculated 2 . Rotation angles are +38°49' (for thickness expansion mode) and +48°48' (for thickness shear mode) at the funda-

0-7803-1945-1/94 $3.00 © 1994 IEEE

mental frequency and +71°52' and +68°17' at the third overtone consequently.

In this paper we report the results of temperature-frequency characteristics calculations for Li tetraborate rotated Y-cut and for resonators manufactured of this material and tested with various sized electrodes. Experimental data for these resonators and filters correlating with theoretical predictions are also presented.

Acoustic Wave Propagation in a Rotated Y-cut Plate

Tiersten [3,4] had used more than once Maxwell equations and linear piezo-elastic formulas to describe piezo-electric crystal under the effect of the external electric field in rectangular system of coordinates. Let us use these equations to describe wave propagation in a Li tetraborate sample. Introducing the linear piezo-elastic equations into the movement equations and taking in account electrostatic equations and the crystall symmetry one can obtain a uniform system of movement equations with the following determinant:

$$\begin{vmatrix} \bar{C}_{66} - C & 0 & 0 \\ 0 & \bar{C}_{22}-C & \bar{C}_{24} \\ 0 & \bar{C}_{24} & \bar{C}_{44}- C \end{vmatrix} = 0 \qquad (2.1)$$

where: $\bar{C}_{66}=C_{66}$,
$\bar{C}_{22}=C_{22}+(e_{22}\, e_{22})/\varepsilon_{22}$,
$\bar{C}_{24}=C_{24}+(e_{22}\, e_{24})/\varepsilon_{22}$, (2.2)
$\bar{C}_{44}=C_{44}+(e_{24}\, e_{24})/\varepsilon_{22}$,

and C_{ik}, e_{ik} и ε_{ik} are transformed under the condition of turning the coordinate system around X_1-axis.

Acoustic waves propagating across thickness of an infinite thin plate of a rotated Y-cut have vastly smaller wave numbers in X_1 and X_3 directions (along the plate) than in the thickness direction. Due to neglectability of these wave numbers we have not considered electrical variables along X_1 и X_3 axes.

The (2.1) system can be solved in the following way:

$$C^{(1)}=C_{66} \ ;$$

$$C^{(2)}=[\bar{C}_{22}+\bar{C}_{44}+\sqrt{(\bar{C}_{22}-\bar{C}_{44})^2+4\bar{C}_{24}^2}]/2; \qquad (2.3)$$

$$C^{(3)}=[\bar{C}_{22}+\bar{C}_{44}-\sqrt{(\bar{C}_{22}-\bar{C}_{44})^2+4\bar{C}_{24}^2}]/2;$$

where $C^{(i)}=\rho\omega^2/\eta_i^2$,

ρ - crystal material density,

ω - cyclic frequency,

η_i - thikness wave propagation constant.

The mechanical displacement in an infinite plate can be described as follows:

$$U_i = A_i \, \mathrm{Sin}(\eta_i x_2) \exp(i\omega t) \qquad (2.4)$$

Since the first expression (fast thickness shear mode) does not depend linearly on the other two in (2.3) further we'll take into account only the second expression (expansion shear mode) and the third one (slow thickness shear mode) that is prime practical interest.

Introducing (2.4) into (2.3) for the thickness shear mode one obtain:

$$A_3 / A_2 = \bar{C}_{24} / (C^{(3)} - \bar{C}_{44}) \qquad (2.5)$$

Propagation of Acoustic Waves in a Thin Finite Plate of Li Tetraborate

Taking into account above assumption one can transform a system of movement equations describing mechanical stresses and electrical induction in a turned Y-cut plate of Li tetraborate in following way:

$$T_{6,1}+T_{2,2}+T_{4,3}=\rho\ddot{U}_2 \ ,$$
$$T_{5,1}+T_{4,2}+T_{3,3}=\rho\ddot{U}_3 \ ; \qquad (3.1)$$

$$
\begin{aligned}
T_1 &= C_{12}U_{2,2}+C_{13}U_{3,3}+C_{14}(U_{2,3}+U_{3,2})-\\
&-e_{21}E_2-e_{31}E_3 \quad ,\\
T_2 &= C_{22}U_{2,2}+C_{23}U_{3,3}+C_{24}(U_{2,3}+U_{3,2})-\\
&-e_{22}E_2-e_{32}E_3 \quad ,\\
T_3 &= C_{23}U_{2,2}+C_{33}U_{3,3}+C_{34}(U_{2,3}+U_{3,2})-\\
&-e_{23}E_2-e_{33}E_3 \quad ,\\
T_4 &= C_{24}U_{2,2}+C_{34}U_{3,3}+C_{44}(U_{2,3}+U_{3,2})-\\
&-e_{24}E_2-e_{34}E_3 \quad ,\\
T_5 &= C_{55}U_{3,1}+C_{56}U_{2,1}-e_{15}E_1 \quad ,\\
T_6 &= C_{56}U_{3,1}+C_{66}U_{2,1}-e_{16}E_1 \quad ;
\end{aligned}
\tag{3.2}
$$

$$
\begin{aligned}
D_1 &= e_{15}U_{3,1}+e_{16}U_{2,1}+\varepsilon_{11}E_1 \quad ,\\
D_2 &= e_{22}U_{2,2}+e_{23}U_{3,3}+e_{24}(U_{2,3}+U_{3,2})+\\
&+\varepsilon_{22}E_2+\varepsilon_{23}E_3 \quad ,\\
D_3 &= e_{32}U_{2,2}+e_{33}U_{3,3}+e_{34}(U_{2,3}+U_{3,2})+\\
&+\varepsilon_{23}E_2+\varepsilon_{33}E_3 \quad .
\end{aligned}
\tag{3.3}
$$

As a result of substitution of (3.2) and (3.3) in the second equation of (3.1) and into electrostatic expressions (2.5) and due to above assumptions about E_1 and E_3 one can obtain:

$$U_{3,11}\hat{C}_{55}+U_{3,22}\hat{C}_{44}+U_{3,33}\hat{C}_{33}+U_{3,23}\hat{C}_{34}=\rho\ddot{U}_3,$$

where

$$
\begin{aligned}
\hat{C}_{55} &= C_{55}+\beta C_{56}+e_{24}(e_{15}+\beta e_{16})/\varepsilon_{22},\\
\hat{C}_{44} &= C_{44}+\beta C_{24}\ e_{24}(e_{24}+\beta e_{22})/\varepsilon_{22},\\
\hat{C}_{33} &= C_{33}+\beta C_{34}+e_{24}(e_{33}+\beta e_{34})/\varepsilon_{22},\\
\hat{C}_{34} &= 2C_{34}+\beta C_{44}+\beta C_{23}+e_{24}(e_{23}+e_{34}+\beta e_{24}+\\
&+\beta e_{32})/\varepsilon_{22},\\
\beta &= 1/(C_{24}/(C^{(3)}-\bar{C}_{44}) \quad .
\end{aligned}
\tag{3.4}
$$

If the complex partial derivative $U_{3,23}$ is excluded from the equation (3.4) and if (3.2) is substituted into the first equation of (3.1) one can obtain:

$$\ddot{U}_{3,23}=\ddot{U}_3(\beta\rho/\hat{\hat{C}}_{34})-U_{3,11}(\hat{\hat{C}}_{55}/\hat{\hat{C}}_{34})-$$

$$-U_{3,22}(\hat{\hat{C}}_{44}/\hat{\hat{C}}_{34})-U_{3,33}(\hat{\hat{C}}_{33}/\hat{\hat{C}}_{34}),$$

where

$$
\begin{aligned}
\hat{\hat{C}}_{55} &= C_{56}+\beta C_{66}+e_{22}(e_{15}+\beta e_{16})/\varepsilon_{22},\\
\hat{\hat{C}}_{44} &= C_{24}+\beta C_{22}+e_{22}(e_{24}+\beta e_{22})/\varepsilon_{22},\\
\hat{\hat{C}}_{33} &= C_{34}+\beta C_{44}+e_{22}(e_{33}+\beta_{34})/\varepsilon_{22},\\
\hat{\hat{C}}_{34} &= C_{23}+C_{44}+2\beta C_{24}+e_{22}(e_{23}+e_{34}+\\
&+\beta e_{24}+\beta e_{32})/\varepsilon_{22} \quad .
\end{aligned}
\tag{3.5}
$$

Thus the movement equation transforms to the form:

$$U_{3,11}\tilde{C}_{55}+U_{3,22}\tilde{C}_{44}+U_{3,33}\tilde{C}_{33}=\rho\ddot{U}_3$$

where

$$
\begin{aligned}
\tilde{C}_{55} &= (\hat{C}_{55}-\hat{\hat{C}}_{55}\hat{C}_{34}/\hat{\hat{C}}_{34})/(1-\beta\hat{C}_{34}/\hat{\hat{C}}_{34}),\\
\tilde{C}_{44} &= (\hat{C}_{44}-\hat{\hat{C}}_{44}\hat{C}_{34}/\hat{\hat{C}}_{34})/(1-\beta\hat{C}_{34}/\hat{\hat{C}}_{34}),\\
\tilde{C}_{33} &= (\hat{C}_{33}-\hat{\hat{C}}_{33}\hat{C}_{34}/\hat{\hat{C}}_{34})/(1-\beta\hat{C}_{34}/\hat{\hat{C}}_{34}).
\end{aligned}
\tag{3.6}
$$

The solution procedure for (3.6), considering conditions in the interface regions of metallized and non-metallized resonant structures in thin finite plates, had been earlier described in [3,4]. Here this method is used to produce software to calculate temperature-frequency characteristics.

Calculation of Temperature-Frequency Characteristics

The temperature-frequency characteristics for Li tetraborate resonators have been calculated using the first and the second order temperature coefficients for: elasticity, piezo-moduli, dielectric permeability, density as well as for linear expansion factors taken from [5].

Fig.1 presents theoretically calculated temperature-frequency characteristics for Li tetraborate devices under conditions of various orientations of the piezoelement in question (50°, 54°, and 58°). All piezoelements are 0.089 mm thick. Electrodes are as large as 1.2x0.7 mm. In

the calculations it has been shown that the temperature-frequency curve turnover point is shifted from -10°C to +30°C when the piezoelectric element rotates around X_1-axis from +50° to +58°.

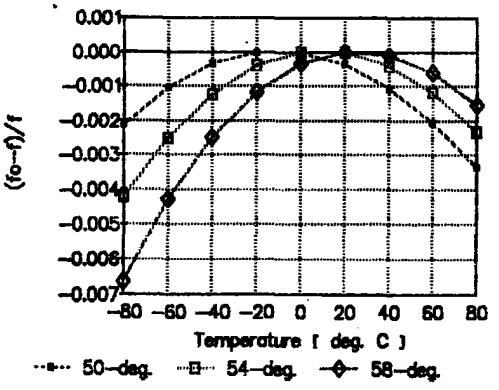


Fig.1 The theoretical piezoelement orientation dependence of temperature-frequency characteristics

Other temperature-frequency data of elements made of 54° turned Y-cuts whith various dimensions of electrodes are presented at Fig.2. The temperature-frequency curve turnover point moves due to the electrode area changes from -10°C (fs-line for the case of infinite non-metallized plate) to +18°C (fmin-line for the case of infinite metallized plate).

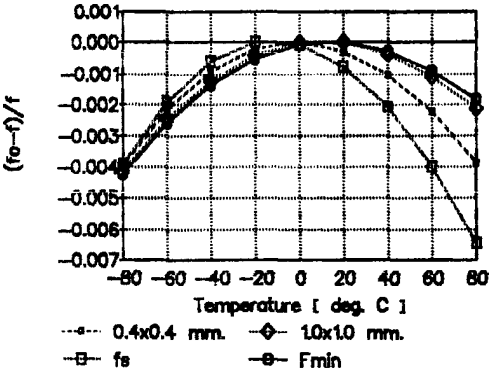


Fig. 2 The theoretical electrode dimensions dependence of resonator temperature-frequency characteristic

Thus, if our goal is to minimize frequency drift in Li tetraborate resonators and filters the cut angle of the piezo-electric element should be corrected depending on the applied electrode area.

Experimental Results

To check the theoretical results obtained we have manufactured Li tetraborate resonators of 51°, 53°, and 54° rotated Y cut with electrodes of various dimensions.

The Li tetraborate crystals were grown by the Czochralski technique in Pt crucibles using the setup for resistanct heating. The raw material is pre-synthesized in accordance to the stoichiometry of Li tetraborate. The seeds were oriented in <001> direction. The growth features: air ambient, the pulling rate 0.3 - 2 mm/h and the rotation rate 15 - 20 rpm. In above conditions we obtain Li tetraborate boules in diameter up to 45 mm and 150 mm long.

The plates have been lapped and polished by chemical etching to obtain the specified frequency value with the accuracy of +/- $2x10^{-5}$. Rectangular electrodes have been coated on to the main surfaces by vacuum deposition technique. Figures 3 and 4 and the table 1 demonstrate temperature-frequency characteristics and some parameters of manufactured resonators.

Comparing theoretical (Fig.1 and 2) and experimental (Fig.3 and 4) results one can make conclusion that the temperature-frequency curve turnover point can be calculated with required precision. However the parabolic curve slope for experimental samples is better than theoretically predicted one. Evidenlty it may be caused by using in the case of calculations of the second order temperature-frequency coefficients thouse are not precise enough to describe piezoelectric crystal. We pro-

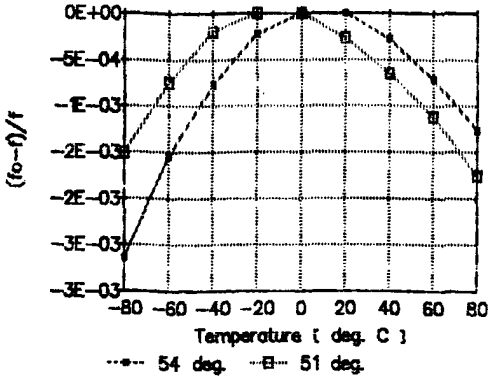


Fig.3 The experimental piezoelement orientation dependence of resonator temperature-frequency characteristic

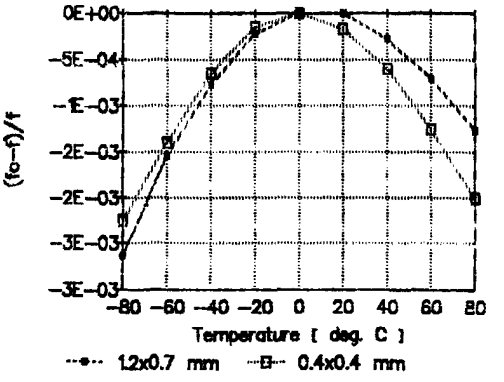


Fig.4 The experimental electrode dimensions dependence of resonator temperature-frequency characteristic

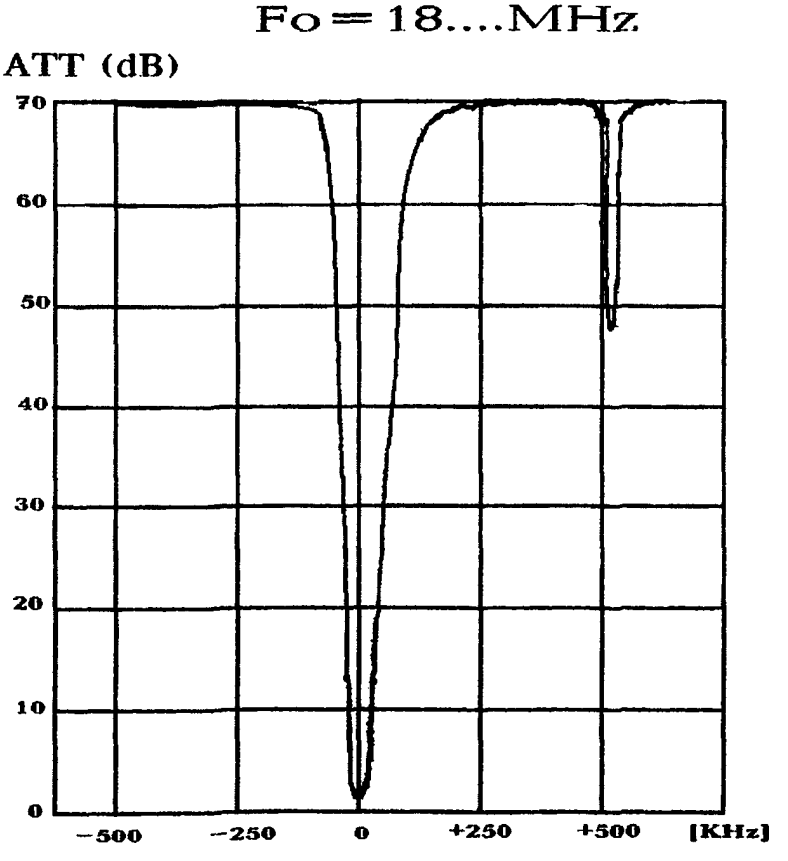


Fig.5. 4-pole filter amplitude-frequency characteristic

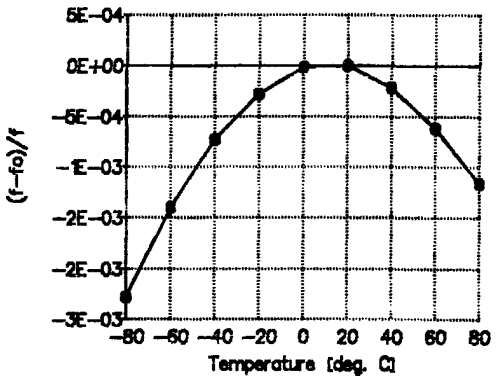


Fig.6. The experimental temperature-frequency characteristic of 4-pole filter with 54° rotated Y-cut piezoelement

Table 1

Frequency	Electrode size	esr	Q	Turnover temper.	Cut
MHz	mm	ohms		deg. C	deg
10.706	1.9x0.8	25	6973	+18	54
18.141	1.2x0.7	20	9100	+18	54
18.215	1.14x0.46	40	6200	0	51
11.315	1.3x0.7	35	2300	-17	51
18.752	1.14x0.46	40	5700	+10	53

Table 2

No. of poles	Passband (dB)	Passband (kHz)	Stopband (dB)	Stopband (kHz)	Stopband (dB)	Stopband (kHz)	Max. ripple (dB)	Max. loss (dB)	Attenuat gurrantd (dB)	Terminating impedance (Om/pF)
2	3	310	20	900	-	-	1.0	0.5	40	1600/1.5
4	3	70	40	210	60	360	0	1.0	70	590/4
4	3	35	40	120	60	250	0.2	2.0	70	750/4
6	3	33	40	90	80	216	0.2	1.5	80	750/4

pose at a later time to carry out more extended study of Li tetraborate crystals manufactured in VNIISIMS.

Figures 5 and 6, and the table 2 show amplitude-frequency and temperature-frequency characteristics of Li tetraborate filters with acoustically coupled resonators. Obtained values make it possible to use these filters in intermediate frequency range for various radiorevers.

Conclusions

Above computations have shown a necessity to take in consideration an electrode area when selecting the angle of cutting piezoelectric plate. Experimental results are in a good agreement with calculated ones. These results are confirmed the statement it is possible to design bandwidth piezofilters for intermediate frequencies of radio-receivers and to produce resonators having large resonant spacing for voltage-operated oscillators. The resonators developed of this material have the temperature-frequency characteristics passed through a maximum at room temperature as well as the second order temperature coefficient of frequency is equalto $0,30x10^{-6}$ $^{o}C^{-2}$. The filters have transmission band of 0,2-1,5%.

Literature

1. J.Fajimara. Proc. 39th Annual. Freguensy.Contr.Symp.May.1985.,pp 351-355.

2. C.D.Y. Emin. Proc. 37th Annual. Freguency. Control. Symp., June 1983, pp. 136 - 143.

3. Tiersten H.F. Linear Piezoelectric Plate Vibrations.- New York: Plenum Press, 1969.- 212p.

4. Tiersten H.F. Analysis of overtone modes in monolithic crystal filters// J. Acoust. Sos. Amer.- 1977.- Vol.62, No. 6.-P.1424-1430.

5. M.Adachi IEEE Ultrasonics Symp. Proc., 1985 pp., 228 - 232.

Sensitivity Analysis of One Port and Two Port BAW and SAW Resonator Model Parameters

Blaine D. Andersen*, Mark Cavin**, Madjid A. Belkerdid, and Donald C. Malocha

University of Central Florida, Electrical and Computer Engineering Dept., Orlando, FL 32816
*Piezo Technology, Inc., P.O. Box 547859, Orlando, FL 32854-7859
**Sawtek, Inc., P.O. Box 609501, Orlando, FL 32860-9501

Abstract--**A great deal has been written on the subject of equivalent circuit modeling of bulk acoustic wave (BAW) and surface acoustic wave (SAW) resonators and on the extraction of their equivalent circuit parameters from measured S-parameter data. However, it is often difficult to obtain reliable and repeatable extraction of the equivalent circuit parameters due to random errors in the measured data itself. Several factors contribute to measurement uncertainty including temperature variation, characterization of fixture parameters, instrument noise, and operator error. This paper first presents a theoretical sensitivity analysis of the extraction of equivalent circuit paramters for single port resonators. For two port resonators, measured S-parameter data was randomly perturbed to simulate the effect of white Gaussian noise in the measurement system. Algorithms were developed using this technique to develope prediction curves giving standard deviation to mean ratios of the extracted equivalent circuit parameters in terms of the noise standard deviation.**

I. INTRODUCTION

The equivalent circuit for a single port quartz resonator is shown in figure 1. The parameters R_1, L_1, and C_1 are termed the motional elements since they model the vibration of the crystal. The capacitance C_0 represents the static capacitance of the crystal [2].

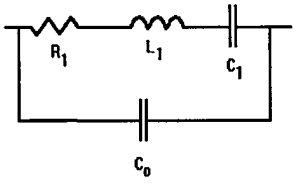


Figure 1. Single port narrowband resonator model. R_1, L_1, and C_1 are the motional resistance, inductance, and capacitance; C_0 is the static capacitance.

A number of different methods have been used to extract the values of the equivalent circuit parameters. Two methods were used for the purposes of this reaearch. The first follows the EIA-512 standard [3] and the second uses a point by point search for the zero phase point.

A first order sensitivity analysis of the extraction of the motional resistance from S-parameter data was reported by Park [1]. In her work, approximations were given for the standard deviation to mean ratio for measured values of R_1 as a function of the average magnitude and standard deviation of S_{11} at resonance. It was found that these approximations were very good for high figure of merit resonators, but were not satisfactory for lower figure of merit resonators. The figure of merit is the ratio of the conductance of R_1 (G_1), to the susceptance of C_0 (B_0) at series resonance. This problem is especially significant for higher frequency resonators since typically, as frequency increases, the susceptance of C_0 increases therefore the figure of merit decreases for a given C_0 and R_1.

This works extends the first order sensitivity analysis of Park to a general sensitivity analysis shown to be valid for all single port resonators. It is also shown that the general equations developed herein reduce to the first order approximations for resonators with high G_1 to B_0 ratios. In addition, a first order sensitivity analysis is presented for two port SAW resonators.

II. SENSITIVITY ANALYSIS FOR EXTRACTION OF EQUIVALENT CIRCUIT PARAMETERS FOR SINGLE PORT RESONATORS.

Referring to the equivalent circuit of figure 1, and defining X_1 as the combined series reactance of L_1 and C_1. The admittance of the resonator can then be written as:

$$Y = jB_o + \frac{R_1 - jX_1}{R_1^2 + X_1^2} \qquad (1)$$

At series resonance, $X_1 = 0$, therefore, at resonance:

$$Y = jB_o + \frac{1}{R_1} \qquad (2)$$

0-7803-1945-1/94 $3.00 © 1994 IEEE

Solving for R_1 in terms of Y at resonance:

$$R_1 = \frac{1}{Y - jB_o} = \frac{1}{yY_0 - jB_o} \quad (3)$$

The normalized admittance (y) is given in terms of S_{11} and the characteristic admittance Y_0 by:

$$y = \frac{Y}{Y_0} = \frac{1 - S_{11}}{1 + S_{11}} \quad (4)$$

Differentiating R_1 with respect to y, and y with respect to S_{11} and applying the chain rule:

$$\frac{dR_1}{dS_{11}} = \frac{dR_1}{dy}\frac{dy}{dS_{11}} = \frac{2Y_0}{(yY_0 - jB_o)^2(1 + S_{11})^2} \quad (5)$$

To obtain the fractional change in R_1 for a given change in S_{11}, multiply through by dS_{11} and divide through by R_1.

$$\frac{dR_1}{R_1} = \frac{2Y_0 dS_{11}}{(yY_0 - jB_o)^2(1 + S_{11})^2 R_1} \quad (6)$$

Substituting equation (3) into the right hand side of equation (6) for R_1 and simplifying gives:

$$\frac{dR_1}{R_1} = \frac{2Y_0 dS_{11}}{(yY_0 - jB_o)(1 + S_{11})^2} \quad (7)$$

Substituting equation (4) into the right hand side of equation (7) for y and simplifying gives:

$$\frac{dR_1}{R_1} = \frac{2Y_o dS_{11}}{(Y_o - jB_o) - j2B_o S_{11} - S_{11}^2(Y_o + jB_o)} \quad (8)$$

Equation (8) gives the fractional change in R_1 from its average value in terms of the average value of S_{11} and the differential change in its value (dS_{11}). If the resonator has a figure of merit much greater than 1 then $G_1 >> B_0$. When this condition is met the preceding derivation is greatly simplified by neglecting B_0 starting with equation (2). Equation (8) then reduces to:

$$\frac{dR_1}{R_1} = \frac{2dS_{11}}{1 - S_{11}^2} \quad (9)$$

Assume B_0 is constant over the narrow measurement frequency range. Further assume that random Gaussian variations in the magnitude of S_{11} at resonance cause random variations in the extracted value of R_1 with dR_1 and dS_{11} approximated as:

$$dR_{1_i} = R_{1_i} - \overline{R_1} \quad (10)$$

$$dS_{11_i} = S_{11_i} - \overline{S_{11}} \quad (11)$$

Substituting the relationships of (10) and (11) into equation (8) and taking the expected value gives:

$$\frac{\sigma_{R_1}}{\overline{R_1}} = \frac{2Y_o\sigma_{S_{11}}}{(Y_o - jB_o) - j2B_o\overline{S_{11}} - \overline{S_{11}}^2(Y_o + jB_o)} \quad (12)$$

Equation (12) gives the fractional standard deviation for extracted values of R_1 in terms of the average measured value and standard deviation of S_{11} at series resonance.

Again note that a large figure of merit justifies neglecting B_0 in the above derivation and allows the simplification of equation (12) to:

$$\frac{\sigma_{R_1}}{\overline{R_1}} = \frac{2\sigma_{S_{11}}}{1 - \overline{S_{11}}^2} \quad (13)$$

Equation (13) is identical to the result reported by Park for a first order approximation. Figure 2 shows plots of this equation with σ_{s11} ranging from 10^{-3} to $5x10^{-5}$. Plotted on the same graph with a darker line is the normalized motional resistance r_1 which is the ratio of the motional resistance R_1 to the characteristic impedance of the measurement system (typically 50 ohms).

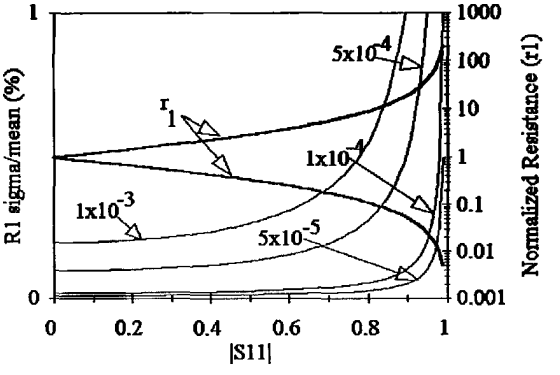


Figure 2. Approximate standard deviation to mean ratio of R_1 and normalized resistance r_1 versus $|S_{11}|$.

From the plots of figure 2 we see that as S_{11} at resonance increases in magnitude, the uncertainty in the extracted value of R_1 increases. In otherwords, resonators which are closely matched to the characteristic impedance of the measurement system have the least degree of uncertainty in the extracted values of R_1.

Figure 3 compares the results predicted by equations (12) and (13) with $\sigma_{s11} = 5.8 \times 10^{-4}$. Observe that as the figure of merit decreases (in otherwords as B_0 increases) the curves bend downward and flatten out indicating less uncertainty in the extracted value of R_1 for low figure of merit units.

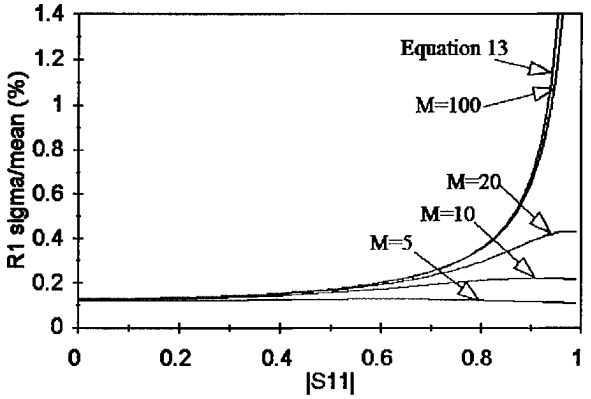


Figure 3. Comparison between first order prediction of equation (13) and general result of equation (12).

Two factors should be considered when using the approximation of equation (13), over the general result of equation (12). First, the resonators figure of merit, and second, the resonators reflection coefficient (S_{11}) at resonance (or equivalently, its motional resistance). The difference between these two equations was found to be less than two tenths of one percent for typical resonators with figures of merit greater than about 25 and low reflection. However, for resonators with low figures of merit and/or high reflection, the difference between the two equations becomes more significant.

For example, from the plots of figure 2, if the average magnitude of S_{11} at resonance is 0.8 and $\sigma_{|S11|}$ is 5.8 x 10^{-4}, then the approximation of equation (13) gives a σ/mean ratio for R_1 of about 0.35%. If the figure of merit for this resonator is 5, then the general result of equation (12) gives this ratio to be about 0.12%. In this case, accurate results can only be obtained from equation (12). However, regardless of the figure of merit and the magnitude of S_{11}, if all that is required is a worst case estimation of the motional resistance standard deviation to mean ratio, then the use of equation (13) and the plots of figure 3 are sufficient.

The accuracy of the preceding analysis was first investigated by computer simulation. Values for the equivalent circuit elements were chosen to give similar Q's, a wide variety of figures of merit, and S_{11} magnitudes at resonance ranging from near zero to near one. Values for S_{11} at series resonance were then calculated for each of the equivalent circuits. The magnitude of each S_{11} value was then randomized 100 times using a Gaussian random number generator with various standard deviations. R_1 was then calculated for each of the S_{11} values, simulating 100 measurements of the same resonator. The σ/mean ratios were then calculated for comparison to the predicted results from equation (12) and (13). Table 1 shows typical results.

The data in table 1 shows excellent correlation between predicted and simulated results. The differences between the predicted values and the simulated values are insignificant for each unit except the last two. These two have low figures of merit (7.869 and 4.541). For these two, the first order approximation is not as accurate in predicting the simulation results, however, the general result of equation (12) still predicts the simulated results with a maximum difference of about three tenths of one percent.

To further investigate the accuracy of the models, several single port BAW resonators were measured. Table 2 shows the comparison between the actual motional resistance standard deviation to mean ratio and values predicted by the first order approximation of equation (13) and the general result of equation (12).

Inspection of the results of the simulated data in table 1 as well as the actual measurement data in table 2 shows that the sensitivity analysis model predicts the actual standard deviation to mean ratio for the motional resistance with excellent accuracy. The measured data and the computer simulated data support the conclusions drawn from the analysis. That is, the first order sensitivity analysis which neglects the effect of C_0 in analyzing the effects of variations in measured S_{11} data on the extracted motional resistance is sufficient for high figure of merit resonators. However, resonators such as Units # 7, 8, and 9 in the computer simulation and the actual 80 MHz and 173 MHz resonators require the more general solution due to their low figure of merit, their high reflection coefficient at resonance, or a combination of the two.

IV. DUAL PORT SAW RESONATOR MODEL.

In order to extend the sensitivity analysis of single port resonators to dual port resonators, the dual port SAW resonator model and extraction technique presented by Cavin was used [4]. The equivalent circuit model is shown in figure 4, where C_t is the transducer capacitance, C_f is the pin to header capacitance, C_{12} is the coupling capacitance between input and output transducers, L_b is the bond wire inductance, R models the bond resistance, and R_1, L_1, and C_1 are the motional elements as before.

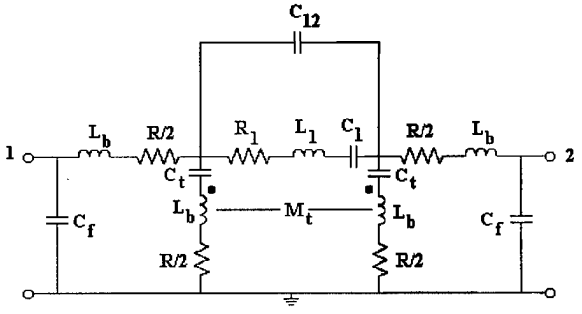


Figure 4. Surface Acoustic Wave Resonator Model.

The benefit of this model is that it accurately represents the resonators wideband response as well as its narrowband response near the main resonance. In addition, longitudinal modes can be modeled by simply adding another series motional RLC combination in parallel with the main resonance. In order to extract equivalent circuit parameters for the longitudinal modes, the narrowband S-parameter file must to wide enough to include them.

The modeled and measured wideband response for a 217 MHz resonator is shown in figures 5. It shows very good agreement between the equivalent circuit model and the actual SAW resonator.

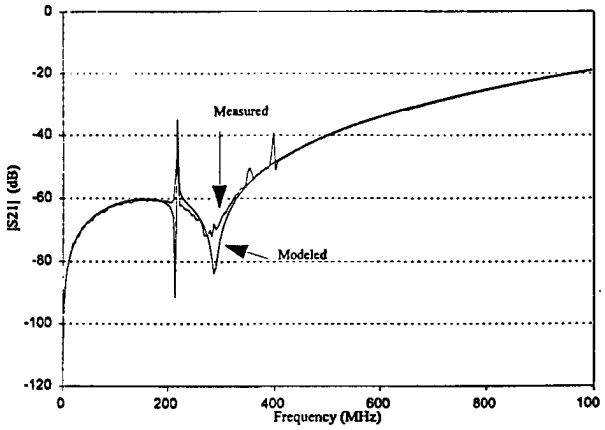


Figure 5 Wideband comparison between measured and modeled magnitude of S_{21} for a 217 MHz SAW resonator.

V. SENSITIVITY ANALYSIS FOR EXTRACTION OF EQUIVALENT CIRCUIT PARAMETERS USING TWO PORT SAW RESONATOR MODEL.

Due to the large number of variables involved in full two port reflection and transmission measurements, it was considered impractical at this point to develop detailed analytical expressions for the standard deviation of a resonator parameter as a function of deviation in measured S-parameters. Rather, a method was devised to simulate large numbers of measurements, analyze the statistical variations in the data, and draw conclusions regarding the uncertainty in measured parameters.

A computer program was written to simulate additive white Gaussian noise in the extracted S-parameter data (S_{11}, S_{12}, S_{21}, and S_{22}). A Gaussian random number generator with zero mean and user selected standard deviation was used to perturb sets of S-parameter data stored on disk. The routines for extracting equivalent circuit parameters were then run using the perturbed data as input. It was then possible to simulate a statistically significant number of measurements without the need to perform numerous time consuming measurements.

Devices were chosen to give a variety of frequencies and resonator parameters. A full set of S-parameter data was measured for each device over a narrow band near its main resonance and over a wide band from 300 KHz to 3 GHz. From this data, the equivalent circuit parameters were extracted for each device.

The S-parameter data used to extract the equivalent circuit parameter was then perturbed and the circuit parameters were extracted again. This process was repeated for each device using several different standard deviations for the simulated noise.

Figures 6, 7, and 8 display the results of this process graphically. Figure 6 plots the standard deviation to mean ratio of the motional resistance versus the standard deviation of the noise. Similarly, figures 7 and 8 plot the standard deviation to mean ratios of the motional inductance and capacitance versus noise standard deviation.

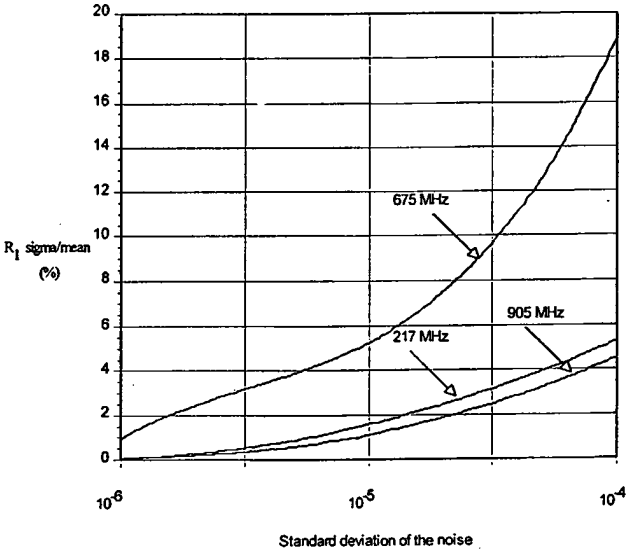


Figure 6. R_1 uncertainties for SAW resonators.

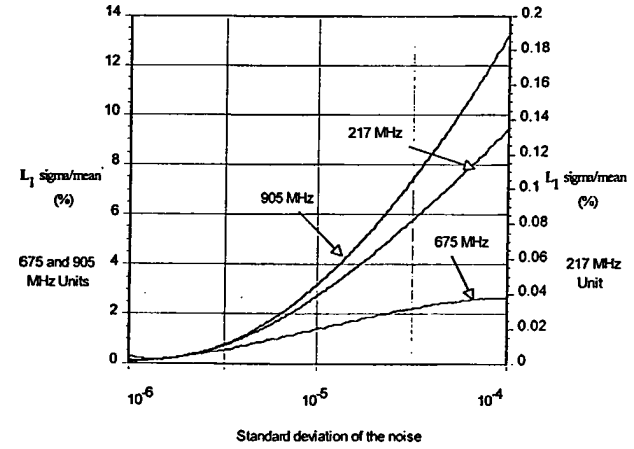


Figure 7. L_1 uncertainties for SAW resonators.

Table 1. Average results from 100 computer simulated extractions of resonator equivalent circuit parameters and comparisons between predicted and simulated results (with $\sigma_{|S11|}$=0.001).

Unit #	R_1 (Ohms)	L_1 (mH)	C_1 (fF)	M	$\lvert S_{11} \rvert$	$\frac{\sigma_{R_1}}{R_1}$ % Computer simulated results.	$\frac{\sigma_{R_1}}{R_1}$ % Predicted by first order model	$\frac{\sigma_{R_1}}{R_1}$ % Predicted by general model
1	50.000	50	2.5	44.7	0.01120	0.202618	0.202567	0.202568
2	75.019	145	3.2	57.4	0.20023	0.209406	0.209428	0.209441
3	99.993	300	3.7	66.6	0.33334	0.221137	0.221142	0.221158
4	125.02	450	3.5	63.5	0.42866	0.219711	0.219702	0.219725
5	150.03	850	4.1	78.7	0.50008	0.249506	0.249351	0.249371
6	200.03	1000	3.1	55.7	0.60004	0.286926	0.286607	0.286662
7	300.22	1500	2.1	37.4	0.71447	0.412502	0.412568	0.412778
8	500.28	2600	1.3	23.3	0.81827	0.551509	0.551340	0.552174
9	1000.8	3600	0.43	7.9	0.90483	1.109194	1.089930	1.105855
10	1999.3	8250	0.25	4.5	0.95118	2.222050	2.122815	2.220319

Table 2. Averaged results from multiple measurements taken on several single port BAW resonators and comparisons between predicted and simulated results.

f_s (MHz)	R_1 (Ohms)	$\lvert S_{11} \rvert$	σ_{S11}	M	$\frac{\sigma_{R_1}}{R_1}$ actual	$\frac{\sigma_{R_1}}{R_1}$ Predicted by first order model	$\frac{\sigma_{R_1}}{R_1}$ Predicted by general model
1.00	1717.2	0.94341	0.00033	92.5	0.59817	0.60319	0.60322
3.57	36.857	0.15133	0.00200	269.8	0.41012	0.40921	0.40922
5.00	70.188	0.16799	0.00069	162.9	0.14228	0.14222	0.14222
10.0	49.051	0.01198	0.00005	70.2	0.00886	0.00955	0.00955
44.7	12.540	0.59900	0.00025	120.2	0.79350	0.79340	0.79340
80.0	33.406	0.20198	0.00064	16.8	0.13261	0.13234	0.13257
173	41.66	0.16099	0.00008	4.05	0.016035	0.015569	0.01603

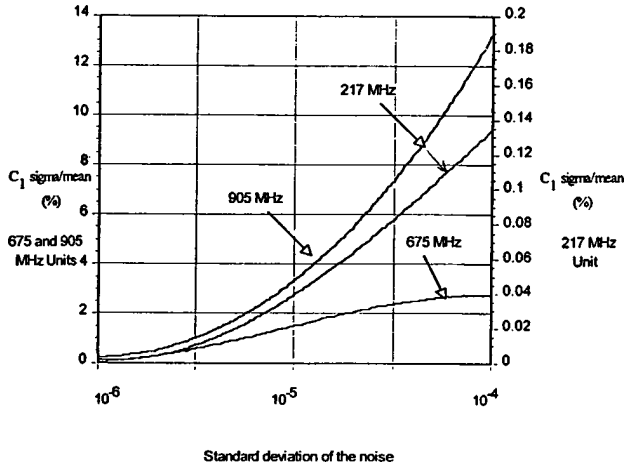


Figure 8. Motional capacitance uncertainties for SAW resonators.

In the sensitivity analysis of one port resonators, it was shown that the uncertainty in the extracted value of R_1 increases as the mismatch between R_1 and the characteristic impedance of the measurement system increases. One would expect the same type of relationship for two port resonators. Of the three two port resonators measured, the 905 MHz device was best matched to 50 ohms, and indeed it had the lowest uncertainty in the extracted value of R_1. However, the 675 MHz device is better matched to 50 ohms than the 217 MHz device, yet it has a much higher uncertainty in the extracted value of R_1, indicating that other factors besides the mismatch between R_1 and Z_0 contribute to this uncertainty.

In order to more closely model the physical behavior of the resonators, the first order longitudinal modes were extracted along with the fundamental mode in performing the simulations discussed above. It is suspected that the presence of the longitudinal modes effects the uncertainty in the extraction of the equivalent circuit parameters. Further investigation is needed to determine the details of these effects.

Since the values of L_1 and C_1 are determined from the reactance slope at resonance, it is reasonable to assume that the uncertainty in their extracted values is a function of this slope. Since reactance slope at resonance is not a commonly used resonator parameter, it is written here in terms of resonator quality factor (Q), series resonant frequency (f_s) and motional resistance (R_1):

$$\frac{dX_1}{df}\Big|_{f=f_s} = 4\pi L_1 = \frac{2R_1Q}{f_s} \qquad (14)$$

Equation (14) shows that higher Q, lower frequency resonators have higher reactance slopes at resonance. Table 3 shows the reactance slope at resonance for the three SAW resonators used in the simulation. Comparing the results of table 3 to figures 9 and 10 shows that the greater the reactance slope at resonance, the less uncertainty there is in the extracted values of L_1 and C_1.

The average standard deviation of the noise on an HP 8753 Network Analyzer with an S-parameter test set was measured to be between 10^{-4} and 10^{-5}. It was observed however that this value was somewhat less for frequencies less than 1 GHz where the data was taken on the devices reported here, and somewhat higher as frequency approached the 3 GHz mark. Unfortunately, the computer model which was used to perturb the data did not account for the frequency dependance of the standard deviation of the noise. Despite this weakness, it is expected that the computer model will give a reasonable first order approximation for the desired standard deviation to mean ratios.

To test this hypothesis, the results of ten measurements on a single resonator reported by Cavin [4] were compared the computer simulated results of table 2. Cavin's results are shown in table 4. These results fall within the range seen for the three devices used in the computer simulation with a noise standard deviation around 10^{-5}, supporting the hypothesis stated above.

Smaller noise variations can be obtained by reducing the receiver resolution bandwidth with a tradeoff is in the time it takes to perform the measurement. From the graphs of figure 8 we see that there is much more uncertainty in the extraction of the motional resistance for the 675 MHz resonator than for the other two. One might consider using a lower resolution bandwidth at the expense of a longer measurement time in order to reduce the noise standard deviation and therefore the uncertainty in the extracted value of R_1.

Table 3. Quality factor and Reactance Slope at Resonance for three SAW resonators.

	Q	Reactance slope at Resonance
217 MHz SAW	23.2×10^3	20.6 KΩ /MHz
675 MHz SAW	25.9×10^3	2.46 KΩ /MHz
905 MHz SAW	1.67×10^3	0.276 KΩ /MHz

Table 4. Results from ten measurements of a dual port SAW resonator [5].

Run #	R_1 (Ohms)	L_1 (mH)	C_1 (fF)
1	202.1	2.051606	1.506671
2	207.6	2.098436	1.473076
3	207.6	2.125072	1.454613
4	208.1	2.168308	1.425617
5	208.1	2.103242	1.469715
6	206.6	2.109979	1.465017
7	206.5	2.129476	1.451595
8	206.1	2.103148	1.469776
9	204.6	2.084518	1.482911
10	205.8	2.089020	1.479716
Mean	206.31	2.1062805	1.4678707
Std.Dev./ Mean (%)	0.8507%	1.3918%	1.3862%

VI. CONCLUSION

A general sensitivity analysis for single port resonators was developed which extends the previous work done on a first order analysis [5]. The model predicts the standard deviation to mean ratio of motional resistance values extracted from measured S-parameter data obtained through the use of an automatic network analyzer. Excellent agreement between simulated, measured, and theoretically predicted results are observed.

A computer program was written which takes an actual set of S-parameter data measured for two port SAW resonators and randomly perturbs the magnitude of each data point. This process models the effect of electrical noise and allows the user to quickly determine how much uncertainty is introduced into his or her measurements as a result of noise. The program then re-extracts the equivalent circuit parameters. The standard deviation of the perturbations can be adjusted to allow the user to determine the effects of different noise standard deviations.

VII. REFERENCES

[1] Park, K., Malocha, D, and Belkerdid., "Quartz Crystal Resonator Model Measurement and Sensitivity Analysis." 43rd Annual Frequency Control Symposium Proceedings, Denver, CO, May 1989, pp. 372-376.

[2] Hafner, E., "The Piezoelectric Crystal Unit - Definitions and Methods of Measurement." Proceedings of the IEEE, Vol 57, February 1969, pp 179-201.

[3] 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.

[4] Cavin, M.S., "Surface Acoustic Wave Resonator Parameter Extraction and Oscillator Design." Thesis: University of Central Florida, 1991.

[5] Malocha, D and Belkerdid, M, "First-Order Sensitivity Analysis of Quartz Crystal Resonator Model Parameters." IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol 37, No. 6, November 1990

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

An Analysis of the S.A.W. Displacements in Quartz and Lithium Niobate by X-ray Topography.

A. ZARKA*, B. CAPELLE*, M. PILARD*, J. SCHWARTZEL**, J. DETAINT** & M. SOLAL***

* L.M.C.P. Université P. et M. CURIE. 4 Place Jussieu. 75252 Paris Cedex 05.France

**FRANCE TELECOM CNET/PAB Laboratoire de Bagneux 196 Av. H. Ravera BP 107. 92225. Bagneux.France.

*** THOMSON SINTRA DTAS 399 Route des Crêtes, 06903. Sophia-Antipolis. France.

ABSTRACT.

In this report we develop investigations which have been made to study the energy distribution of different type of surface acoustic waves in different devices using quartz and lithium niobate.The observations were performed with X-ray topographs obtained by the use of the synchrotron radiation. The time structure of the X-rays has permitted to obtain stroboscopic observations of the amplitude and phase distributions of the Rayleigh waves across the devices. Bragg and Laue patterns were obtained in order to analyse the components of the displacements. The most important results concern the interactions between the dislocations and the S.A.W.

1. INTRODUCTION

The S.A.W. devices are important components in modern signal processing and many new developments have been made to increase their performances. In this report we develop investigations which have been made to study the energy distribution of different type of surface acoustic waves in different devices using quartz and lithium niobate.

The observations were performed with X-ray topographs obtained by the use of the synchrotron radiation at L.U.R.E.(Orsay, France). The time structure of the X-rays has permitted to obtain stroboscopic observations of the amplitude and phase distributions of the Rayleigh waves across the devices. Bragg and Laue patterns were obtained with the use of the white X-ray beam in order to analyse the components of the displacements.

2. EXPERIMENTAL SETTING

The used technique (figure 1) was that employed in previous experiments and given in details for example in (1). The stroboscopic X-ray topographs are obtained using a double axis spectrometer at L.U.R.E. (2). This spectrometer presents the advantage that when the sample is placed in the incident white beam a Laue pattern of topographs can be obtained (3). Another interesting aspect of this apparatus consists in the monochromatic setting. In this case, on the first axis is adjusted a 110 sample of germanium which selects from the white beam only one wavelength. Furthermore in the two previous cases, using a very fine slit section topographs can also be obtained. For crystals or filters too thick to be examined in a transmission Laue setting (4), reflection setting (4) with an oblique incidence was used.The film was adjusted parallel to the sample at a distance (10-20 cm) choosen to obtain in each case a good compromise between the resolution and the separation of the diffraction spots.

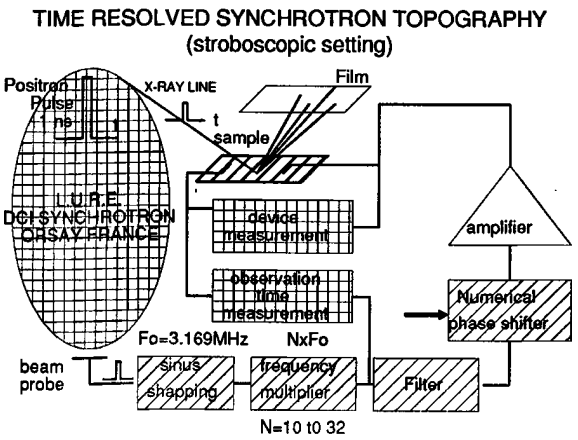


FIGURE 1

0-7803-1945-1/94 $3.00 © 1994 IEEE

3. QUARTZ

S.A.W. PROPAGATION IN ST QUARTZ

PROPERTIES.

The displacement vector of the quasi-Rayleigh wave propagating on an ST-cut plate in the X direction has three components in contrary to the "pure" Rayleigh waves existing on isotropic material. All of the three components are important.

These waves are very interesting for practical application due to the cancellation of the temperature coefficient of the delay at room temperature.

All the components of these waves are represented in the figure 2a using the coordinates system represented below. These are given in function of the normalised X3 coordinate (Λ=L is the wavelenght).

The properties of these waves are shown in the table 1 and the corresponding curve 2b.

An example of a 9.507 MHz delay line is given on the figure 2c.

EXPERIMENTAL

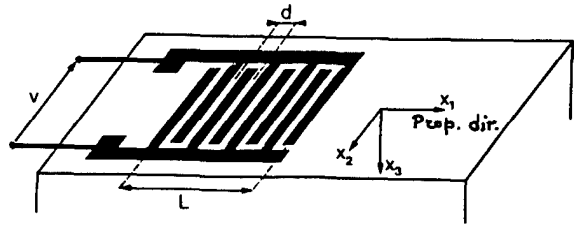
Different X-ray stroboscopic topographs have been obtained either by transmission or by reflexion. Furthermore several stroboscopic topographs have been obtained either by transmission or reflexion.

Furthermore, section topographs have been done to better understand the propagation of the waves inside the samples.

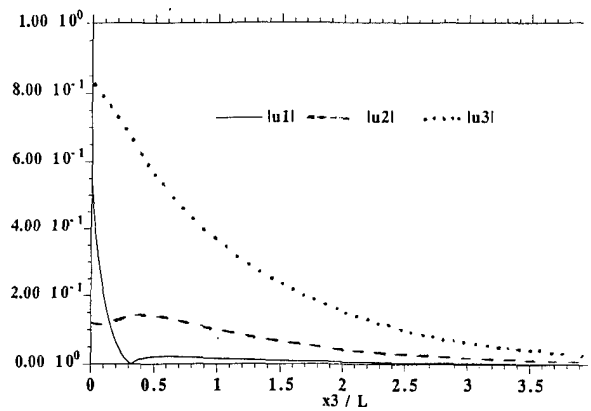
The stroboscopic observation permits to analyse the progressive wave by recording the image during a very short time slots (1ns) repeated every period of the surface wave.

	V	h_1	ϕ	CTf_1	CTf_2
QUARTZ ST	3159.2	0.14	0	0	-4.10^{-8} K^{-2}

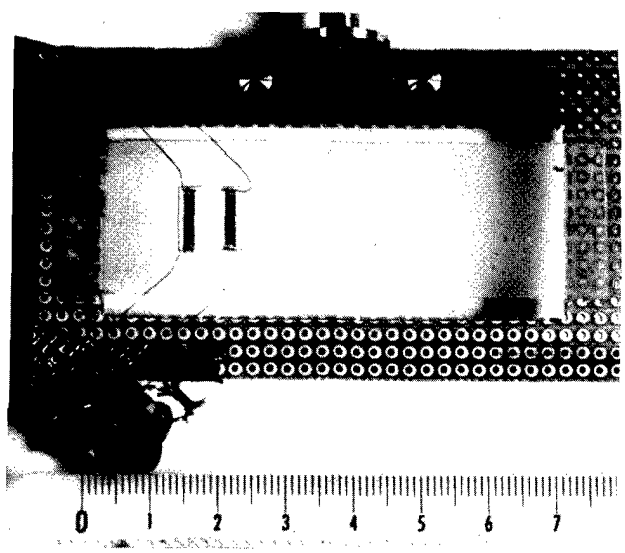
TABLE 1



a



b



c

FIGURE 2

Two types of quartz have been particularly studied and they correspond to devices working at 9MHz or at 22 MHz.
Examples of the waves propagation observed by X-ray topography are given in the following figures and concern a sample working at 9 MHz.

In figure 3 it may be observed a reflexion topograph in which the u3 component could be evidenced: the straight crested waves are seen as black contrast. One can also observed contrasts due to the emergence of the dislocations.

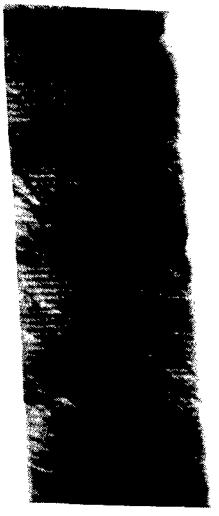


FIGURE 3

In figure 4 which corresponds to a "translation" topograph obtained using the 210 reflexion the u1 component can be observed.

In figure 5 which corresponds to a topograph obtained with the 032 reflexion the u2 component is very weak.

In figure 6 it may be observed a contrast which is due to mix of the different components obtained with the 101 reflexion.

The section topograph presented in fig 7 was obtained using the same previous reflexion. simulation of the obtained contrasts have been done and there is a good agreement between the calculated and the experimental images.

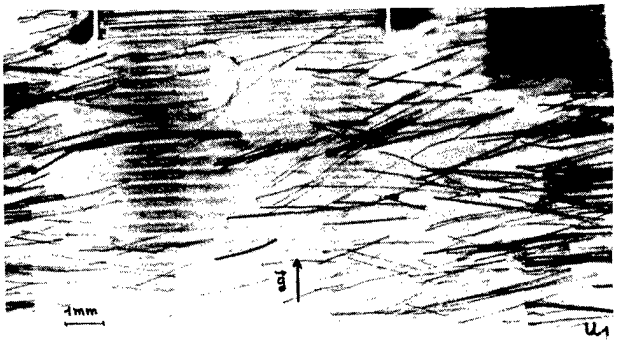


FIGURE 4

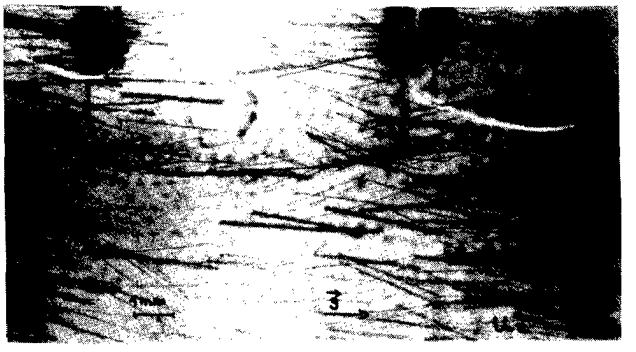


FIGURE 5



FIGURE 6



FIGURE 7

INTERACTION WITH THE DISLOCATIONS

In the different figures presented in the following it may be observed that the dislocations are present inside the crystal and they interact with the propagation of the S.A.W.

On figure 8 a strong interaction between the dislocation bundles which are organized in cellular form interact with the S.A.W. beam and modifies the energy distribution of the wave.

On the figure 9a is represented a video image of another crystal. The sample contains dislocations but is also perfect in some regions. The figure 9b is a stroboscopic video topograph of the same device under excitation. Surface waves appear as black vertical lines equally spaced. To observe the interactions between the surface waves and the defects, it is easy to compute the difference between the two images. This is shown in the figure 9c. The vibration fringes vanish where the density of the dislocations is high and they present a maximum contrast where the crystal is rather perfect. In the same manner an accute observation of the sample presented on the figure 8 permits to observe that the distance between the fringes are multiplied by two and even by four in the vicinity of the dislocations.

This was also observed in other cases and particularly for the 22 MHz. However, this effect is more important for the lower frequencies. (larger delay linewavelenght and depth penetration).



FIGURE 8

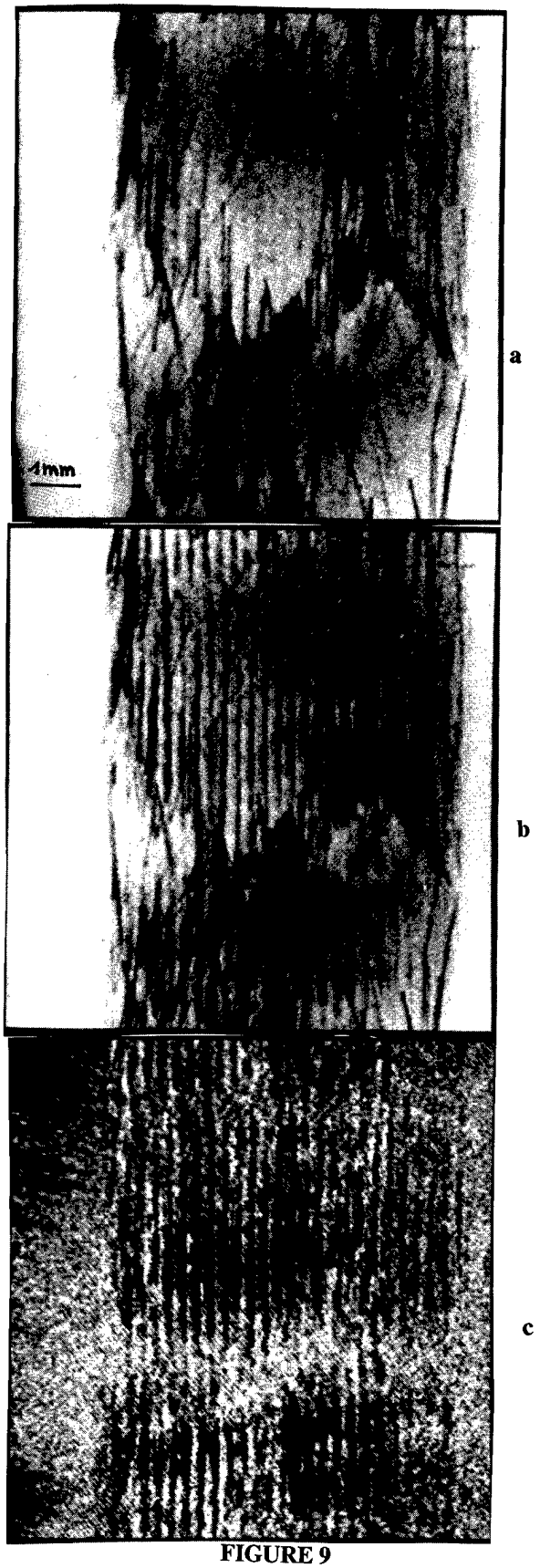


FIGURE 9

4. LITHIUM NIOBATE

Lithium niobate is a piezoelectric material displaying a very large coupling coefficient which is frequently used to obtain large bandwidth S.A.W. filters (5-6). Most often these devices make use of Rayleigh waves propagating on y plates in the Z direction.

Properties of the YZ propagating S.A.W.

The components of the displacement of these waves are represented in figure 10. The other properties of these waves are displayed in the table 2.

velocity (m/s)	3491.1
coupling coefficient	4.86
temperature coefficient	94 ppm
angle of power flow	0.0

Table 2

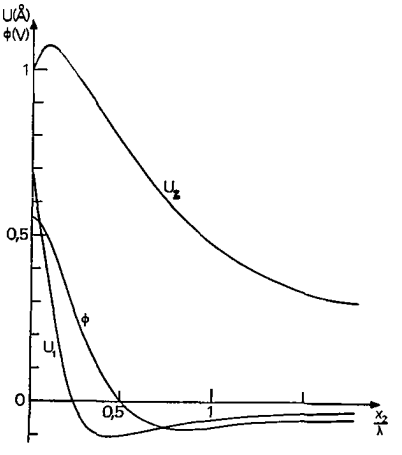


FIG. 6. 14. — *Onde de Rayleigh dans $LiNbO_3$ (coupe Y, propagation suivant Z). Variation des déplacements et du potentiel en fonction de la distance à la surface non métallisée.* (D'après la figure 1 de la référence 11).
Réf. Dieulesaint.

FIGURE 10

STUDIED DEVICES

Several high performance wide bandwidth filters made using multistrip couplers have been studied. The electrical response of one of these filters is given on figure 11. This filter has a 4.5 MHz bandwidth with low ripples at a center frequency of 60.216MHz.

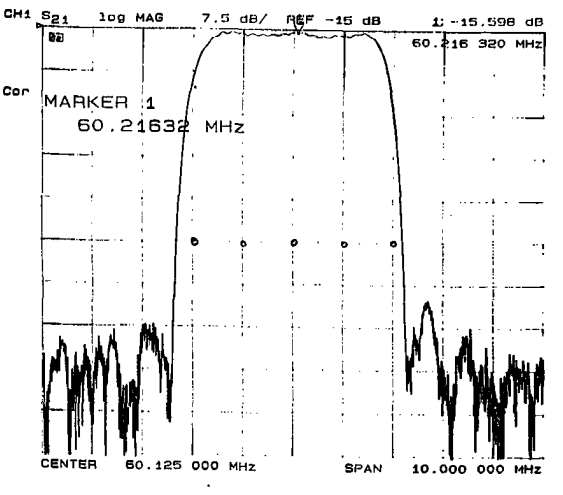


FIGURE 11

STUDIES OF S.A.W. FILTERS.

1. In order to find the best observation conditions to image the u3 component of the displacement we have varied the topographic parameter. Using different. X-ray wavelenghts we have obtained different configuration of the topographic contrasts due to the acoustic wave (figure 12).
On these topographs we can observe that the wave fronts don't present straight lines but distortions which correspond to the variations of the amplitude in the direction parallel to the fingers of the transducers.

2. Using a non synchronous excitation we have investigated the variation of the energy distribution in function of the frequency. On figure 13 which imagine the component normal to the surface it can be observed the variation of the radiation pattern of the transducer with the excitation frequency. in the same way, the S.A.W. beam is decomposed in several sub-beams whose number and shape vary also with the frequency.

3. Different patterns are observed if the excitation acess of the filter are changed ; this is represented on figure 14a it can be easily observed the modification of the pattern when the condition of the excitation vary. In figure 14b and 14c are shown the enlargement of the two different concerned zones.

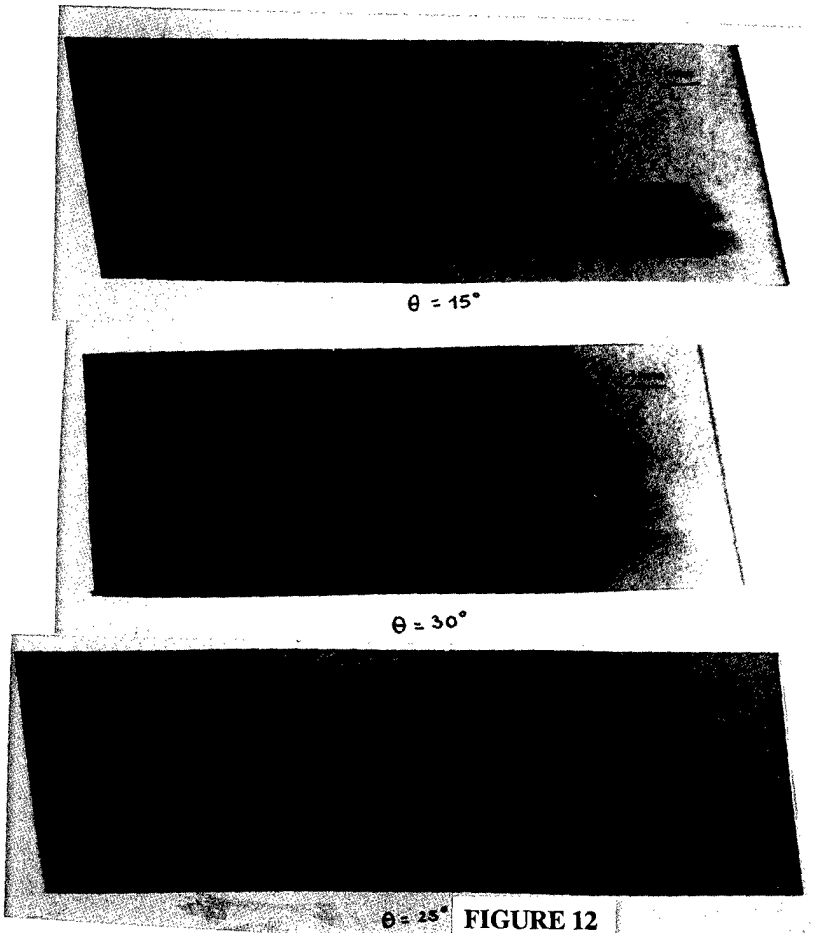


FIGURE 12

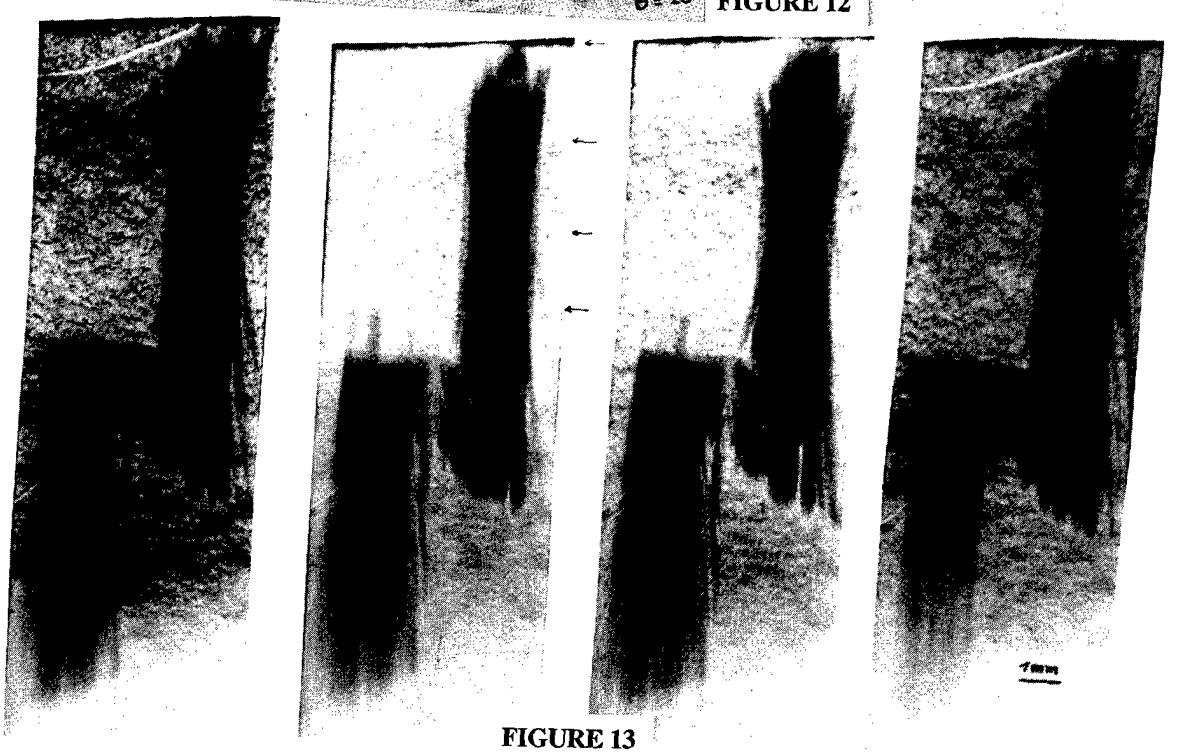
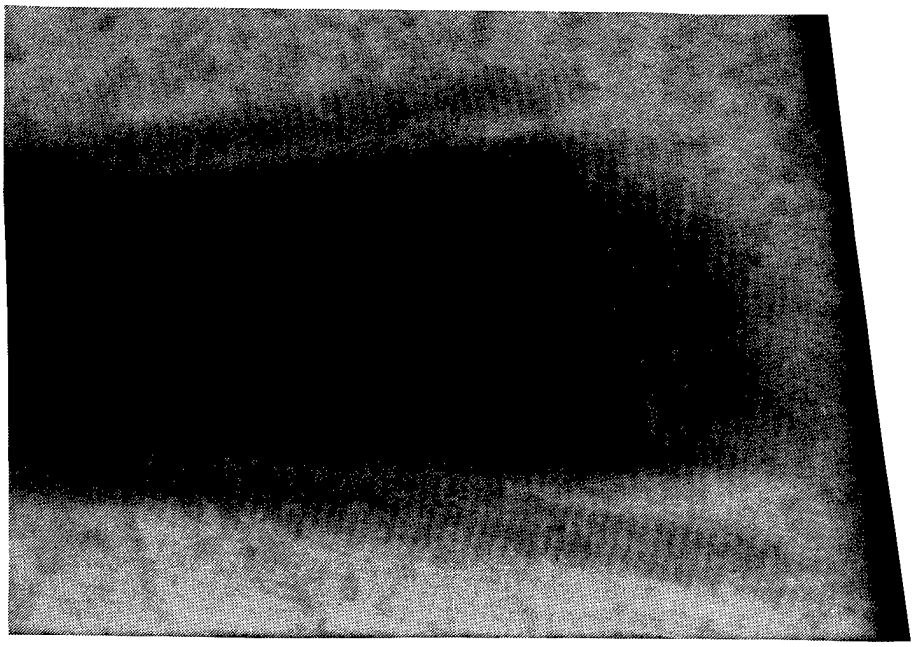


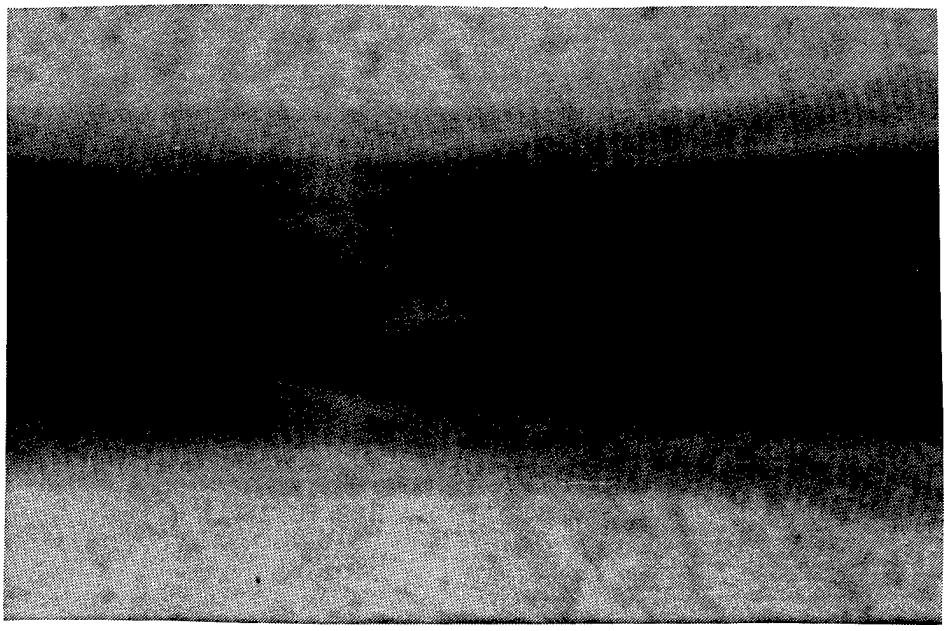
FIGURE 13



a



b



c

FIGURE 14

4. This topograph corresponds to a large excitation voltage. in this conditions different features are revealed. among them, small radiation lobes making an important angle with the main propagating direction can be observed. Another observation is made concerning the working mechanism of the acoustic absorbant placed at the open end of the transducer (figure 15a). An enlargement of the central part is given in figure 15b. After a computed treatment the fringes are evidenced in figure 15c.

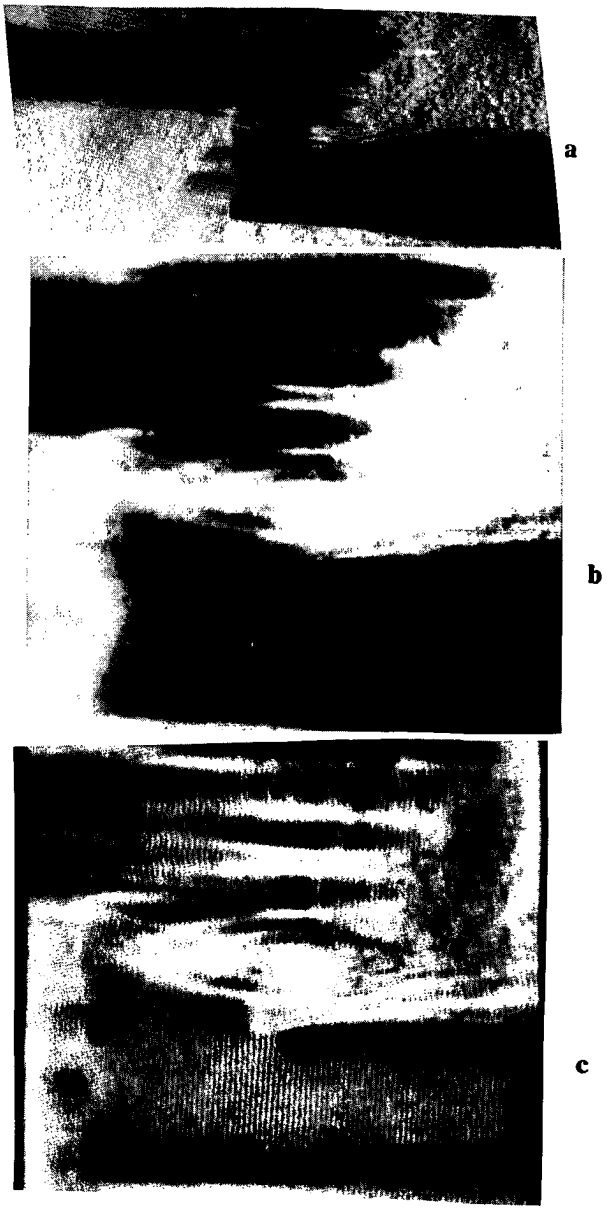


FIGURE 15

5. CONCLUSION

X-ray topography has permitted to investigate different features of S.A.W. propagation in several types of devices.

1/ This technique allows to visualise the different components of the S.A.W.. displacement.
2/ The whole device can be imaged in a single topograph.
3/ The stroboscopic setting allows to resolve the individual waves and to get information on their phase and amplitude distribution.
We have, for the first time, shown that strong interactions exist between the S.A.W. and the defects existing in the materials.
Similar observations have been made on other materials and devices and further experiments are in project.

6. REFERENCES

(1) ZARKA, A. ; CAPELLE, B. ; DETAINT, J.and SCHWARZEL, J. Proceeding of the 41st Annual Symposium on Frequency Control - 1987 ; 236-240 - Philadelphia.(U.S.A.).

(2) L.U.R.E. (Laboratoire pour l'Utilisation du RayonnementElectromagnétique) - Orsay - FRANCE.

(3) TUOMI, T. ; NAUKKARINEN, K. and RABE, P. Phys. Stat. Sol - 1974 - A.25, 93.

(4) CAPELLE. B.,DETAINT. J., SCHWARTZEL. J., ZHENG. Y. and ZARKA. A. Proc. IEEE Frequency Control Symposium (1992). 553-560.

(5) FELDMAN. M. and HENAFF. J. Traitement du signal par ondes élastiques de surface. Ed. Masson. (1986).

(6) MAERFELD. C. "Multistrip Couplers". Wave Electronic 2. (1976). 82-110.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THEORETICAL ANALYSIS OF THERMAL FREQUENCY BEHAVIOR OF A SAW DEVICE BEING CAUSED BY LASER BEAM IRRADIATION AND ITS DEPENDENCE ON ANGLE OF CUT

Nobunori Oura and Tatsuya Nakase
Precision and Intelligence Laboratory
Tokyo Institute of Technology
4259 Nagatsuta, Midoriku, Yokohama 227, Japan

Abstract

The frequency excursion due to intermittent laser beam irradiation on the surface of a SAW device plate is analyzed by assuming that only one dimensional thermal stress in the propagation direction of SAW on the surface of the plate exists and a diameter of the beam is infinitesimal. As a result, specimen plates immune to transient environmental temperature change are made and their frequency trends better than those of the ST-cut plate are shown.

1.Introduction

The SAW devices as frequency selecting and generating devices with small size and stable frequency performance are in great demand with the growth of mobile communications and consumer applications[1][2][3]. However, their frequency stabilities are inferior to those of the crystal oscillators using the bulk wave resonator by one to three order of magnitude.

In order to improve the frequency stabilities of the SAW devices, there appeared many papers [4][5][6] which deal with stress compensated cut of the plate such as SC-cut in the bulk wave resonator. The frequency temperature characteristics of the SAW device are divided into two categories; the static frequency temperature characteristics and transient(dynamic) ones[6][7][8]. The static frequency temperature characteristics[9] are those measured at the temperature where the temperature distribution on the device is quasi-homogeneous. The transient frequency temperature characteristics are known as the oscillating frequency excursion phenomenon in the bulk wave resonator with the environmental transient temperature change around the resonator. The phenomenon is caused by the thermal stress arisen in the resonator plate with temperature change.

A laser beam as a thermal probe was employed to make measurements of the sensitivity of the device to transient temperature change in it[11]. The oscillating frequency of a SAW oscillator was measured, while the laser beam intermittently irradiated every part of the surface of the device plate with cover removed. The frequency excursions due to the intermittent beam irradiation were plotted at every point on a map conformed to the device plate corresponding to the irradiation positions of the beam. The positions with the same frequency excursions were connected with line segments to form a contour map of frequency excursion. The map characterizes the points on the device plate sensitive to temperature change.

The paper deals with the frequency excursions due to heat transfer resulted from the laser beam irradiation by taking into account the thermal stress, which causes the frequency excursions, provided that for simplicity one dimentional thermal stress in the propaga-

0-7803-1945-1/94 $3.00 © 1994 IEEE

tion direction of the elastic wave is only considered, although the stress propagates in every direction.

2.Theoretical

2.1.Wave velocity without thermal stress

The oscillating frequency f of the SAW oscillator is roughly determined by the device employed as follows[11][12];

$$f = v/2p \tag{1}$$

where v is the propagation velocity of SAW, p the pitch of the interdigital electrodes. In this study the righthanded Euler's angle is employed to express the angle of cut of the device plate. The displacement components U_i of SAW and electric potential ϕ^I in the medium (I)(quartz) ($x_3 > 0$) are expressed by the following equations.

$$U_i = a_i exp(-\alpha k x_3) \cdot exp\{j(\omega t - k x_3)\} \quad (i = 1 \sim 3) \tag{2}$$

$$\phi^I = a_4 exp(-\alpha k x_3) \cdot exp\{j(\omega t - k x_1)\} \tag{3}$$

where $a_i (i = 1 \sim 3)$ are unknown amplitudes, k wave number ω/v, α attenuation constant, ω angular frequency, t time.

On the other hand, no displacement exists in the medium (II)(air) only an electrostatic potential ϕ^{II} exists.

$$\phi^{\mathrm{II}} = a_5 exp(-\gamma k x_3) \cdot exp\{j(\omega t k x_1)\} \tag{4}$$

The wave velocity is obatained by solving the equation of motion substituted piezoelectric equation and Laplace equation under the boundary conditions.

$$\rho \frac{\partial^2 U_i}{\partial t^2} = \sum_{k=1}^{3} \frac{\partial T_{ik}}{\partial x_k} \quad (i = 1 \sim 3) \tag{5}$$

$$\sum_{k=1}^{3} \frac{\partial D_k}{\partial x_k} = 0 \quad (i = 1 \sim 3) \tag{6}$$

$$T_{i3}\Big|_{x_3=+0} = 0 \quad (i = 1 \sim 3) \tag{7}$$

$$\phi^I\Big|_{x_3=+0} = \phi^{II}\Big|_{x_3=0} \tag{8}$$

$$D_3^I\Big|_{x_3=+0} = \varepsilon_0 \nabla \phi^{II}\Big|_{x_3=0} \tag{9}$$

The flowchart for computing the wave velocity is shown in Fig.(1)[11]. The elements X_{ij} of $detX$ in Fig.(1) are functions of stiffness, α, mass density of quartz and wave velocity. The elements Y_{ij} of $detY$ are functions of stiffness, α, β_{ij} unknown amplitude ratio a_i/a_4 dielectric constants and piezoelectric constants.

2.2.Wave velocity under thermal stress

It is assumed in the analysis that the cross section of the laser beam is infinitesimal and the absorption coefficient of laser light to quartz is 50%, though the cross section of a real laser beam has an area. The medium is regarded as an isotropic homogeneous body with the same elastic stiffness as that of quartz in the wave propagation direction.

When a heat source with heat capacity Q is generated at an instant at $t = 0$ at the origin of an infinite half space elastic body $z \geq 0$, the heat distribution τ at a point (x, y, z) in the body in a quasi-static state is expressed by the following equation[13].

$$\tau = \frac{Q}{(4\pi k t)^{3/2}} e^{-R^2/4kt} \tag{10}$$

where k thermal conductivity, Q heat capacity, $R^2 = x^2 + y^2 + z^2$.

The thermoelastic displacement potential is given by Equation (11).

$$\Phi = -\frac{K_1}{2G} \cdot \frac{1}{R} erf\left(\frac{R}{2\sqrt{kt}}\right) \tag{11}$$

where $K_1 = (1 + \nu)\alpha Q G / 2\pi(1 - \nu)$, α expantion coefficient, G shear modulus, ν Poisson's ratio.

The stress component σ_{rr} is obtained by solving the thermoelastic equations with Φ under the boundary condition.

$$\sigma_{rz}\big|_{z=0} = 0 \tag{12}$$

Therefore the thermal stress σ_{rr} to be obtained is given as follows:

$$\sigma_{rr} = \sigma'_{rr} + \sigma''_{rr} \tag{13}$$

where σ' and σ'' are given as follows;

$$\begin{aligned}\sigma'_{rr} &= -\frac{K_1}{R^3}\left[(2-\frac{3z^2}{R^3}erf(\frac{R}{2\sqrt{kt}})\right. \\ &\left. -\frac{R}{\sqrt{\pi kt}}e^{-R^2/4kt}\{2-\frac{3z^2}{R^2}(1+\frac{R^2}{6kt})\}\right]\end{aligned} \tag{14}$$

$$\begin{aligned}\sigma''_{rr} &= \frac{2G}{1-2\nu}\int_0^\infty B\lambda^3\Big\{(1-\lambda z)J_0(\lambda r) \\ &+(2\nu-1+\lambda z)\frac{J_1(\lambda r)}{\lambda r}\Big\}e^{-\lambda z}d\lambda\end{aligned} \tag{15}$$

The unknown quantity $B(\lambda t)$ is obtained to satisfy the boundary condition $\sigma' + \sigma'' = 0$ at $z = 0$.

$$B(\lambda, t) = \frac{12\nu}{2G\lambda}erf(\lambda\sqrt{kt}) \tag{16}$$

The elastic constants under stress are given by the following equation[4][6][14][15].

$$G_{L\gamma Q\alpha} = c_{L\gamma Q\alpha} + \hat{c}_{L\gamma Q\alpha} \tag{17}$$

where

$$\begin{aligned}\hat{c}_{L\gamma Q\alpha} &= T^1_{LQ\delta\gamma\alpha} + c_{L\gamma Q\alpha AB}E^1_{AB} \\ &+c_{L\gamma PQ}w_{\alpha,P} + c_{LPQ\alpha}w_{\gamma,P}\end{aligned} \tag{18}$$

where T^1_{LQ} is the static external stress, $c_{L\gamma Q\alpha AB}$ third order stiffness, E^1_{AB} strain, $c_{L\gamma PQ}$,$c_{LPQ\gamma}$ second order stiffness, $w_{\alpha,P}$,$w_{\gamma,P}$ displacement slope.

The frequency excursion of the device intermittently exposed to the laser beam irradiation is calculated by using the foregoing equations concerning the thermal stress components and the elastic constants under the stress. Figure (2) shows the flowchart of the calculation.

3.Comparison of experimental frequency excursion with calculated excursion

3.1 Experimental

An ST-cut device plate with cover removed was irradiated with a He-Ne laser beam with 21mW at room temperature along the line A shown in Fig.(3). Figure (4) shows the frequency excursion profile along the line A of which origin 0 is indicated by the black circle.

3.2 Calculated excursion

The wave velocity on the surface of the device plalte depends upon the magnitude of the thermal stress due to the irradiation at a surface point under consideration. The thermal stress differs with the distance from the heat source so that the wave velocity differs from point to point on the line A. In this calculation, the wave velocities were calculated at an interval of 0.1mm on the line A and averaged as follows;

$$T = \frac{0.1}{v_1} + \frac{0.1}{v_2} + \cdots + \frac{0.1}{v_n} \tag{19}$$

$$v' = 2.0/T \tag{20}$$

where T is the propagation time of the wave along the exciting electrode, v_n is a propagation velocity at point n, v' is the averaged SAW velocity.

The calculation was carried out under the following conditions; only stress component in the SAW propagation direction is considered. The electrode effect on the velocity is neglected. The absorption coefficient of the beam is assumed to be 50%. The heat source generated by the beam is a point source, though a diameter of the real beam is about 0.4mm. The quartz plate is an infinite half space.

The calculated values are superposed on the experimental so as to meet the maximum values of both the profiles as shown in Fig. (4) because the real absorption coefficient of the beam is unknown. In reality, the beam light of He-Ne laser is almost transparent for quartz. The beam light is absorbed to the electrode and adhesive of the plate to the holder base and the generated heat by absorption transfers to the plate. These heat processes make it difficult to calculate the

excursions. Nevertheless it is enough for comparing the profiles of the experimental relative excursion and the calculated relative excursion.

4. Dependence of the frequency excursion on the angle of cut of the device plate

The wave velocity at a point apart 1mm from the heat source was calculated and the frequency excursions at an instant of the irradiation were calculated for various angles of cut. Figure (5) shows the excursions for the propagation direction $\psi = 0°, 20°, 40°, 60°, 80°$. Figure (6) shows the excursion in the same plane as an ST-cut plate($\phi = 0, \theta = 128.5°, \psi = 0$) as a function of the propagation direction.

The specimen plates more sensitive or immune to the thermal stress were made as shown Table (1) by referring to the results in Fig. (5). The specimen were lapped and polished following the process in [16]. The size of the specimen, the electrode deposition process and the package are all the same as the plate shown in Fig.(3).

Figure (7) shows the frequency excursion of specimen F. The frequency jumped by 0.68ppm at the instant of the irradiation and decreased gradually with time owing to the static frequency temperature characteristics. Figure (8) shows the excursion of specimen G. Both the excursions show the same sense as predicted from the calculation.

On the other hand, specimen B with great immunity to the thermal stress predicted by the calculation shows the least excursion shown in Fig. (9). The frequency trends of the oscillators employing specimen B's or the ST-cut plate for one hour are shown in Fig.(10). The enlarged trends for three minutes are shown in Fig.(11). The trend measurements were made on the specimens with cover removed. The air often flowed onto the specimen plates with human movement in the room during the trend measurement.

The Allan variance of the specimens is listed in Table (2). The trends of the ST-cut specimens after 50 minutes are removed on computing the Allan variance.

The frequency trends suggest that the thermal transient frequency temperature characteristics are more important than those of the static ones for stabilizing the oscillating frequency of the SAW oscillators for a short term, though the existing SAW devices are manufactured from the viewpoint that the better the static temperature frequency characteristics, the stabler the frequency stability.

5.Conclusions

The thermal frequency excursion being irradiated a laser beam was analyzed under some of assumption. The frequency excursion profile of an ST-cut plate as a function of the position on the plate is of good agreement with the experimental and the calculated.

The specimens less immune to the thermal stress predicted by the calculation are superior to an existing STcut device in frequency stability by nearly one order of magnitude. The thermal transient frequency characteristics should be taken into account on manufacturing the stabler SAW oscillators rather than the static frequency characteristics.

The intermittent irradiation of laser beam on a SAW plate is applicable to such an application as laser power meter.

The authors greatly acknowledge Messrs. R. Murakwa and M. Yamakita for their help for preparing the manuscript.

6.References

[1] C.C.W.Ruppel et al.,"SAW Devices for Consumer Communication Application," IEEE Trans. UFFC, vol.40, NO.5.pp.438-452, September 1993

[2] I.D.Avramov,"Gigahertz Range Resonant Devices for Oscillator Applications Using Shear Horizontal Acoustic Waves,"IEEE Trans. UFFC, vol.40, NO.5.pp.459-468, September 1993

[3] G.K.Montress and T.E.Parker,"State-of-the Art SAW Oscillators:Design and Performance," Proc. 7th EFTF, pp.439-454, 1993

[4] E.K.Sinha;"A Stress Temperature Compensated Orientation and Propagation Direction for Surface Acoustic Wave Devices," IEEE Trans. UFFC, vol.UFFC34, No.1, January, pp.6474,1987

[5] S.Ballandras and E.Bigler,"Doubly Rotated SAW Quartz Cuts with Low Sensitivty to Stress and Temperature Effects Exibiting a Power Flow Angle Close to Zero," J.Acoust.Soc.Am., NO.95, Vol.3, pp.1390-1395, March 1994

[6] E.K.Sinha and W.J.Tanski;"Stress Induced Effects on the Propagation of Surface Waves," Proc. 37th AFCS, pp.415-422,1983

[7] D.Hauden et al.;"High Order Temperature Coefficients of Quartz SAW Oscillators," Proc. of the 28th ASFC, pp.7781.1974

[8] D.Hauden et al.;"Temperature Effects on Quartz Crystal Surface Wave Oscillators," Applied Physics Letters, vol.31, No.5, pp.315-317, 1977

[9] Y.Shimizu and Y.Yamamoto;"SAW Propagation Characteristics on α-Quartz with Arbitrary Cut," Trans. vol.J63A11, pp.799-806, 1980 (in Japanese)

[10] N.Oura et al.,"Transient Thermal Frequency Behavior of SAW Devices with a Laser Beam as a Thermal Probe," Proc. 7th EFTF,pp.239-242,1993

[11] The 150th Committee of Japan Society for the Promotion of Science ed.;"Elastic Surface Wave Devices Handbook," Ohom Press,Tokyo,1991 (in Japanese)

[12] K.Shibayama ed.;"Elastic Surface Wave Engineering," Corona Press, Tokyo, 1983 (in Japanese)

[13] Y.Takeuchi and N.Noda;"Analyses of Thermal Stresses," Nisshin Press, 1989 (in Japanese)

[14] H.F.Tiersten;"Perturbation Theory for Linear Electro Elastic Equations for Small Fields Superposed on a Bias,"J.Acoust.Soc.Am.,64, pp.832-837, 1978

[15] R.N.Thurston et al.;"Third Order Elastic Coefficients of Quartz,"J.Appl.Phys., 371,pp.267-265,1966

[16] H.Fukuyo et al.;"The RefractiveIndex Distribution Normal to the Polished Surface of Fused Quartz Measured by Ellipsometry," J.Appl.Phys. 50(5), pp.3653-3657, May 1979

Table 1: The specimen plates more sensitive or immune to the thermal stress

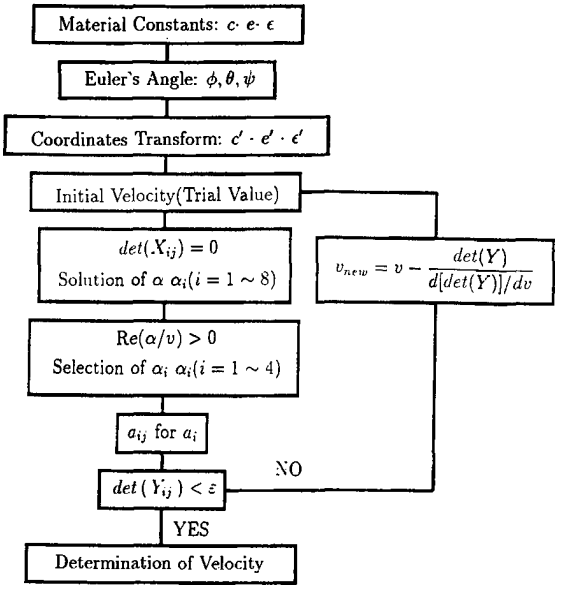
Specimen	θ[°]	ϕ[°]	Δ F(calculation)[ppm]	Powerflow angle[°]
F	128.5	9.7	0.328	3.224
G	128.5	42.6	2.112	1.674
B	140	0	0.003	0
I	40	22.6	-1.148	-0.765
ST-cut	128.5	0	0.089	0

Table 2: Frequency excursion of specimens of various angles of cut

	Measurement[ppm]	Calculation[ppm]
ST-cut	0.52	0.089
F	0.65 ~ 0.76	0.328
G	0.93 ~ 1.07	2.112
B	0.03	0.003
I		-1.148

Table 3: The Allan variances of the specimens

Specimen	Allan variance
B-u	2.29×10^{-9}
B-i	2.40×10^{-9}
B-e	2.41×10^{-9}
ST-cut1	1.05×10^{-8}
ST-cut2	1.30×10^{-8}



c: stiffness e: piezoelectric constant ϵ: dielectric constant

Fig.1 Flowchart for Computing Rayleigh Wave Velocity

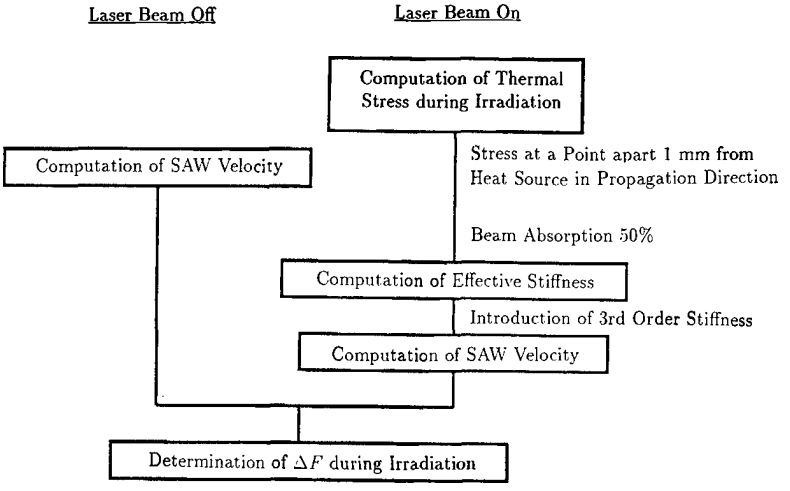


Fig.2 Flowchart for Calculating Theoretical Velocity under Thermal Stress

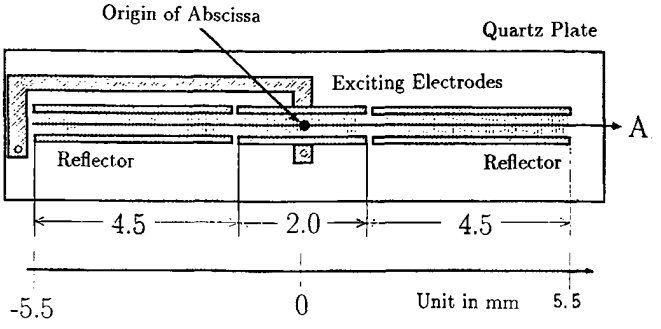


Fig.(3)Plan View of a Specimen Plate

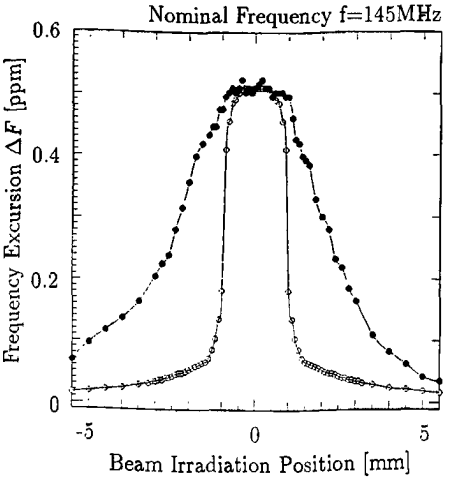


Fig.(4)Relation between Irradiation Position and Frequency Excursion ΔF

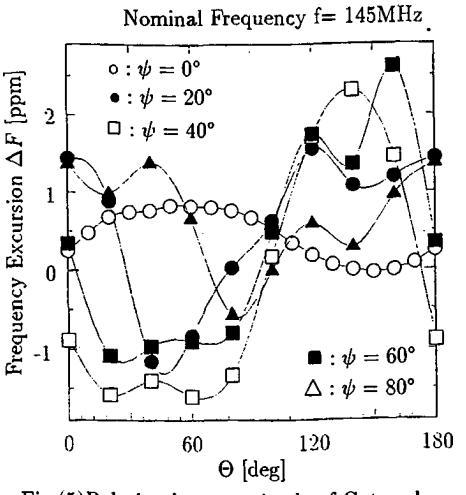


Fig.(5)Relation between Angle of Cut and Frequency Excursion ΔF due to Irradiation ($\varphi = 0°$ for all the specimens)

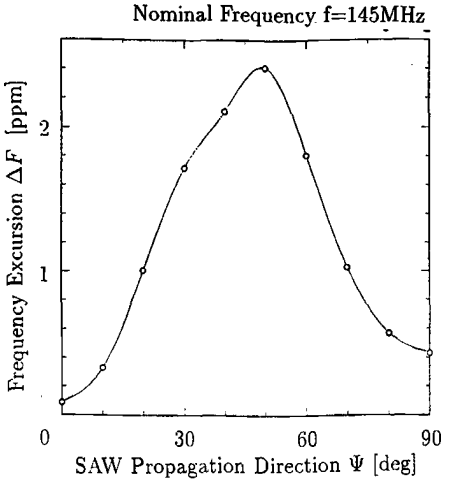


Fig.(6)Relation between Propagation Direction Ψ and Frquency Excursion ΔF ($\phi = 0°, \theta = 128.5°$)

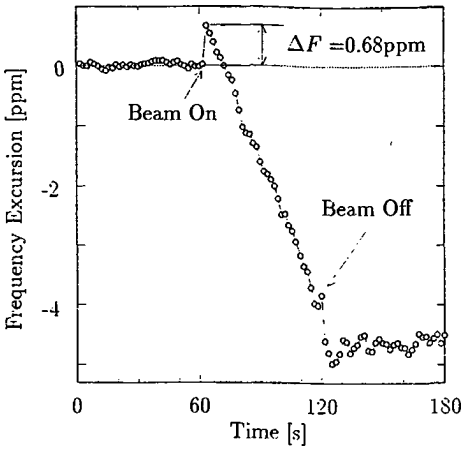


Fig.(7)Frequency Excursion with Beam Irradiation (specimen F)

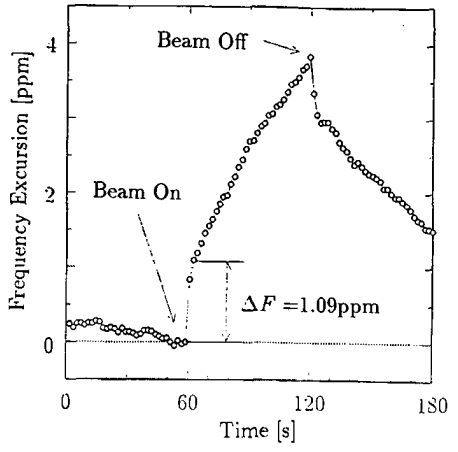


Fig.(8)Frequency Excursion with Beam Irradiation (specimen G)

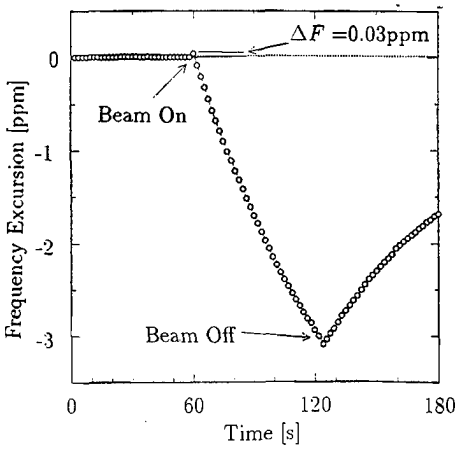


Fig.(9)Frequency Excursion with Beam Irradiation (specimen B)

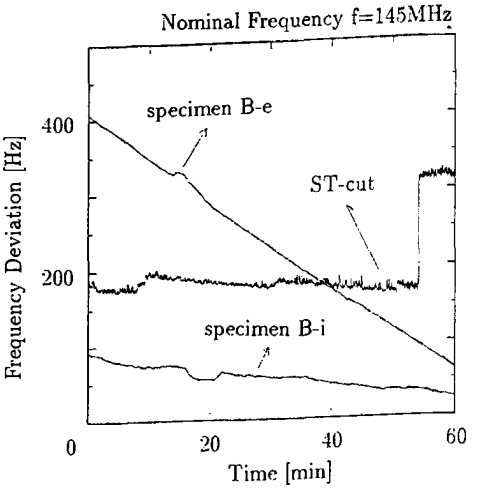


Fig.(10)Frequency Trends of Specimen B's and ST-cut with Cover Removed

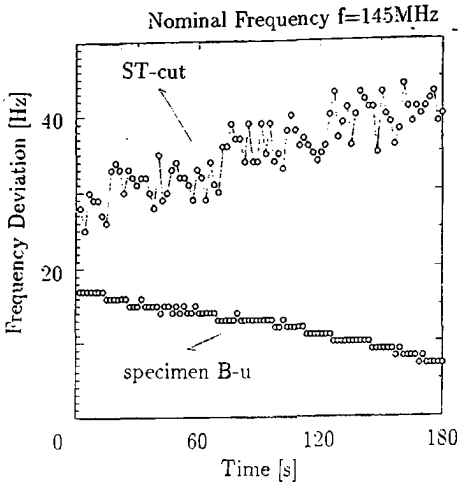


Fig.(11)Frequency Trends of Specimen B and ST-cut with Cover Removed

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SAW RING FILTERS WITH INSERTION LOSS OF 1 DB

Sergei A. Dobershtein and Victor A. Malyukhov

Omsk Research Electronic Equipment Institute
231 Maslennikov Street,
Omsk 644063, Russia

Abstract: This paper presents low loss ring SAW filters on 49°YX, 64°YX, 128°YX $LiNbO_3$ realized on the basis of the reflecting multistrip couplers (RMSCs). Using the RMSCs with 3 electrodes per λ (λ -is the SAW wavelength at the center frequency) and self-resonance approach when filter specified real input/output impedances are achieved by IDT static capacitance compensation with radiation susceptance, ring filters have shown very low insertion losses of 0.8-1 dB, 3 db fractional bandwidth of 2-5 % with very low ripple of 0.1 dB, stopband attenuation over 50 dB at 10-33 % offset from the center frequency of 45 MHz. At 50 Ohm 148, 164, 172 MHz ring filters on 128°YX for the low power transceivers have provided insertion loss of 1 dB, 1 dB bandwidth of 2-2.3 MHz, the stopband attenuation over 55 dB at 25 MHz offset from the center frequency. Two cascaded filters at 164.5 MHz have shown insertion losses below 3 dB and stopband attenuation over 90 dB. The chip size is 5x4x0.7 mm.

Introduction

Decreasing of SAW filter insertion losses up to 1 dB will bring to wide applications of SAW technology in modern communication systems [1]. Different techniques are used for low loss SAW filter design. The conventional way of decreasing the insertion losses is the use of different constructions of unidirectional interdigital transducers (IDTs), but this way encounters great difficulties. The major difficulty lies in the fact that complexity of a specific construction and topology of such IDTs decreases essentially the flexibility of the design of SAW filters with limiting characteristics [2]. In this connection the use of a ring construction, in which the acoustic energy radiated in both directions by the input IDT localizes by the output IDT, is very perspective [3]. This is achieved for example by use of two reflecting multistrip couplers (RMSCs) [2,4] (Fig. 1). With this design it is easy to realize the high selectivity (above 50 dB [4]) as here the filter frequency response is defined by the product of the responses of RMSCs and input and output IDTs. Low insertion losses are defined by a structure of the RMSCs and

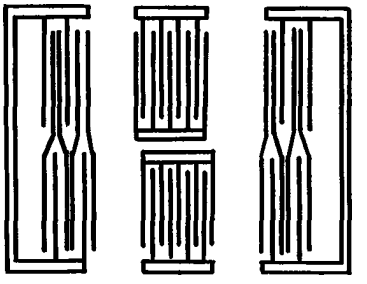


Fig. 1. SAW ring filter with RMSCs.

matching of the input/output IDT with the loads. Using the RMSCs with 3 electrodes per λ (Fig. 1) and self-resonance (self-matching) approach when filter specified real input/output impedances are achieved by IDT static capacitance compensation with radiation susceptance, ring filters can have the insertion losses under 3 dB [4,5]. Restrictions concerning ring filter fractional bandwidth and associated with the fact that the passband is proportional to K^2 (K^2 - is electromechanical - coupling coefficient) can easily be overcome by choosing a suitable cut of piezoelectric.

This paper presents low loss SAW ring filters on 49°YX, 64°YX, 128°YX $LiNbO_3$ realized on the basis of the above design aspects. The optimization of previously developed 45 MHz SAW ring filters [5] was provided for achieving insertion losses up to 1 dB. To increase the selectivity up to 50-80 dB at 5-10 % offset from the center frequency the phase weighting [6] and cascading are used. SAW filters for 146-174 MHz low power transceivers are presented to illustrate the applications of the developed filters. The results of investigation of the filter phase characteristics are given which are of great practical importance for low noise SAW oscillators and voltage controlled oscillators.

0-7803-1945-1/94 $3.00 © 1994 IEEE

Analysis of previously developed low loss SAW ring filters

Design and technology aspects of self-matched low loss ring filters on 49°YX, 64°YX, 128°YX $LiNbO_3$ are discussed in detail in work [5]. According to [5] 45 MHz unweighted ring filters on mentioned cuts of $LiNbO_3$ have shown insertion losses below 2 dB. We shall restrict our consideration to the analysis of losses and determine the methods of their further decreasing. A large body of performed experimental and calculation research showed, that the ring filter losses mainly include: losses in RMSCs; load mismatch losses; losses due to parasitic signals; conduction losses of Al-electrodes; propagation losses of leaky SAW on 49°YX, 64°YX. Thus the following investigations for every $LiNbO_3$ cut should be carried out with the view of decreasing these losses: determination of optimal number of RMSC electrodes providing minimum transmission losses between input and output IDT; obtaining filter specified pure real input/output impedance at the center frequency by self-matching effect with regard to influence of RMSCs, package, bond wires and other parasitic effects; reduction of tripple transit echoe (TTE) level; reduction of the conduction losses by choosing optimal Al-electrode thickness and width of bus-bars.

Determination of optimal number of RMSC electrodes

Investigations are carried out for 45 MHz filters with topology described in [5]. Unweighted IDTs had $N=1.5/K^2$ finger pairs. Number of RMSC electrodes for every LiNbO cut was chosen according to the table. All filter patterns were fabricated with identical metallization thickness and were connected to identical loads for every $LiNbO_3$ cut. The optimal number N_{opt} of RMSC electrodes was detemined by minimum insertion losses of the filter. Fig. 2 shows the frequency response of filter insertion losses on 64°YX vs number of RMSC electrodes. As will be seen from Fig. 2 $N_{opt}=90$ for 64°YX because in this case insertion losses are minimum. Similarly it was found $N_{opt}=225$ for 128°YX and $N_{opt}=62$ for 49°YX and this presented in Table 1.

Table 1. Insertion losses (I.L.) of the ring filter vs number of RMSC electrodes (N_{RMSC}) on 128°YX, 64°YX, 49°YX $LiNbO_3$.

128°YX		64°YX		49°YX	
N_{RMSC}	I.L., dB	N_{RMSC}	I.L., dB	N_{RMSC}	I.L., dB
150	2.0	79	2.3	48	2.4
225	1.3	90	1.4	62	1.3
300	2.2	97	2.3	75	2.5

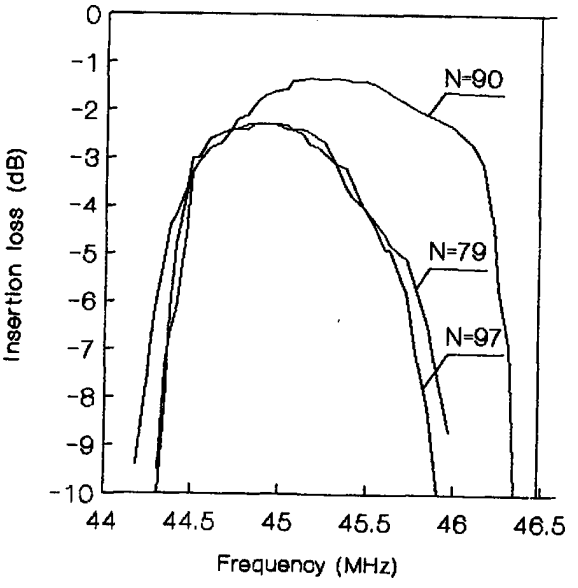


Fig. 2. Frequency response of filter insertion losses on 64°YX vs number of RMSC electrodes.

Reduction of TTE level

Performed experiments showed that the reduction of TTE level in previously developed ring filters [5] according to method [7], when input and output IDTs were shifted by $S=\lambda/4$, has not been adequately efficient. This method was evolved. The idea of the evolution is that tripple transit signals arriving right and left to the output IDT (Fig. 1) were summed there with opposite phases and useful signals with the same phases. For this purpose, as the simple calculations show, it is necessary that input and output IDTs were shifted by the distance $S=n\cdot\lambda/6$, n is odd. In this case (as experiments have shown) compared to method [7] amplitude and group delay passband ripple reduced, input/output isolation and self-matching improved and as a result the filter insertion losses decreased (Fig. 3).

Obtaining filter specified pure real input/output impedance by self-matching

Real input/output impedance of former ring filters is provided by choosing the number of IDT finger pairs N from the equation $N=1.5/K^2$ [5, 8]. This equation was obtained theoretically and experimentally for the single IDT with electrode width of $\lambda/4$, when IDT static capacitance is compensated with radiation susceptance. In the actual device - SAW ring filter self-matching effect is distorted due to reflected signals, influence of RMSCs, package, bond wires and other parasitic effects. Fig. 4 shows experimental input impedance characteristics of the filter on 49°YX described in [5]. As will

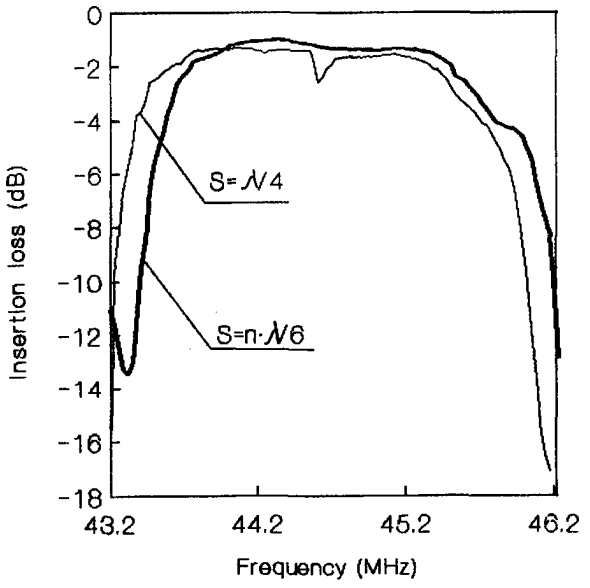


Fig. 3. Frequency response of filter on 49°YX within the passband vs shift S between input/output IDTs.

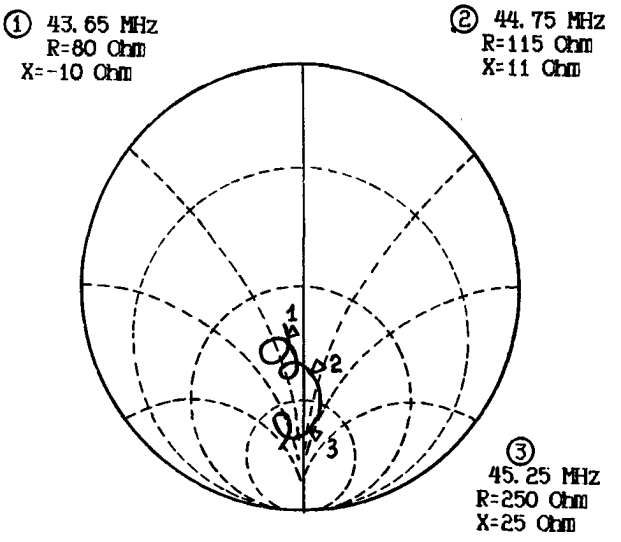


Fig. 4. Input impedance characteristics of the filter on 49°YX before the optimization.

be seen from Fig. 4 at the center frequency of 44.75 MHz input impedance has an inductive component, which is undersirable, since this causes the additional filter insertion losses. This effect can be eliminated by connecting of additional capacitance to input/output IDTs or changing the number of finger pairs in IDTs. In this case it is necessary to decrease the number of the IDT finger pairs as compared with the value $N=1.5/k^2$, and thus to decrease the inductive component of radiation susceptance. Similar effects were observed in the filters on 128°YX and 64°YX. The net number of the IDT finger pairs is chosen after several iterations till the filter measured pure real input/output impedance is obtained at the center frequency.

Choosing optimal Al-electrode thickness

At filter low insertion losses about 1 dB the conduction losses of IDT and RMSC Al-electrodes play essential role. Since filter topology (Fig. 1) eliminates the parallel connection of the IDTs, long IDT apertures (several tens of λ [4]) are required to obtain low input/output impedance (for example 50 Ohm). This causes the conduction losses in IDT and RMSC electrodes. These losses can be decreased only by increasing metal thickness and width of bus-bars. But photolithography is difficult under the excessive thickness: undercut of the electrodes is increased, reproducibility of the geometrical dimensions of IDT and RMSC topology is impared. This can cause the additional filter insertion losses, since the reflection condition in RMSC, and self-matching effect in IDT are disrupted. Thus optimal metal thickness should be defined, under which filter minimum insertion losses and specified tolerance on the topology reproducibility accuracy are achieved. Fig. 5 shows measured insertion losses of the filter on 128°YX vs metal thickness. Filter input impedance is about 85 Ohm. As will be seen from Fig. 5 at 45 MHz the filter has minimum insertion losses with metal thickness of 0.6 μm. For 64°YX, 49°YX the minimum insertion losses are recorded when metal thickness is 0.5-0.6 μm and input impedances are about 120 Ohm (in this case IDT apertures are the same as for filter on 128°YX with input impedance about 85 Ohm).

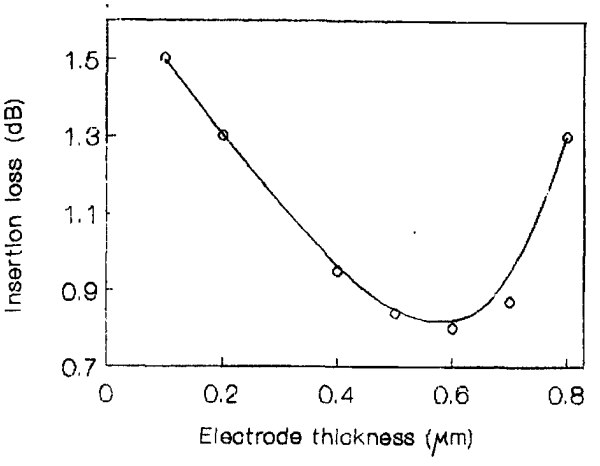
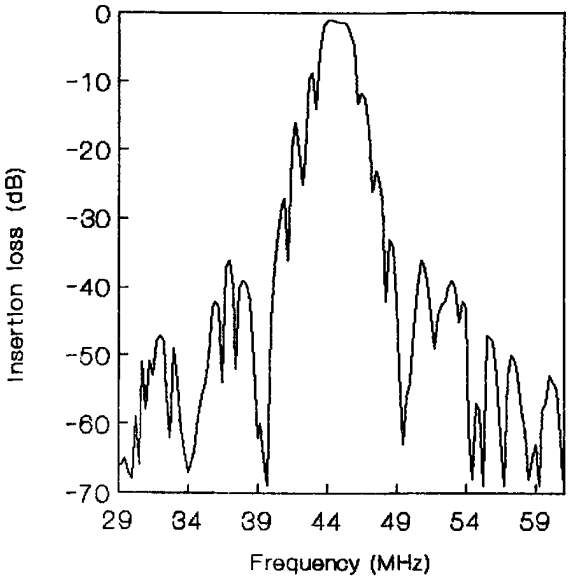
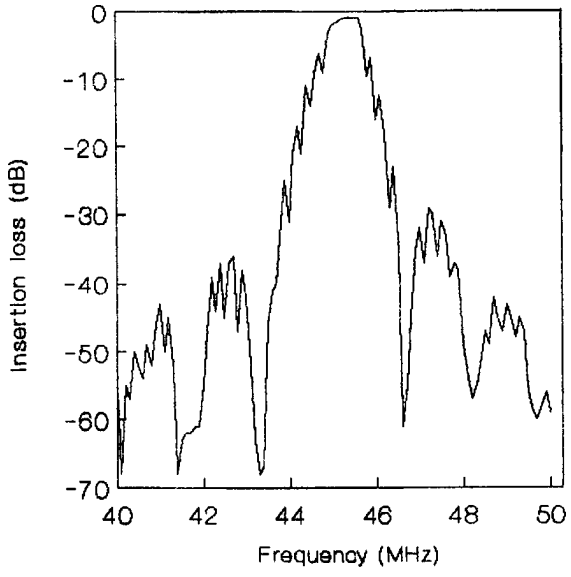


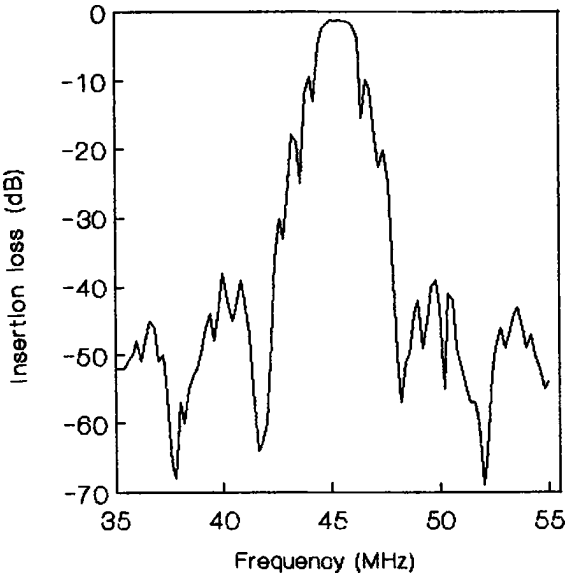
Fig. 5. Measured insertion losses of 45 MHz filter on 128°YX vs Al-electrode thickness.



a)



c)

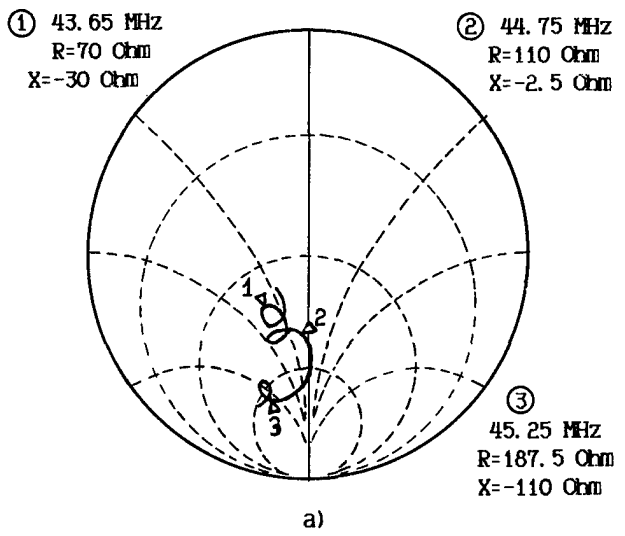


b)

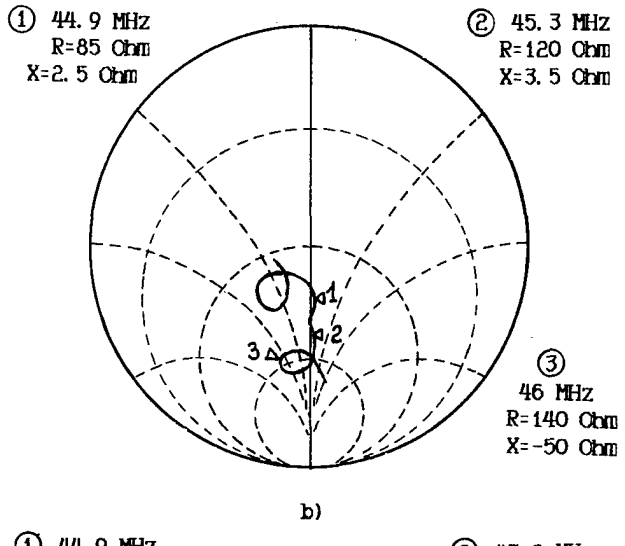
Fig. 6. Frequency response of the filters on 49°YX (a), 64°YX (b), 128°YX (c).

Development of ring filters with insertion loss of 1 dB

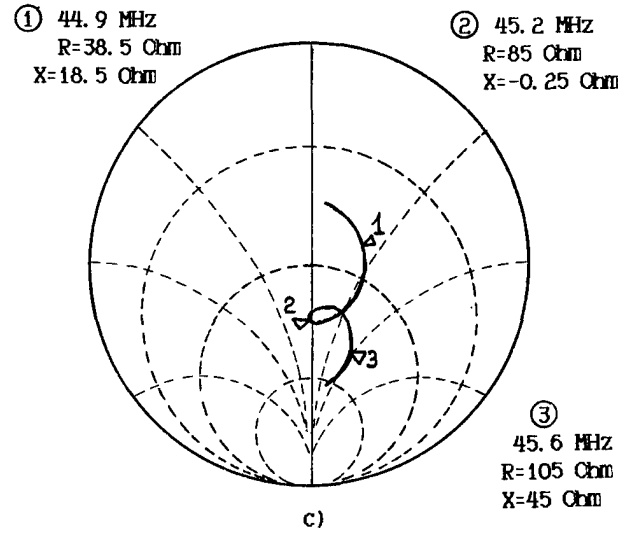
Development of ring filters taking into account mentioned investigations was realized by successive approximations using computer programs on the basis of the equivalent circuit model [5]. Besides in leaky SAW filters on 49°YX, 64°YX the insertion losses are determined under all things being equal by propagation losses which depend essentially on the distance between IDTs and metallization area between them and this was the reason to optimize filter topology by decreasing the distance between IDTs and RMSCs and width of bus-bars. The experimental patterns were fabricated in the intervening stages of the development. Insertion losses and filter input/output impedances were carefully measured and the calculations and filter topology were corrected. Frequency responses of ring filters on 49°YX, 64°YX, 128°YX are shown in Fig. 6, 8. Input/output IDTs are unweighted. Measured input impedance charateristics of the filters are shown in Fig. 7. As will be seen from Fig. 7 at the center frequency filter input impedances are close to real and are 85, 120, 110 Ohm for 128°YX, 64°YX, 49°YX respectively. The filters have been connected to loads equal to measured input/output impedance of the filters at the center frequency (Fig. 7), which has allowed to eleminate the mismatch losses and losses in the matching networks. At 45 MHz the ring filters have shown very low insertion losses from 0.8 to 1 dB, 3 dB fractional bandwidth from 2 to 5 % with very low ripple of 0.1 dB, stopband attenuation over 50 dB at 10-33 % offset from the center frequency. The filter characteristics are tabulated in Table 2. Measured phase responses for developed filters are presented in Fig. 8. Ring filters have linear phase characteristics with the phase shift of $\pm 180°$ at the 3 dB fractional bandwidth from 2 to 5 % depending on $LiNbO_3$ cut. Using such frequency dependent elements with insertion loss of 1 dB in the amplifier feedback loop it is possible to develop low noise and low power consumption SAW oscillators and voltage controlled oscillators. SAW filters for 146-174 MHz low power transceivers are presented in Fig. 9-11 as examples of applications of developed filters. Fig. 9 shows



a)

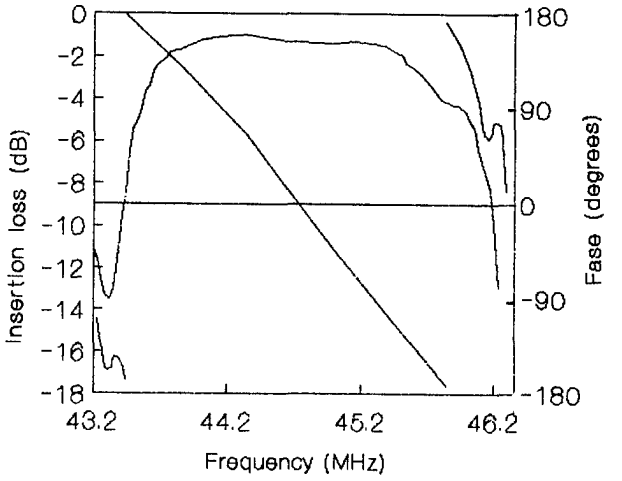


b)

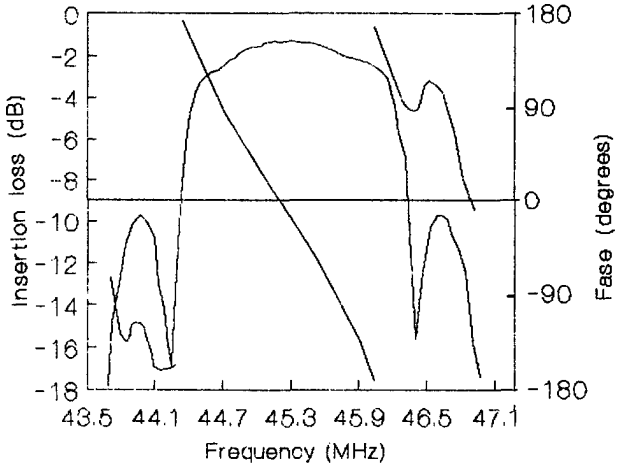


c)

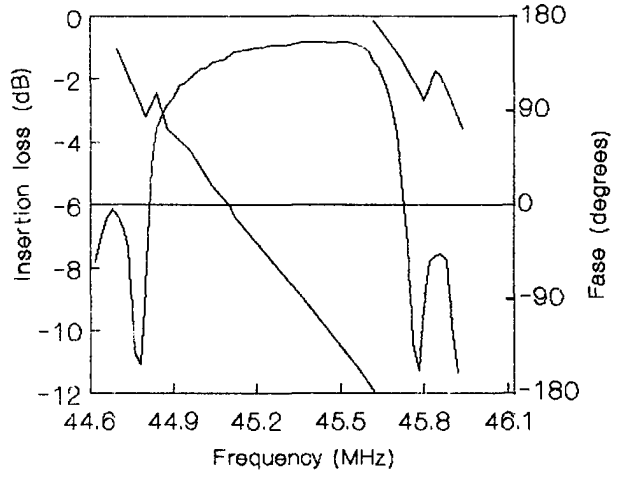
Fig. 7. Input impedance characteristics of the filters on 49°YX (a), 64°YX (b), 128°YX (c).



a)



b)



c)

Fig. 8. Frequency and phase response within the passband of the filters on 49°YX (a), 64°YX (b), 128°YX (c).

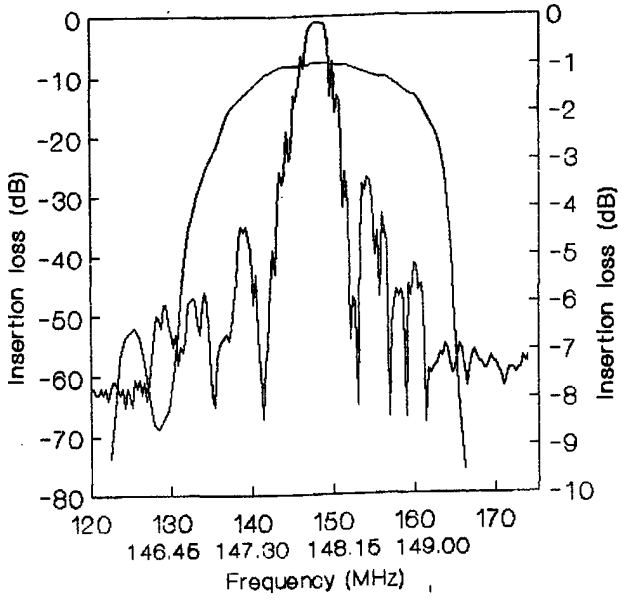


Fig. 9. Frequency response of 148 MHz unweighted filter on 128°YX.

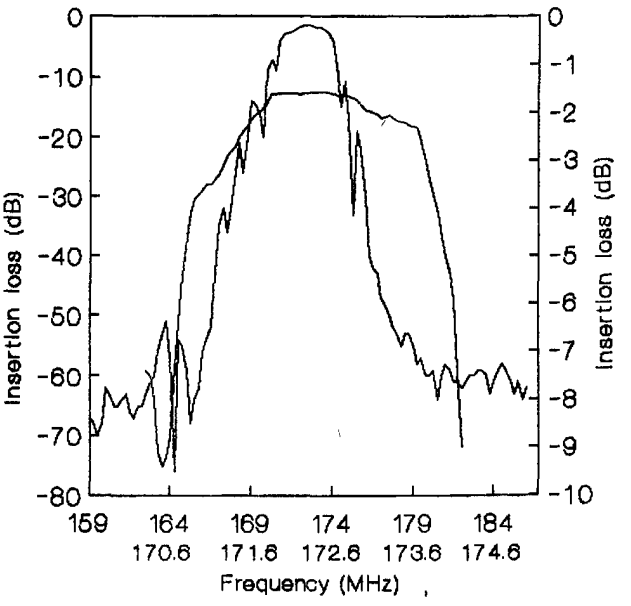


Fig. 10. Frequency response of 172.5 MHz weighted filter on 128°YX.

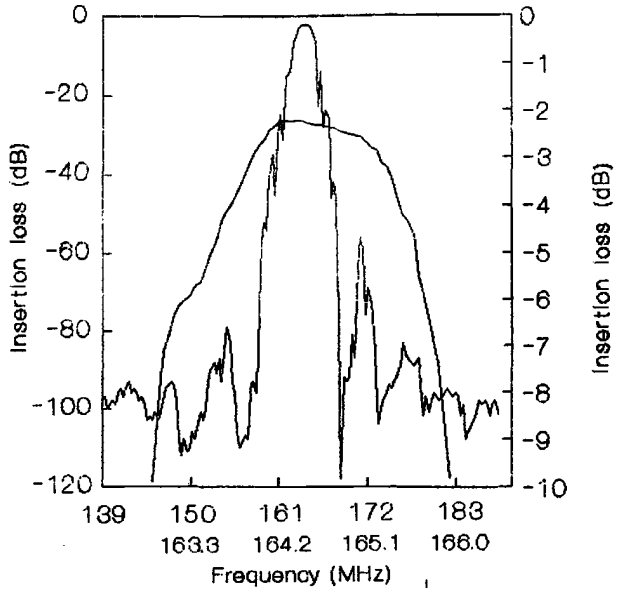


Fig. 11. Frequency response of the cascade of two 164.5 MHz unweighted filters on 128°YX.

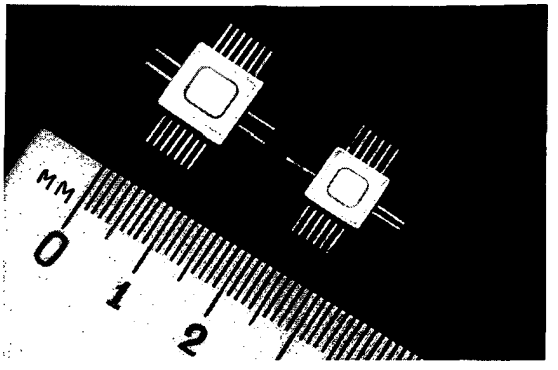


Fig. 12. Photograph of the developed ring SAW filters.

frequency response of 148 MHz filter on 128°YX. Similar filters were developed for the frequencies of 164.5 and 172.5 MHz. At 50 Ohm filters have shown insertion losses about 1 dB, 1 dB bandwidth of 2-2.3 MHz, the stopband attenuation around 60 dB at 25 MHz offset from the center frequency. Chip size is 5x4x0.7 mm. To increase the selectivity at low offsets from the center frequency the phase weighting should be used and at great offsets - cascading. Frequency response of 172.5 MHz filter with phase weighted IDTs is presented in Fig. 10. The Hamming weighted IDTs were 43 λ long. Stopband attenuation at 5 % offset from the center frequency increased up to 55 dB and the insertion losses impared up to 1.5 dB. Fig. 11 shows frequency response of the cascade of two unweighted filters at 164.5 MHz. At 50 Ohm such structure has shown insertion losses below 3 dB and stopband attenuation over 90 dB at $\pm$ 25 MHz offset from the center frequency. The filter characteristics are tabulated in Table 2. The filters can be mounted in planar packages compatible with SMT (Fig. 12). The developed low loss SAW filters can be used both in front-end stages of the mobile radio transceivers, increasing their sensitivity and image-frequency rejection, and in final stages reducing their power consumption, size and weight.

Table 2. Summary of results for low loss ring SAW filters.

Characteristics		Y-cut plates of $LiNbO_3$					
		49°YX	64°YX	128°YX			
Center frequency,	MHz	44.5	45.3	45.3	146-174	172.5*	164.5**
3 dB fractional bandwidth,	%	5	3.7	1.8	1.8	1.8	1.4
Insertion loss,	dB	1	1	0.8	1	1.5	2.5
Stopband attenuation at offset (%) from center frequency,	dB	55 (33%)	55 (21%)	55 (10%)	55 (10%)	55 (5%)	90 (10%)
Shape factor (3-40 dB)		3.2	3.4	3.3	3.3	2.5	3
Passband ripple,	dB	0.1	0.1	0.1	0.1	0.1	0.1
Chip dimensions,	mm	9x6x0.7	9x7x0.7	16x9x0.7	5x4x0.7	5x5x0.7	8x5x0.7

* - filter with phase weighted IDTs, ** - the cascade of two filters

Conclusion

Developed SAW ring filters with RMSCs have demonstrated very low insertion losses of 0.8-1 dB and fractional bandwidths of 2-5 % on different $LiNbO_3$ cuts. To extend the bandwidth leaky SAW on 49°YX and 64°YX are used. Careful analysis of the losses in RMSCs, TTE suppression, mismatch, conduction, and propagation losses allowed to obtain very low insertion losses and low ripple. Presented filter phase characteristics with linear phase shift of ±180°in the passband are very promising for low noise SAW oscillators and voltage controlled oscillators. The series connection of these filters yields excellent selectivity as well: 164.5 MHz cascade of two filters on 128°YX has provided insertion losses below 3 dB and stopband attenuation over 90 dB. The presented filters for 146-174 MHz transceivers have small-sized chip and no matching networks which allow to achieve optimum mass, dimensions and cost for the similar class of the low loss SAW filters. All these features allow to infer that developed SAW ring filters will be widely used in the electronic equipment.

Acknowledgement

Authors gratefully acknowledge the assistances of T. Marina for fabricating the SAW chips, V. Nokhrina and T. Slobodskikh for editing the manuscript.

References

[1] C.S. Hartmann, "Future High Volume Applications of SAW Devices", in Proceedings of the IEEE Ultrasonics Symposium, 1985, pp. 64-73.

[2] R.B. Brown, "Low-Loss Device Using Multistrip Coupler Ring Configuration With Ripple Cancellation", in Proceedings of the IEEE Ultrasonics Symposium, 1986, pp. 71-76.

[3] F. Sandy and T.E. Parker, " Surface Acoustic Ware Ring Filter", in Proceedings of the IEEE Ultrasonics Symposium, 1976, pp. 391-396.

[4] W. Pollock, J. Schofield, R.F. Milson, R.J. Murray and I. Flinn, "Low-Loss SAW Filter Using Single-Phase IDTs And No External Tuning", in Proceedings of the IEEE Ultrasonics Symposium, 1983, pp. 87-92.

[5] S. Dobershtein, V. Malyukhov, K. Nikolaenko and V. Razgonyaev, "Use of Self-Resonance Approach For High Selectivity Low Loss SAW Filters", in Proceedings of the IEEE Ultrasonics Symposium, 1992, pp. 151-154.

[6] M. Hikita, Y. Kinoshita, H. Kojima and T. Tabuchi, "Phase Weighting For Low Loss SAW Filters", in Proceedings of the IEEE Ultrasonics Symposium, 1980, pp. 308-312.

[7] P. Qui, C. Zhao and Z. Shi, " SAW Group-Type Multistrip Coupler", in Proceedings of the IEEE Ultrasonics Symposium, 1987, pp. 107-110.

[8] M. Hikita, T. Tabuchi, H. Kojima, A. Nakagoshi and Y. Kinoshita, "High Performance SAW Filters With Several New Technologies For Cellular Radio", in Proceedings of the IEEE Ultrasonics Simposium, 1984, pp. 82-92.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A SAW RESONATOR FILTER EXPLOITING RMSCs

E. Danicki
Polish Academy of Sciences, 21 Świętokrzyska Str., Warsaw 00-049, Poland

Abstract

Reversing multistrip couplers are applied as reflecting structures in the double channel SAW resonator filter. The reflection is purely electrical. Double channel operation of the resonator is advantageous for many reasons, one is that the resonator performance does not depend on the position of interdigital transducers in the cavity region, the feature that otherwise can be obtained by using much more complicated unidirectional transducers. The theory of RMSC is presented based on spectral analysis of the field in a general srutucture with nonidentical acoustic channels (as in RMSC beam compressor). Scattering matrix is derived. Final relations for SAW resonator filter are presented as well as some experimental results, including that obtained for "chirp" dispersive RMSC.

1 INTRODUCTION

Usefulness of reversing multistrip couplers (RMSC) [1] as components of SAW devices has long been unrecognized, probably because of lack of its theory. Recent theoretical and experimental results [2-6] show however that their unique properties can be advantageous in some applications, primarily in SAW resonators of low Q factor and wide tuning range.

The theory presented in next Section is based on the Bløtekjær's method of analysis [7]. RMSC is considered operating at its fundamental frequency, however, neglecting bulk waves, the results can be extended to the second overtone (twice the fundamental frequency) where RMSC works equally well. The theory presented below is a generalization of that given in [8] by allowing different strip dimensions in both acoustic channels. The main difficulty in the theory is caused by small parameters which must be carefully treated.

In reward, the theory of SAW resonator utilizing RMSCs as the resonator reflecting structures is exceptionally simple. This is because of double-channel operation of the resonator so that SAW excited by an IDT placed in the cavity region makes full round before it returns back to the IDT from its other side. In typical SAW resonator exploiting grating reflectors, SAW enters the IDT, from the same side, immediately after being reflected by the reflector. This makes such typical resonator sensitive to the IDT position within its cavity region [11], while the resonators exploiting RMSCs are not sensitive on that position.

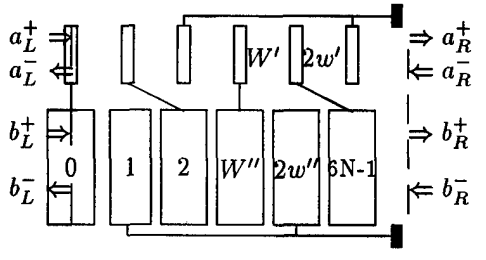


Fig.1. *General pattern of RMSC.*

2 THEORY OF RMSC

There are two acoustic channels in the considered RMSC (Fig.1), having aperture widths W' and W'', strips are $2w'$ and $2w''$ wide there. The strip period in both channels is $\Lambda = 2\pi/K$. Strips are periodically interconnected between the channels, and some are grounded, the pattern is periodic with period 3Λ which will be considered close to the SAW wavelength. This allows us to neglect Bragg reflection of SAW resulting from the strip elasticity and perfect conductivity, and to neglect bulk waves. Thus we can take full advantage of the theories presented in [7-10].

2.1 Theory of Periodic Strips

A harmonic wave-field $\exp(j\omega t - jkx)$ on the surface of piezoelectric body is characterized by angular frequency ω, wave-number k, and complex amplitudes of electric potential ϕ, electric field $E_{\|} = -\nabla_x \phi$, surface electric charge density $\Delta D_{\perp}$, a quantity proportional to the particle displacement U [10] and to SAW amplitude $a = \sqrt{\omega/2}U$, which is related to the SAW Poynt-

0-7803-1945-1/94 $3.00 © 1994 IEEE

ing vector magnitude $\Pi = |a|^2/2$. The relations between these quantities, which sufficiently characterize the stress-free piezoelectric halfspace, are [10]

$$U = \sqrt{\epsilon_e}(-\phi + \tfrac{1+\kappa^2}{|k|\epsilon_e}\Delta D_\perp)$$
$$|k|(|k| - k_v)\phi - (|k| - k_o)\Delta D_\perp/\epsilon_e) = 0$$

where $k_{v,o}$ - SAW wavenumbers at free and metallized substrate surface, $\kappa^2 = \Delta v/v = (k_o - k_v)/k_v$, ϵ_e - surface dielectric permittivity.

Following the method [7] the wave-field under the periodic strips is represented by series

$$E_\parallel = E_n e^{-j(r+nK)x}, \quad E_n = \alpha_m S_{n-m} P_{n-m}(\Delta)$$
$$\Delta D_\perp = D_n e^{-j(r+nK)x}, \; D_n = -j\epsilon_e \alpha_m P_{n-m}(\Delta) \quad (1)$$

and similarly for U, where $r \in (0, K)$, $m = -1, 0, 1$, $\Delta = Kw$ ($w = w'$ or w''), and P_ν is Legendre function. Proceeding further with the method, one obtains that the strip admittance $Y = I/V$ where I is the strip current and V is the strip electric potential, depends on r. That is also the case of $U_{0,-1}$, which are quantities related to the amplitudes of forward, and backward propagating SAWs ($a_{0,-1}$). In the considerations below we apply assumption of *small* value of $|\Delta|$, that is $P_\nu(\Delta) \approx P_\nu(0) + \Delta\nu P_{-\nu}(0)$ [12], to obtain

$$Y = \frac{I}{V} = j2\omega\epsilon_e P \frac{(r - r_v)(K - r - r_v)}{(r - r_o)(K - r - r_o)} \sin\pi\frac{r}{K}$$
$$P = W\frac{P_{-r/K}(\Delta)}{P_{-r/K}(-\Delta)}, \; k' = \frac{k_o + k_v}{2} + \Delta\frac{k}{2}\kappa^2$$
$$U_p = V\frac{2\kappa\sqrt{\epsilon_e}}{\Lambda P_{-r/K}(-\Delta)} \frac{K - (r - pK) - k'}{(r - r_o)(K - r - r_o)} \sin\pi\frac{r}{K} \quad (2)$$

where r_v and r_o are wavenumbers of SAWs propagating under free and sfort-circuited strips, correspondingly,

$$r_v = k[1 - \tfrac{\kappa^2}{2}(1 - \tfrac{r}{K})] - k\tfrac{\kappa^2}{2}\Delta\tfrac{r}{K}(2 - \tfrac{r}{K})$$
$$r_o = k[1 + \tfrac{\kappa^2}{2}(1 - \tfrac{r}{K})] - k\tfrac{\kappa^2}{2}\Delta\tfrac{r}{K}(2 - \tfrac{r}{K}) \quad (3)$$

2.2 RMSC Beam Compressor

In RMSC, the strip pattern has period 3Λ, thus the pattern wave-number is $K' = K/3 = K''$, equal in both acoustic channels (' and " mark quantities in the upper, and lower acoustic channels, for example be r'_o and Δ' should be applied in the above relations concerning upper channel). Thus, following the Floquet theorem, we apply following expansion for *discrete* wave-field components at the n-th strip center

$$\begin{Bmatrix} V'(n) \\ V''(n) \end{Bmatrix} = \sum_{m=0}^{2} \begin{Bmatrix} V'_m \\ V''_m \end{Bmatrix} e^{-j(\delta + mK')n\Lambda} \quad (4)$$

and similarly for $I'(n)$ and $I''(n)$, where δ is unknown yet. Below, we will consider RMSC working near its center frequency, where $r_{o,v} \approx K'$. Noticing that r (see previous Section) is now $r_m = \delta + mK'$, we expect a solution in $\delta \approx 0$. Substitution of r_m into Eqs.(2) yields Y_m and $U_p^{(m)}$ as follows (indices ' , " dropped)

$$Y_0 \approx 0, \; U_0^{(0)} \approx 0, \; U_{-1}^{(0)} \approx 0, \; U_{-1}^{(1)} \approx 0, \; U_0^{(2)} \approx 0$$
$$Y_{1,2} = j\sqrt{3}\omega\epsilon_e P(K' \pm \delta - r_v)(K' \pm \delta - r_o)^{-1}$$
$$U_{0,-1}^{(1,2)} \approx \frac{\kappa\sqrt{3\epsilon_e}}{\Lambda P_{-1/3}(-\Delta)} \frac{V_{1,2}}{K' \pm \delta - r_o}. \quad (5)$$

The eigenvalue problem is formulated as a set of equations obtained from the Kirchhoff's laws applied to strips in one period, in our case to strips 0,1,2 in both acoustic channels (Fig.1, below we apply notations $\alpha = \exp(-jK'\Lambda)$, $z = \exp(-j\delta\Lambda)$; useful identities are $\alpha^3 = 1$, $1 - \alpha = -j\sqrt{3}\alpha^2$, $1 + \alpha = -\alpha^2$)

$$V'_0 = -\alpha^2 V'_1 - \alpha V'_2, \; V''_0 = -\alpha V''_1 - \alpha^2 V''_2$$
$$\begin{bmatrix} 1 & -\alpha & \alpha & -1 \\ 1 & -1 & z & -z \\ Y'_1 & Y'_2 & Y''_1 & Y''_2 \\ Y'_1 & \alpha Y'_2 & \alpha z Y''_1 & z Y''_2 \end{bmatrix} \begin{bmatrix} V'_1 \\ V'_2 \\ V''_1 \\ V''_2 \end{bmatrix} = 0 \quad (6)$$

With new notations (' and " dropped, $p = 1, 2$)

$$y_p = \frac{K' - (-1)^p\delta - r_v}{K' - (-1)^p\delta - r_o}, \; X = \sqrt{\frac{P'}{P''}} \quad (7)$$

the determinant of Eqs.(6) that must be equal zero is

$$[y'_1 \sin(\tfrac{\delta\Lambda}{2} + \tfrac{\pi}{3}) + y'_2 \sin(\tfrac{\delta\Lambda}{2} - \tfrac{\pi}{3})]$$
$$\times[y''_1 \sin(\tfrac{\delta\Lambda}{2} + \tfrac{\pi}{3}) + y''_2 \sin(\tfrac{\delta\Lambda}{2} - \tfrac{\pi}{3})] +$$
$$(Xy'_1 + \tfrac{1}{X}y''_1)(Xy'_2 + \tfrac{1}{X}y''_2)\sin^2\tfrac{\pi}{3} = 0 \quad (8)$$

Below we apply possible choice $X = 1$ for the symmetry reason, and that there are $\Delta = \cos Kw'$ in the upper channel and $-\Delta = \cos Kw''$ - in the lower channel, as well as $r'_{v,o} = r_{v,o} - c\Delta$ and $r''_{v,o} = r_{v,o} + c\Delta$, correspondingly, where c is constant (3). These applied parameters allow to rewrite the left-hand side of the dispersive Eq.(8) in form

$$[p_v m_o \sin(\frac{\delta\Lambda}{2} + \frac{\pi}{3}) + m_v p_o \sin(\frac{\delta\Lambda}{2} - \frac{\pi}{3})]^2 + 3p_o p_v m_o m_v$$

where $p_{v,o}$ and $m_{v,o}$ are evaluated for $\Delta = 0$ as follows

$$d_v = K' - r_v, \quad p_v = d_v + \delta, \quad p_o = d_o + \delta,$$
$$d_o = K' - r_o, \quad m_v = d_v - \delta, \quad m_o = d_o - \delta.$$

To solve Eq.(8), we appply substitution $\delta^2 = d_o d_v(1 + x)$ that helps us to obtain two solutions

$$\delta_1^2 = d_o d_v, \; \delta_2^2 \approx d_o d_v[1 + \frac{\Lambda}{\sqrt{3}}(r_v - r_o)] \quad (9)$$

One easily notices that there is a stopband where $d_o d_v < 0$. This is the common stopband for all four wavenumbers $\pm\delta_1$ and $\pm\delta_2$ resulting from the above solutions. The corresponding modes propagating in the area of RMSC are discussed in next Section.

2.3 Uniform RMSC

For the simplicity reason, further considerations will concern uniform RMSC having two *identical* acoustic channels, $W' = W'', \Delta' = \Delta'' = 0, r'_v = r''_v$ etc. This simplifies solutions for eigenvectors associated with eigenvalues δ_n. From (6), one obtains eigenvectors (independent on the sign of δ_n)

$$\pm\delta_1 : \begin{cases} V'_1 = A \\ V'_2 = A \\ V''_1 = A \\ V''_2 = A \end{cases} \quad \pm\delta_2 : \begin{cases} V'_1 = \beta B \\ V'_2 = \beta^* B \\ V''_1 = \beta^* B \\ V''_2 = \beta B \end{cases} \tag{10}$$

where $\beta = 1 + j2/\sqrt{3}$, and A, B are arbitrary constants. Most important however, are SAW amplitudes, or U, which can be evaluated from (6) (A_n- constant)

$$\pm\delta_n : U_0^{'(n)} = \frac{A_n}{K' \pm \delta_n - r_o},\ U_{-1}^{'(n)} = \frac{A_n}{K' \mp \delta_n - r_o}$$
$$U_0^{''(n)} = \frac{A_n}{K' \pm \delta_n - r_o},\ U_{-1}^{''(n)} = \frac{A_n}{K' \mp \delta_n - r_o}. \tag{11}$$

In boundary conditions at both bounds of the RMSC area, we will compare wave components having the same *phase* velocities (see discussion [10] about boundary conditions being the approximations to true ones concerning both particle displacement $\sim U$, and stress $\sim \partial_x U$. The discussion that follows helps to interpret wave-fields associated with $U_n^{(m)}$ which interpretation is independent on the channels thus allowing indices ' and " to be temporarily dropped.

On the strength of Eqs.(1) and (4) we have these ampplitudes involved in the wave-field

$$U_0^{(n)} e^{-j(\delta+K')x} + U_{-1}^{(n)} e^{-j(\delta-K')x}.$$

For $\delta = +\delta_n$ this can be rewritten in form

$$U_0^{(n)} [e^{-j(K'+\delta_n)x} + \gamma_n e^{j(K'-\delta_n)x}]$$

where the first component is a forward propagating wave which decay along its propagation path (provided that the sign of δ is properly chosen), and the second component is a backward propagating wave, which phase velocity is opposite to direction of its decaying. For $\delta = -\delta_n$ we can write

$$U_{-1}^{(1)} [\gamma_n e^{-j(K'-\delta_n)x} + e^{j(K'+\delta_n)x}]$$

which interpretation is similar, but this time the forward wave propagates left, where

$$\gamma_n = (K' + \delta_n - r_o)(K' - \delta_n - r_o)^{-1} \tag{12}$$

Summarizing, the wavefield in the RMSC area, evaluated at the k-th strip center, is
- in the uppper channel (a_p^q - a wave amplitude)

$$\alpha^k a_1^+ e^{-j\delta_1 k\Lambda} + \alpha^{-k}\gamma_1 a_1^+ e^{-j\delta_1 k\Lambda} +$$
$$\alpha^k \gamma_1 a_1^- e^{j\delta_1 k\Lambda} + \alpha^{-k} a_1^- e^{j\delta_1 k\Lambda} +$$
$$\beta\alpha^k a_2^+ e^{-j\delta_2 k\Lambda} + \beta^*\alpha^{-k}\gamma_2 a_2^+ e^{-j\delta_2 k\Lambda} +$$
$$\beta\alpha^k \gamma_2 a_2^- e^{j\delta_2 k\Lambda} + \beta^*\alpha^{-k} a_2^- e^{j\delta_2 k\Lambda},$$

- and in the lower acoustic channel

$$\alpha^k a_1^+ e^{-j\delta_1 k\Lambda} + \alpha^{-k}\gamma_1 a_1^+ e^{-j\delta_1 k\Lambda} +$$
$$\alpha^k \gamma_1 a_1^- e^{j\delta_1 k\Lambda} + \alpha^{-k} a_1^- e^{j\delta_1 k\Lambda} +$$
$$\beta^*\alpha^k a_2^+ e^{-j\delta_2 k\Lambda} + \beta\alpha^{-k}\gamma_2 a_2^+ e^{-j\delta_2 k\Lambda} +$$
$$\beta^*\alpha^k \gamma_2 a_2^- e^{j\delta_2 k\Lambda} + \beta\alpha^{-k} a_2^- e^{j\delta_2 k\Lambda}.$$

RMSC spans over $x \in (-3N\Lambda, 3N\Lambda)$, and there are incident SAWs at both its boundaries (Fig.1). Boundary conditions concerning wave components propagating right and left in both channels, compared separately, are ($Z_p = \exp(-j\delta_p 3N\Lambda)$, $p = 1, 2$)

$$\begin{bmatrix} a_L^+ & b_L^+ & a_R^- & b_R^- \end{bmatrix}^T = \begin{bmatrix} Z_1^{-1} & \gamma_1 Z_1 & \beta Z_2^{-1} & \beta\gamma_2 Z_2 \\ Z_1^{-1} & \gamma_1 Z_1 & \beta^* Z_2^{-1} & \beta^*\gamma_2 Z_2 \\ \gamma_1 Z_1 & Z_1^{-1} & \beta^*\gamma_2 Z_2 & \beta^* Z_2^{-1} \\ \gamma_1 Z_1 & Z_1^{-1} & \beta\gamma_2 Z_2 & \beta Z_2^{-1} \end{bmatrix} \begin{bmatrix} a_1^+ \\ a_1^- \\ a_2^+ \\ a_2^- \end{bmatrix} \tag{13}$$

and similarly for the scattered waves, which solved, yield the scattering matrix of RMSC

$$\begin{bmatrix} a_L^- & b_L^- & a_R^+ & b_R^+ \end{bmatrix}^T = \begin{bmatrix} \frac{\beta^*(\Gamma_2-\Gamma_1)}{\beta-\beta^*} & \frac{\beta\Gamma_1-\beta^*\Gamma_2}{\beta-\beta^*} & \frac{\beta T_1-\beta^* T_2}{\beta-\beta^*} & \frac{\beta^*(T_2-T_1)}{\beta-\beta^*} \\ \frac{\beta\Gamma_2-\beta^*\Gamma_1}{\beta-\beta^*} & \frac{\beta(\Gamma_1-\Gamma_2)}{\beta-\beta^*} & \frac{\beta(T_1-T_2)}{\beta-\beta^*} & \frac{\beta T_2-\beta^* T_1}{\beta-\beta^*} \\ \frac{\beta T_2-\beta^* T_1}{\beta-\beta^*} & \frac{\beta(T_1-T_2)}{\beta-\beta^*} & \frac{\beta(\Gamma_1-\Gamma_2)}{\beta-\beta^*} & \frac{\beta\Gamma_2-\beta^*\Gamma_1}{\beta-\beta^*} \\ \frac{\beta^*(T_2-T_1)}{\beta-\beta^*} & \frac{\beta T_1-\beta^* T_2}{\beta-\beta^*} & \frac{\beta\Gamma_1-\beta^*\Gamma_2}{\beta-\beta^*} & \frac{\beta^*(\Gamma_2-\Gamma_1)}{\beta-\beta^*} \end{bmatrix} \times \begin{bmatrix} a_L^+ & b_L^+ & a_R^- & b_R^- \end{bmatrix}^T \tag{14}$$

$$\Gamma_n = \gamma_n \frac{1 - e^{-j\delta_n 12N\Lambda}}{1 - \gamma_n^2 e^{-j\delta_n 12N\Lambda}},\ T_n = \frac{(1-\gamma_n^2)e^{-j\delta_n 6N\Lambda}}{1 - \gamma_n^2 e^{-j\delta_n 12N\Lambda}}$$

For small $\Delta v/v$, there is $\delta_1 \approx \delta_2$ and thus $\Gamma_1 \approx \Gamma_2 = \Gamma$ and $T_1 \approx T_2 = T$, which are ($k = (k_o + k_v)/2 = \omega/v$)

$$\Gamma = \frac{j\frac{\bar{\kappa}}{\sqrt{f^2-\bar{\kappa}^2}} \tan kL\sqrt{f^2-\bar{\kappa}^2}}{1 + j\frac{f}{\sqrt{f^2-\bar{\kappa}^2}} \tan kL\sqrt{f^2-\bar{\kappa}^2}},\ f = \frac{k-K'}{k}$$

$$T = [\cos kL\sqrt{f^2-\bar{\kappa}^2} + j\frac{f}{\sqrt{f^2-\bar{\kappa}^2}}\sin kL\sqrt{f^2-\bar{\kappa}^2}]^{-1}$$

where $\bar{\kappa} = \kappa^2/3$ and $L = 6N\Lambda$ is the length of RMSC.

It is interesting, that if $a_L^+ = b_L^+$ and $a_R^- = b_R^-$ then $a_L^- = b_L^-$ does not depend on Γ_2, and $a_R^+ = b_R^+$ does not depend on T_2, and thus on δ_2. In such a plane wave operation ($a_L^+ = b_L^+, a_R^- = b_R^-$), only one mode exists in the RMSC area, namely this associated with δ_1. The reflector stopband is $(2/3)\Delta v/v$ relatively, and, in the center of it, the SAW decaying coefficient is relatively high, equal $(2\pi/3)\Delta v/v$ on a wavelength. In normal double-channel operation, these properties are only slightly disturbed by the small difference between δ_1 and δ_2, leading to a small SAW backward reflection into the channel of incident wave. This backward reflection can be neglected for weak piezoelectrics.

3 RMSC RESONATOR

The simplest case of a double-channel RMSC resonator filter is considered. There are two identical IDTs in both acoustic channels, which are characterized by their P-matrices as illustrated in Fig.2 [10].

If we apply uniform transducers having fingers with the same period Λ as in RMSCs, there will be 3 strips per wavelength (with every third strip 'hot' and others grounded). SAW wavenumber is r_o under such system of strips. The IDT having M 'hot' fingers has SAW radiation conductance $G_o \approx 3\omega\epsilon_e(\Delta v/v)W(M^2-M+1)$ which value can be assumed valid in the entire narrow frequency band of interest. It is convenient to include the static capacitance C_T of the transducer to its loading R_n by substitution $1/R_n$ with $1/R_n + j\omega C_T$.

From Fig.2 one obtains (below we replaced $\Gamma\exp(-jr_oL')$ by Γ for shortening notations)

$$I_n = \sqrt{2G_o}(a_n^+ + a_n^-) + G_oV_n,\ n = 1,2$$
$$a_1^{+,-}(1-\Gamma^2) = \sqrt{G_o/2}(V_1\Gamma + V_2)\Gamma$$
$$a_2^{+,-}(1-\Gamma^2) = \sqrt{G_o/2}(V_2\Gamma + V_1)\Gamma$$

which yield

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \frac{G_o}{1-\Gamma^2}\begin{bmatrix} 1+\Gamma^2 & 2\Gamma \\ 2\Gamma & 1+\Gamma^2 \end{bmatrix}\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \quad (15)$$

Four one-port configurations are

• IDTs connected parallelly ($V = V_1 = V_2$, $I = I_1 + I_2$) or in series ($I = I_1 = I_2$, $V = V_1 + V_2$), where

$$I = G_o'(1+\Gamma)(1-\Gamma)^{-1}V, \quad (16)$$

• or connected antiparallelly ($V = V_1 = -V_2$, $I = I_1 - I_2$) or in reversed series ($I = I_1 = -I_2$, $V = V_1 - V_2$),

$$I = G_o'(1-\Gamma)(1+\Gamma)^{-1}V, \quad (17)$$

where $G_o' = 2G_o$ for parallel, and $G_o' = G_o/2$ in series connections.

For natural two-port operation, the input admittance of the resonator filter, which other IDT is loaded by impedance R_2, is (except $j\omega C_T$)

$$Y = \frac{I_1}{V_1} = G_o\frac{1+X_2\Gamma^2}{1-X_2\Gamma^2},\ X_2 = \frac{1-R_2G_o}{1+R_2G_o} \quad (18)$$

and the insertion loss of the filter defined by $IL \sim -\log|S|$ where $S = 2V_2/E$. and R_1, E characterize the source, results from

$$S = \frac{4R_2G_o}{(1+R_1G_o)(1+R_2G_o)}\frac{\Gamma}{1-X_1X_2\Gamma^2} \quad (19)$$

(X_1 is defined similarly to X_2). It is interesting that $S \sim \Gamma$ if either X_1 or X_2 is equal zero. Matching one port of the filter to make $X_2 = 0$ ($X_1 > 0$ assumed), one will measure the filter frequency response that resembles to the RMSC reflection coefficient $|\Gamma(f)|$, with its flat passband, if N is sufficiently large.

There are two distinctive cases of Γ (including the term $\exp(-jr_oL')$) at the center frequency $f = 0$

• $\Gamma^2 = -1$. In this case the distance between RMSCs counts an entire number of wavelengths. There is

$$\begin{array}{ll} Y = 0 & \text{if } R_2 = 0, \\ Y = \infty & \text{if } R_2 = \infty. \end{array} \quad (20)$$

• $\Gamma^2 = 1$ is the other case, where RMSCs are additionally separated by one eighth of the wavelength,

$$\begin{array}{ll} Y = \infty & \text{if } R_2 = 0, \\ Y = 0 & \text{if } R_2 = \infty. \end{array} \quad (21)$$

In both cases $Y = G_o$ if $R_2 = 1/G_o$.

In conclusion, the second case is proper for a passband filter, which insertion loss is minimal at the center frequency, and the first case is proper for a comb or band-rejecting filter (for $X_n > 0$).

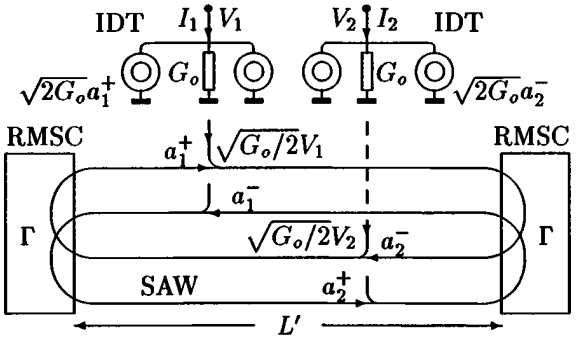


Fig.2. *RMSC symmetric two-port SAW resonator. (Remark: At the second overtone frequency, r/K appearing in Eqs.(3) is twice as much. This reduces $r_o - r_v$, and $\bar{\kappa}$ in Eqs.(14), thus the RMSC relative stopband, by half.)*

4 EXPERIMENT

An experimental resonator has been made on YZ $LiNbO_3$ substrate. A fully periodic structure of strips was applied (the case $\Gamma^2(f=0)=-1$, Sec.3), RMSC included 240 strips each (80 wavelength, however every twelfth pair was short-circuited to prevent electric potentials arrising on the RMSC bus-bars), 192 strips placed within the cavity region included two IDTs with 10 'hot' fingers each (three strips per wavelength; the remaining strips were grounded), similar transducers were placed outside the device for making the RMSC reflection coefficient measurements (Fig.3). The aperture widths of both acoustic channels were 2mm, separated by .1mm. The strip period was 16μm (50% Al metallization). On ST quartz, a similar structure was made, but in spite of 1200 strips included in RMSCs, the number was far too small for obtaining $|\Gamma| \sim 1$.

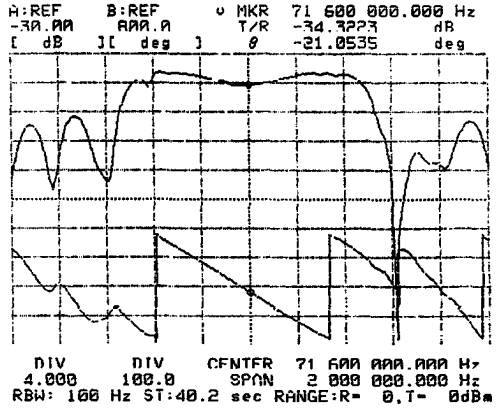


Fig.3. *IL for outside IDTs, a function* $\sim |\Gamma(f)|$.

The saddle in IL$\sim |\Gamma(f)|$ shown in Fig.3 is explained by the strip resistivity (similar effect has been observed in typical Bragg $\Delta v/v$ reflectors [13]). Applying the theory presented in Sec.3, one may check that the current J flowing by the RMSC bus-bars is proportional, in its stopband, to $(\sqrt{\bar{\kappa}^2 - f^2/\bar{\kappa}})a$, where a is the incident SAW amplitude. The power loss on the bus-bars resistance is $\sim |J|^2$, which gets its maximum at the stopband center. This causes the RMSC reflection loss seen in Fig.3. Another experimental results will be presented in a number of self-explaining figures.

Acknowledgments

The author is indebted to John Vig (US Army Res. Lab.) for kind encouragement. Professional technological works by Z.Grzybowski (ITE), L.Dobrzański (ITME), and W.Gregorczyk (PIT) are acknowledged. Work sponsored by KBN under Grant 3 1212 9101.

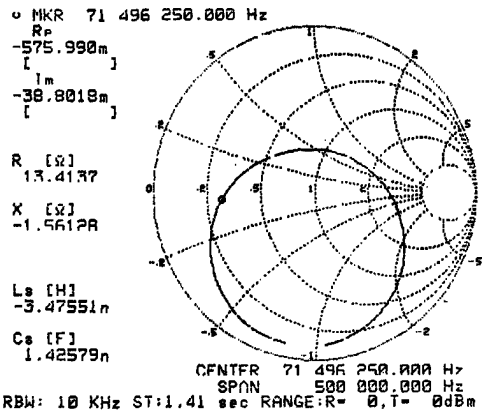


Fig.4 *Smith chart, IDTs in parallel connection*

References

1. W.Pollock, J.Schoefield, R.F.Milsom, R.J.Murray, I.Flinn, Low-loss SAW filter using single-phase IDTs and no external tuning, *1983 IEEE Ultras. Symp. Proc.*, 87–92
2. A.L.Chamorro, Low-loss SAW filter design using reflector multistrip coupler, *1991 IEEE Ultras. Symp. Proc.*, 129–132
3. E.Danicki, Theory and applications of RMSC, *1992 IEEE Ultras. Symp. Proc.*, 195–198
4. S.Doberstein, V.Malyukhov, K.Nikolaenko, V.Rozgonyaev, Using of self-resonance approach for high sensitivity low loss SAW filters, *ibid.*, 151–154
5. E.Danicki, Reversing multistrip coupler, *Ultrasonics*, **31** (1993), 421–424
6. E.Danicki, New configuration of SAW resonator, *Electr. Lett.*, **29** (1993), 1172–1173
7. K.Bløtekjær, K.Ingebrigtsen, H.Skeie, A method for analyzing waves in structures consisting of metal strips on dispersive media, *IEEE Trans.*, **ED-20** (1978), 1133-1146
8. E.Danicki, Synchronous forward to backward surface acoustic wave coupling in reversing multistrip couplers, Photonic Band Gaps and Localization, ed. C.M.Soukoulis, Plenum Press, NY, 1993, 499–508
9. D.P.Morgan, Surface-Wave Devices for Signal Processing, Elsevier, 1991, App.D
10. E.Danicki, Generation and Bragg reflection of surface acoustic waves in nearly periodic system of elastic metal strips on piezoelectric half-space, *J. Acoust. Soc. Am.*, **93** (1993), 116–131
11. P.V.Wright, Analysis and design of low-loss SAW devices with internal reflections using coupling-of-modes theory, *1989 IEEE Ultras. Symp. Proc.*, 141–152
12. A.Erdélyi, W.Magnus, F.Oberhettinger, F.G.Tricomi, Higher Transcendental Functions, McGraw Hill, NY, 1953, v.1, Ch.3
13. E.Danicki, S.Bartnicki, A method for SAW reflection coefficient determination, *1979 IEEE Ultras. Symp. Proc.*, 663–666

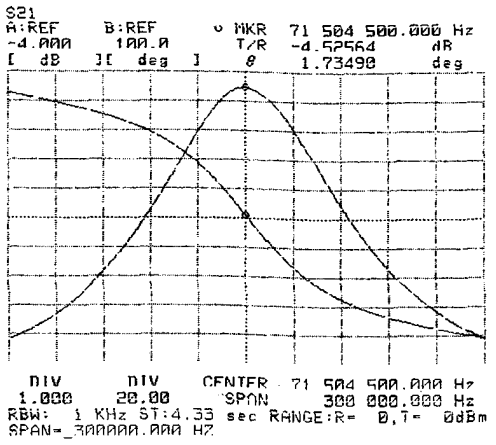


Fig.4. *Narrow-band response, IDTs in parallel.*

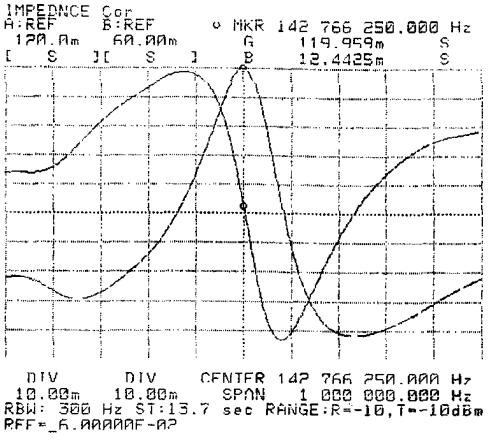


Fig.5. *As above but for second overtone.*

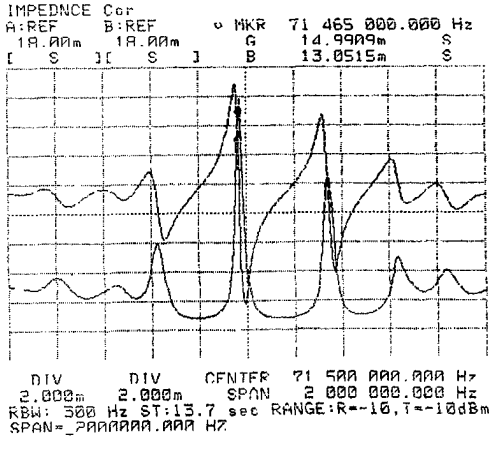


Fig.6. *Wide-band response, IDTs in parallel.*

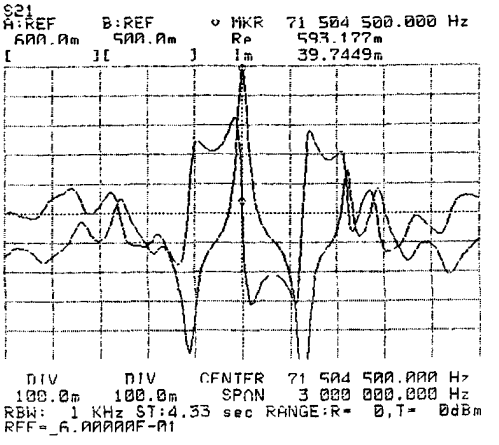


Fig.7. *2-port transadmittance (Re, Im).*

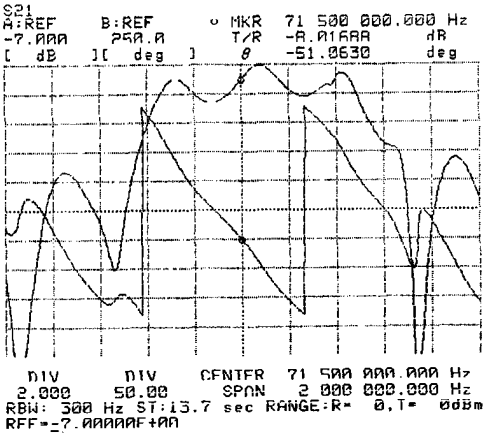


Fig.8. *Transadmittance ~ $|\Gamma|$, output IDT nearly matched.*

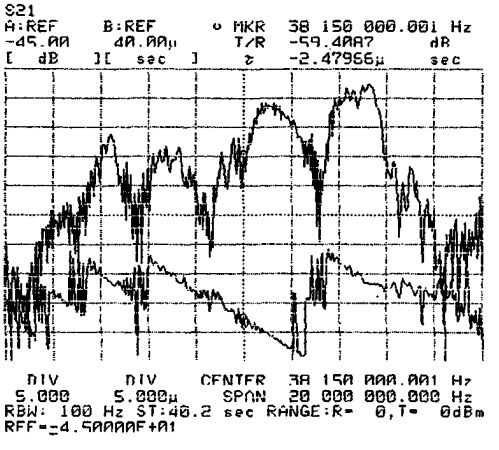


Fig.9. *A preliminary result on RMSC dispersive filter. In spite of severe distortion of amplitude response, the time delay-frequency response is near that expected.*

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Micro-Directional Couplers for Frequency Selection

E. Haruvi-Busnach, M. Zilberstein, N. Croitoru, A. Seidman

Tel-Aviv University, Faculty of Engineering,
Department of Electrical Engineering - Physical Electronics
Tel-Aviv 69978, Israel

Abstract Big adaptors in SAW devices cause cross-talk. This problem was solved employing directional couplers (DC) made of miniaturized micro-interdigital transducers and microacoustic waveguide (WG). They were made by deposition of Al-strips, using microlithographic technology. This enables to avoid dispersion, to increase efficiency, and to integrate several DCs on the same substrate. A coupling of DC of $\sim$ 25dB for the DC was obtained and, by depositing between the WGs, horizontal dashed micro-lines as perturbation, frequency displacement, between transmitting and receiving WGs was obtained.

The RF input circuits employed in electronic warfare are based on the principle of channelizing frequencies. The ideal solution for channelizing frequencies is to employ filters based on surface acoustic wave (SAW) technology. But because of losses and the need of big adaptors, which cause cross-talk, SAW technology has, so far, not been used.[1,2]

In order to solve these problems, directional couplers (DC), able to channelize frequencies, on the basis of frequency selectivity characteristics, combined with periodical perturbations, has been devised.[3]

In this paper, miniaturized SAW directional couplers are described. Micro interdigital transducers and micro-acoustic waveguides were made by employing deposited aluminum strips. Both the width and the spacing between strips are 80-100μm.

The micro-size of the fingers of the interdigital transducers reduces the input capacity of the device to negligible values. The guiding of the acoustic waves avoids dispersion, increases the efficiency of the device, and enables the integration of several devices on a single substrate.

On the basis of micro-waveguides we developed micro-directional couplers. In Fig. 1 the frequency response of the signal in the transmitting and receiving waveguides of DC is shown. As can be seen, the difference between received and transmitted signal at the central frequency is about 25 dB.

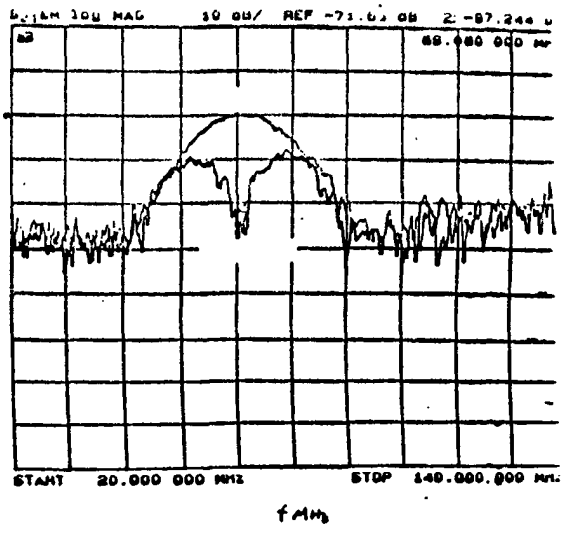


Fig. 1. Signal vs. frequency of transmitting (lower curve) and receiving (upper curve) waveguides.

Assuming that the propagated transmitting (A) and receiving (B) waves in waveguides are:

$$A = D\cos(CZ) \tag{1}$$

$$B = D\sin(CZ) \tag{2}$$

from Eq. (1) and (2) the coupling coefficient (C) is obtained:

$$C = \tan^{-1}(A/B)/L \tag{3}$$

where A is the output signal from the transmitting waveguide, B is the output signal, from the receiving waveguide, D the input signal, and L is the length of the waveguides.

Comparing the characteristics of (1) and (2) to that seen in Fig. (1), a similar behavior was observed.

Inserting perturbations in between the two waveguides of the DC, creates frequency selectivity of the received signal.

0-7803-1945-1/94 $3.00 © 1994 IEEE

The most efficient perturbation which leads to the best, so far, obtained selectivity had the shape of strips, parallel to the waveguides. In Fig. (2a), one type of DC with perturbations is presented. The signal response of both the transmitter and receiver waveguides as a function of frequency is shown in Fig. (2b). A displacement of the central frequency (of about 10 MHz) in the receiving waveguide, as compared to the central frequency in the transmitting waveguide, can be seen.

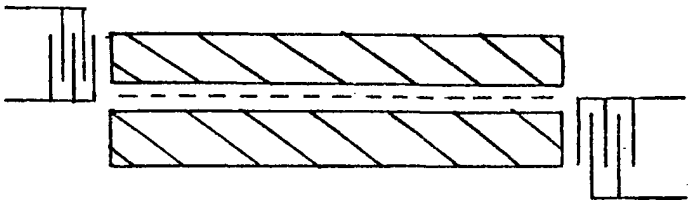


Fig. 2a. Schematic view of a directional coupler. The upper part is transmitting, the lower is receiving, and in between, the dashed microline is used as perturbation.

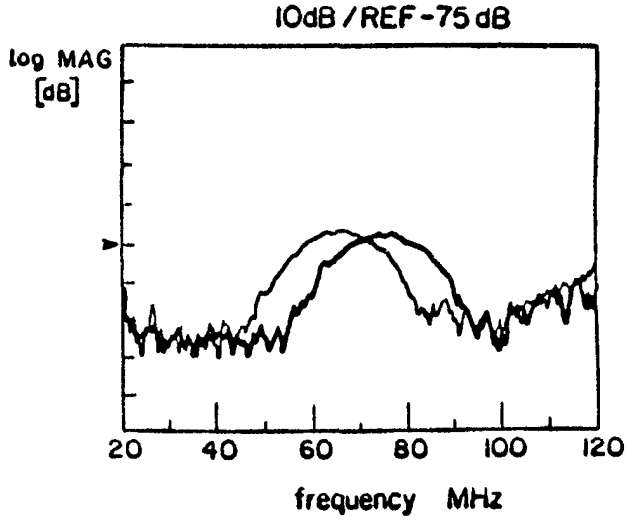


Fig. 2b. Signal response of both the transmitter and receiver waveguides vs. frequency.

The ratio (log MAG, Fig. 2b) between the transmitting to receiving waveguides, is about 1. Vertical and other types of perturbation strips have not given any frequency selectivity of the propagated wave from transmitting to receiving waveguides in the directional coupler. The mechanism, due to perturbations, which affects the transmitted signal in the directional coupler, was studied by modifying the width, length, interval and position in the interval between the waveguides. A very interesting measurement was made for a continuous strip between the waveguides. For small width ($d < 16\mu m$) all the coupled signals in the transmitter are reflected and scattered. This shows that the perturbation strip, for a given critical width, acts as a scattering and reflecting element. For $d > 16\mu m$ the perturbation strip acts as an intermediate waveguide between transmitter and receiver. The studies of the influence of spacing between the strips on the frequency selectivity of the transmitted signal is currently being studied. The frequency selectivity, produced by the perturbation, was studied in order either to devise a frequency channelizer or to improve the existing type of filters. In addition, the realized integrated micro-directional couplers, which enabled to reduce the input capacity, have made the adaptors, which were the source of cross-talk, unnecessary.

References

[1] D.P. Morgan, "Surface-wave devices for signal processing", Amsterdam: Elsevier, 1985, Ch. 8, 183-209, Ch. 10, 281-321.

[2] J.D. Maines and E.G.S. Paige, "Surface acoustic wave components, devices, and application", Proc. of IEE, 120, 1078-1110, 1973.

[3] I. Haruvi-Busnach, M. Zilberstein, G. Golan, A. Seidman, and N. Croitoru, "High transfer-efficiency microsurface acoustic wave directional couplers", J. Appl. Phys., 72(12), 5561-5564, 1992.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

EXTREME PRESSURE-SENSITIVE CUTS FOR SURFACE ACOUSTIC WAVES IN α-QUARTZ

R.M. TAZIEV, E.A.KOLOSOVSKY, and A.S.KOZLOV

The Institute of Semiconductor Physics
Russian Academy of Sciences, Siberian Branch
13, Lavrentyeva, Novosibirsk, 630090, Russia

Abstract

The influence of external static forces on the SAW velocity is computed in a case of circular plates (membranes). Pressure and radial in-plane stress sensitivities of the SAW velocity at any point on the surface of thin membrane as a function of α-quartz crystal anisotropy are presented by sensitivity contour mapping from two independent variables: direction of the SAW propagation and cut angles of α-quartz substrate. Both pressure loading and radial in-plane symmetrical force loading theoretical sensitivity values are compared with experimental ones in the case of circular (membrane) plates. Both experimental and theoretical results are in good agreement.

Introduction

For reduction the sensitivity of SAW devices to mechanical and thermal perturbations several cuts for α-quartz have been proposed [1-3]. One of them so-called "STC-cut" (Φ=0 ; ϑ=41,8 ; Ψ=46,89) have been obtained by B.K.Sinha [4] and another cut so-called "TG-cut" (Φ=0 ; ϑ=-10 ; Ψ=30) have been introduced by G.Theobald et al [5]. These planar isotropic stress compensated cuts exhibit a sensitivity to dynamic thermal effects with one order of magnitude less than that for STX cut of α-quartz[4,5]. But for applications of quartz resonators in the development of pressure sensors one seek the orientations which have a large sensitivity to pressure loading and small sensitivity to in-plane all-round force compression of thin membranes.

The problem of propagation of SAW in pre-strained medium is transformed into separated problems: calculation of the bias and calculation of the wave propagation by using linearized equations. When the solid is subjected to a homogeneous bias, the equations of motion and boundary conditions in terms of effective elastic constants can be solved in a straightforward manner. However, when the biasing state is inhomogeneous, the effective elastic constants are position dependent and a direct solution of the problem is not possible. In the later case, a perturbation procedure is more suitable for computation of small changes in the elastic wave velocity. This method allows to investigate the influence of inhomogeneous stress gradients on the surface wave propagation, when the active length of SAW device expands over region of significant stress gradients.

Moreover, a major problem in such studies is to determine a biasing deformation state of the plate which is strongly influenced by the plate geometry and mounting supports of the device. Last time for explaining some discrepancies between theoretical and experimental sensitivities of SAW devices to acceleration and dynamic temperature effects have been employed the finite-element method (FEM) to obtain the real stress distribution in the part of the plate where SAW device fabricated [6-7].

In this paper we numerically analyze a general expression, relating a fractional velocity shift of SAW to a pressure loading of plate surface and radial all-round and pair diametrical symmetric in-plane force loading of thin circular arbitrarily anisotropic plates. The velocity shifts are calculated by the perturbation method [1]. All necessary independent displacement gradients induced by external forces are derived analytically at any point of the membrane surface. So, we can predict the frequency shift for a SAW resonator at any point on the surface of thin circular plate.

The problem of the pressure sensitivity is considered from a general point of view as location coordinates of SAW resonators on the surface of plate. The cut orientation of substrate and direction of SAW propagation are chosen to achieve an extreme sensitivity. For circular quartz plate results are given by contour mapping of the pressure sensitivity for dual channel differential SAW pressure sensor as a function of the crystal anisotropy. For any SAW propagation direction we, using the analytical relation for relative change of SAW resonator frequency, optimize the location coordinates of two SAW resonators on the surface of membrane to achieve an extreme the sensor pressure sensitivity. For all-round isotropic radial force contour compression of thin anisotropic plate our expressions for stress tensor components are the same as for isotropic plate. In this case the components are not position dependent. However, for pair diametrical force loading of plate the substrate anisotropy plays important role and, as a result, the stress tensor components differ from that for isotropic plate. There are some cuts with both high pressure sensitivity and small all-round in-plane compression sensitivity in quartz (for dual channel SAW sensor).

Also, we present a contour mapping for all-round force compression of thin membrane (disc), a contour mapping of the temperature coefficient of SAW delay (TCD) and SAW electromechanical coefficient contour mapping. Comparing these figures one can choose optimal cuts and SAW propagation directions for SAW pressure sensor applications.

SAW propagation in a pre-strained medium

The propagation of surface waves in presence of quasistatically deforming states is governed by the equations of motion for small dynamic fields superposed on a bias. The perturbation method [8] leads to the first-order relative velocity change

$$\frac{\Delta V}{V_0} = \frac{\int_{\upsilon} u^*_{i,j} \hat{C}_{ijkl}\, u_{k,l}\, d\upsilon}{2\rho\omega_0^2 \int_{\upsilon} u^*_i u_i\, d\upsilon} \qquad (1)$$

$$\hat{C}_{ijkl} = T_{ik}\delta_{jl} + C_{ijklmn}F_{mn} + C_{ijnl}W_{l,n} + C_{inkl}W_{j,n}$$

0-7803-1945-1/94 $3.00 © 1994 IEEE

$\hat{C}_{ijkl}$ are the effective constants that depend on the biasing state, C_{ijkl}, is the second-order elastic, ρ denotes the mass density, Tij, Eij, $W_{i,j}$ are the biasing stress, strain, and displacement gradient components, respectively, C_{ijklmn} are the third-order nonlinear constants, u is the solution of the unperturbed surface wave with the velocity V_0 and the angular frequency ω_0.

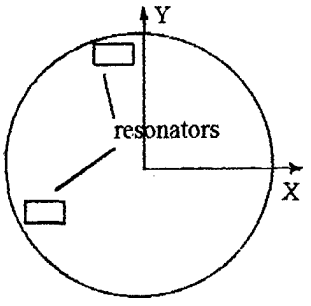


Fig.1. A schematic diagram of quartz membrane with dual channel differential SAW resonator structures with the collinear direction of SAW propagation.

Pressure loading of membrane.

We consider a circular piezoelectric plate of thickness h and in radius R with rigidly fixed boundary which is subjected to uniform pressure on one of its surfaces. Because of the plate is thin, a plane-stress distribution assumption is made. The static strains on the surface of membrane due to application of a uniform pressure P are [10]

$$E_{11} = -\frac{3}{8}\left(\frac{R}{h}\right)^2 \frac{P}{C}\left\{1 - 3\left(\frac{x}{R}\right)^2 - \left(\frac{y}{R}\right)^2\right\}$$
$$E_{22} = -\frac{3}{8}\left(\frac{R}{h}\right)^2 \frac{P}{C}\left\{1 - 3\left(\frac{y}{R}\right)^2 - \left(\frac{x}{R}\right)^2\right\} \qquad (2)$$
$$E_{12} = -\frac{3}{4}\left(\frac{R}{h}\right)^2 \frac{P}{C}\left(\frac{x}{R}\frac{y}{R}\right)$$

here E_{11}, E_{22} and E_{12} are the strain tensor components, C is the coefficient, consisting of the following combination of elastic constants:

$$C=(3(A_{11}+A_{22})+2(A_{12}+2A_{66}))/8.$$

where $A_{ij}=C_{ij}-C_{i3}C_{j3}/C_{33}$, i,j=1,2,6,
Other strain components are obtained from the surface free condition

$$\begin{pmatrix} C_{33} & C_{34} & C_{35} \\ C_{43} & C_{44} & C_{45} \\ C_{53} & C_{54} & C_{55} \end{pmatrix}\begin{pmatrix} E_{33} \\ 2E_{23} \\ 2E_{13} \end{pmatrix} = -\begin{pmatrix} C_{31} & C_{32} & C_{36} \\ C_{41} & C_{42} & C_{46} \\ C_{51} & C_{52} & C_{56} \end{pmatrix}\begin{pmatrix} E_{11} \\ E_{22} \\ 2E_{12} \end{pmatrix}$$

The expression (1) is usually a function of nine displacement gradients. But, the plane-stress assumption and no rigid rotation condition in the center of plate, where the SAW devices is placed, may be reduce the expression (1) to a function only three stress coefficients.

For numerical calculations we used the material constants for a-quartz have been obtained in [11]. For simplicity, we assume a quasi-pointual SAW propagation in small volume of plate part submitted to an homogeneous mechanical bias. Then, using (1) and (2), we can obtain the following relation for a local change of SAW velocity

$$\frac{1}{A}\frac{\Delta V}{V_0} = a_0 + a_1\left(\frac{x}{R}\right)^2 + a_2\left(\frac{y}{R}\right)^2 + a_3\left(\frac{x}{R}\frac{y}{R}\right) \quad (3)$$

from which we can predict the frequency shift of SAW resonators at any point of surface of plate. Here, the coefficients a_j depend on the material constants both explicitly and implicitly through the SAW properties and may be computed for each crystal cut and propagation direction of the SAW, A is the such normalization coefficient that the relation (3) is valid for any membrane dimensions with the same constants a_j and $A=P*(R/h)^2$

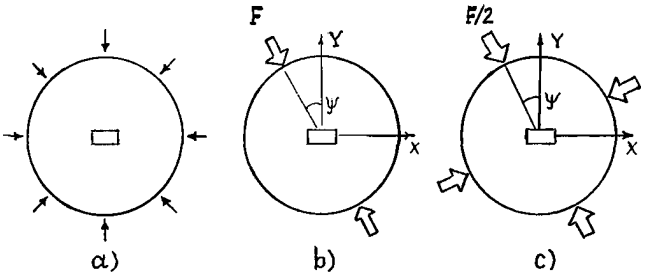


Fig.2. Schematic diagrams of circular plate subject to all-round in-plane compression, a pair of diametrical forces and two orthogonal pair of diametrical forces.

To obtain extreme values of sensitivity of SAW resonator, we must find all local maximum and minimum values of the function (3). It may be easy shown that one of them exists at the center of membrane with coordinates (x=0,y=0). Other extreme coordinates are in the contour line of membrane and they are determined from relation which gives us two roots

$$tg^2\theta + 2(a_1 - a_2)/a_3\, tg\theta - 1 = 0,$$

here, $x=R\cos\theta$, and $y=R\sin\theta$ describes a contour line of membrane. So, we have to choose only two extreme sensitivity coordinates among three ones to obtain the absolute maximum and minimum magnitudes of the function (3) on the surface of membrane. If we consider a case of one SAW resonator sensor, we must choose only the absolute maximum value of (3) to obtain the high sensitivity.

In a case of dual channel differential sensor, we have to choose both the maximum and the minimum values and difference between them allows us to reach a high sensitivity. The dual channel differential scheme of a sensor is more preferable than one channel scheme because of the first allows one to avoid or sufficiently to decrease a temperature influence on a SAW sensor (see Fig.1.).

In-plane compression of thin anisotropic plates.

The stress tensor components for this problem have been derived analytically for isotropic plate in [10]. But for anisotropic case there are some problems which may be resolved only by FEM

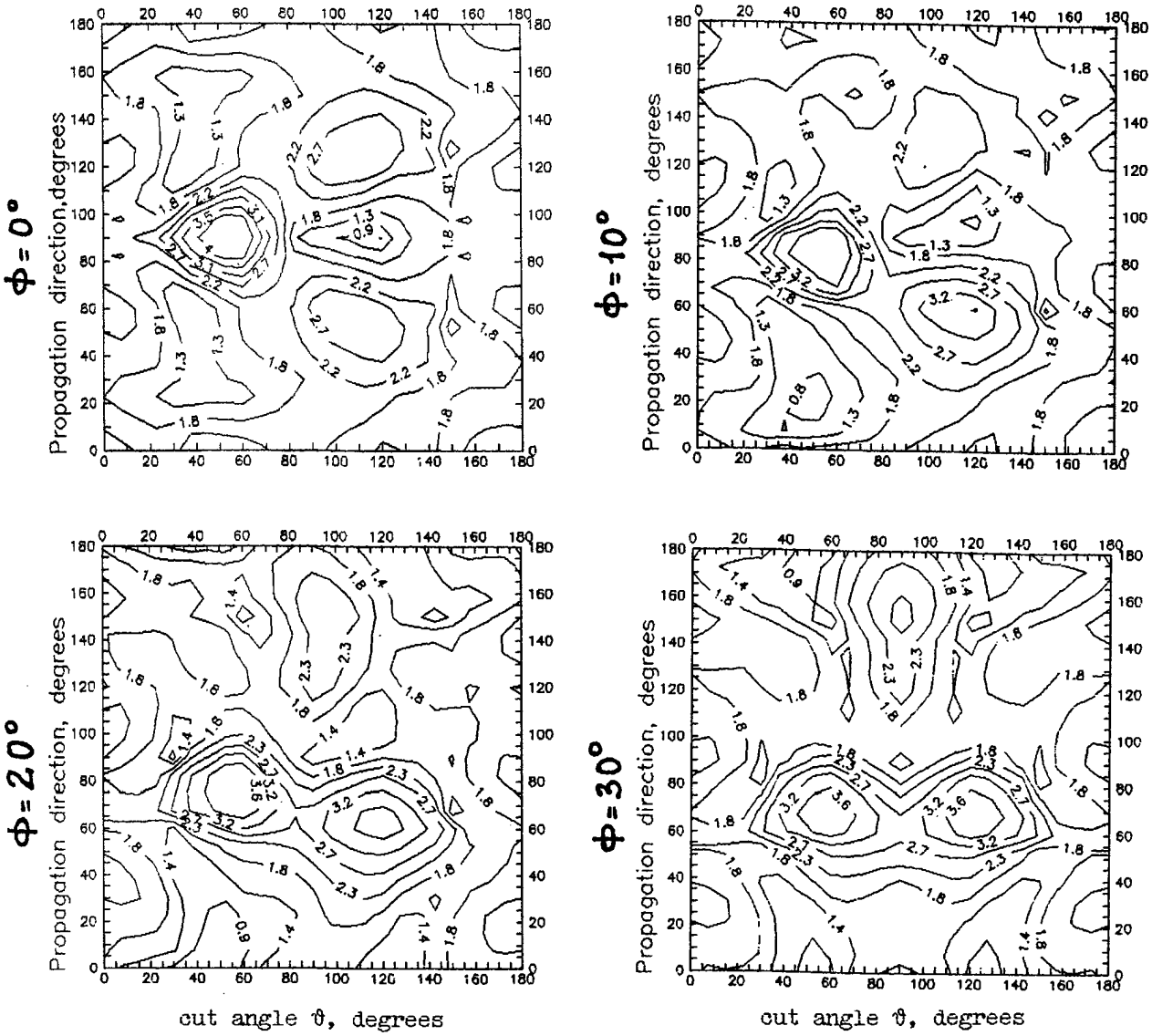


Fig.3a,b,c,d. A contour-mapping of pressure sensitivity of dual channel differential SAW resonator sensor.

elements methods or by complex function theory for two-dimensional system as thin disc. It is known [10] that this problem are reduced to determination of two analytical complex functions F_1 and F_2 which may be written as follows:

$$F_1(z_1)=A_m P_{1m}(z_1),$$

$$F_2(z_2)=B_m P_{2m}(z_2),$$

$$P_{jm}=(\eta_j+Q_j)^m+(\eta_j-Q_j)^m, \qquad j=1,2$$

$$\eta_j=\frac{z_j}{(1-i\mu_j)R}, \qquad Q_j=\left[\eta_j^2\frac{(1+i\mu_j)}{(1-i\mu_j)}\right]^{1/2}$$

where $z_1=x+\mu_1 y$ and $z_2=x+\mu_2 y$, here μ_1 and μ_2 are the complex constants that are the combination of elastic compliances of plate material, R is the radius of disc. A_m and B_m are determined from the following boundary conditions on the contour line of disc:

$$2\mathrm{Re}[F_1(z_1)+F_2(z_2)]=-\int Y_n\,dt=f_1(t)$$

$$2\mathrm{Re}[\mu_1 F_1(z_1)+\mu_2 F_2(z_2)]=\int X_n\,dt=f_2(t)$$

where X_n, Y_n are the coordinate projection values of external force loading the contour line of disc, t is the parameter of boundary line of disc.

The stress tensor components are expressed in terms of derivatives of the functions F_1 and F_2 :

$$T_{11}=2\mathrm{Re}[\mu_1^2 F_1'(z_1)+\mu_2^2 F_2'(z_2)]$$

$$T_{22}=2\mathrm{Re}[F_1'(z_1)+F_2'(z_2)]$$

$$T_{12}=-2\mathrm{Re}[\mu_1 F_1'(z_1)+\mu_2 F_2'(z_2)]$$

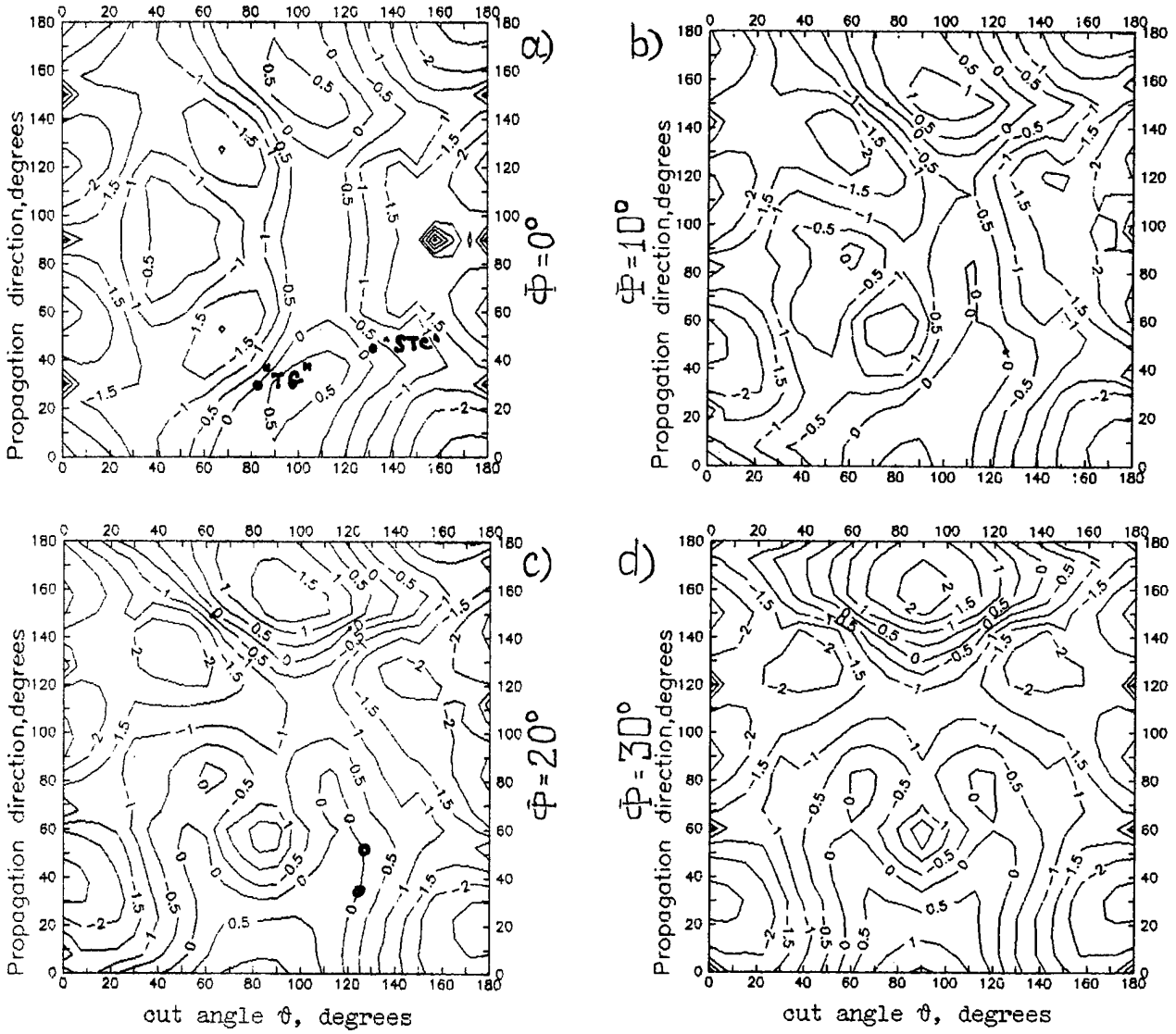


Fig.4a,b,c,d. A contour mapping of SAW sensitivity to all-round in-plane isotropic force compression of circular plates.

For all-round in-plane compression of disc(see Fig.4 a) the stress tensor components T_{ij} are identical with those for isotropic plate:

$T_{11} = T_{22} = F/(h\,d)$, $T_{12} = 0$.

This consideration would be incomplete, if we don't mention following circumstance. The problem of anisotropic plate compression can be resolved in closed form by complex function method in a general case, if one will use the well-known Shwartz's formula, which restores an analytical function inside of unit circle by it's real part on the boundary contour line[.13.]. The resulting expressions for functions F_1 and F_2 may be written as:

$$F_1(z_1(\xi)) = F_1(\xi) = \frac{1}{4\pi(\mu_2-\mu_1)} \int_0^{2\pi} [\mu_2 f_1(t) - f_2(t)] \frac{e^{it}+\xi}{e^{it}-\xi} dt$$

$$F_2(z_2(\xi)) = F_2(\xi) = \frac{-1}{4\pi(\mu_2-\mu_1)} \int_0^{2\pi} [\mu_1 f_1(t) - f_2(t)] \frac{e^{it}+\xi}{e^{it}-\xi} dt$$

where t is the parameter for boundary line which is the same for z, z_1 and z_2 regions; $z_j(\xi)$ is the conform transformation, that transforms the inside region $|\xi| < 1$ into the elliptical region $z_j = x + \mu_j y$. It should be noted that the all elastic anisotropy of plate disk contains only in two values: μ_1 and μ_2.

Computational results and discussion

Fig.3 gives the contour mapping of extreme pressure sensitivity (3) for SAW dual channel differential sensor as a function of Euler's cut angles of quartz plate Φ and ϑ and direction of SAW propagation angle Ψ. For any concrete values of Euler's angles Φ, ϑ and Ψ the positions of two resonators are chosen to achieve an extreme sensitivity as described above such way. So, step by step we have completely investigated all cuts and propagation directions of SAW

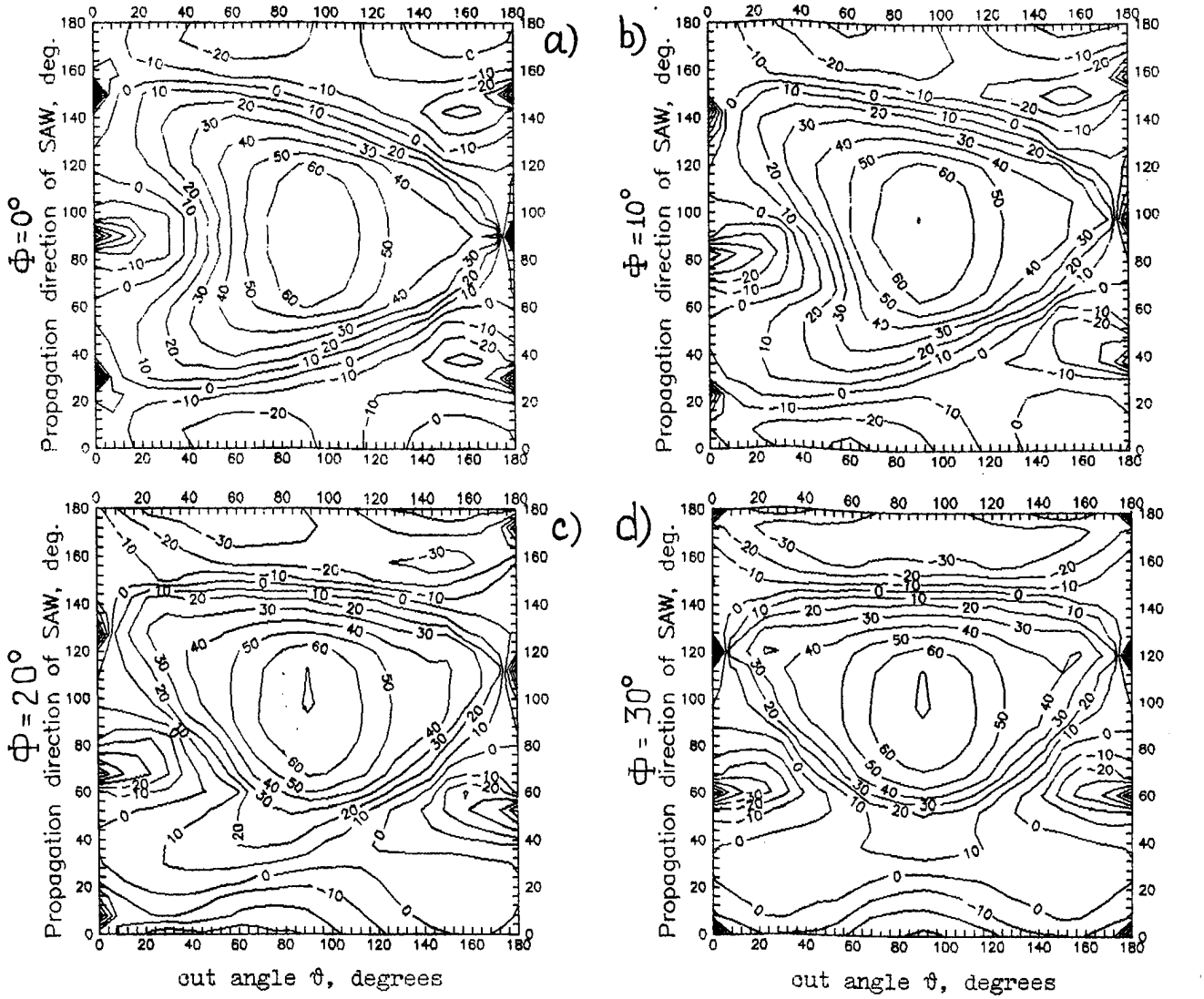


Fig.5a,b,c,d. A contour mapping of SAW temperature coefficient of delay in α-quartz.

in quartz with discrete angle step of 10 degree. It appears that this value of angle step is sufficient because the sensitivity value is a smooth of Euler's angles Φ, ϑ and Ψ. For numerical calculations we used the material constants for α-quartz in [.11.] and the following parameters: h=1mm, R=10mm, P=100 kPa. To obtain the real value of sensitivity δV/V, we must multiply it to 1.E-04. As one can see, there are some cuts of quartz which have nearly two times higher pressure sensitivity than well-known ST,X α-quartz with Euler's angles (Φ=0; ϑ=132,75; Ψ=0). One of them is the cut with angles Φ=0, ϑ=140 with direction of SAW propagation of 40 degrees with respect to X axis of quartz. Previously, for certain cuts and SAW propagation directions in quartz the pressure-sensitivity contour mapping have been presented in [.12.].

Fig.4,5,6 present the contour mapping of all-round force compression, the temperature coefficient of SAW delay (TCD)and SAW electromechanical coefficient contour mapping. Comparing these figures one can choose optimal cuts and SAW propagation directions for SAW pressure sensor applications. For example, using the interception of contour line with zero value in Fig.4 and Fig.5 we may select the cuts that possesses both temperature and all-round force compression insensitivity, simultaneously. Authors [4,5] by this way have selected well-known "TG-cut" and "STC-cut". Fig.4 shows additionally two other similar points with Euler's angles Φ=20 , Ψ=127 , ϑ=55 and Φ=20 , Ψ=125 ,ϑ=36 . Using Fig.6 for contour mapping of SAW electromechanical coefficient, we find that selected above points are suitable for excitation of SAW. These values are approximately 0.04% and 0.05%, respectively. Finally, using Fig.3 for pressure contour mapping we see that these cuts have pressure sensitivity values are 3.6 and 2.3, respectively, which are more than that for "TG-cut" and "STC-cut" approximately at 1.5 times. These sensitivities are realized with the following SAW resonator coordinates: one of them are (0,R), and another are (R,0) for Euler's

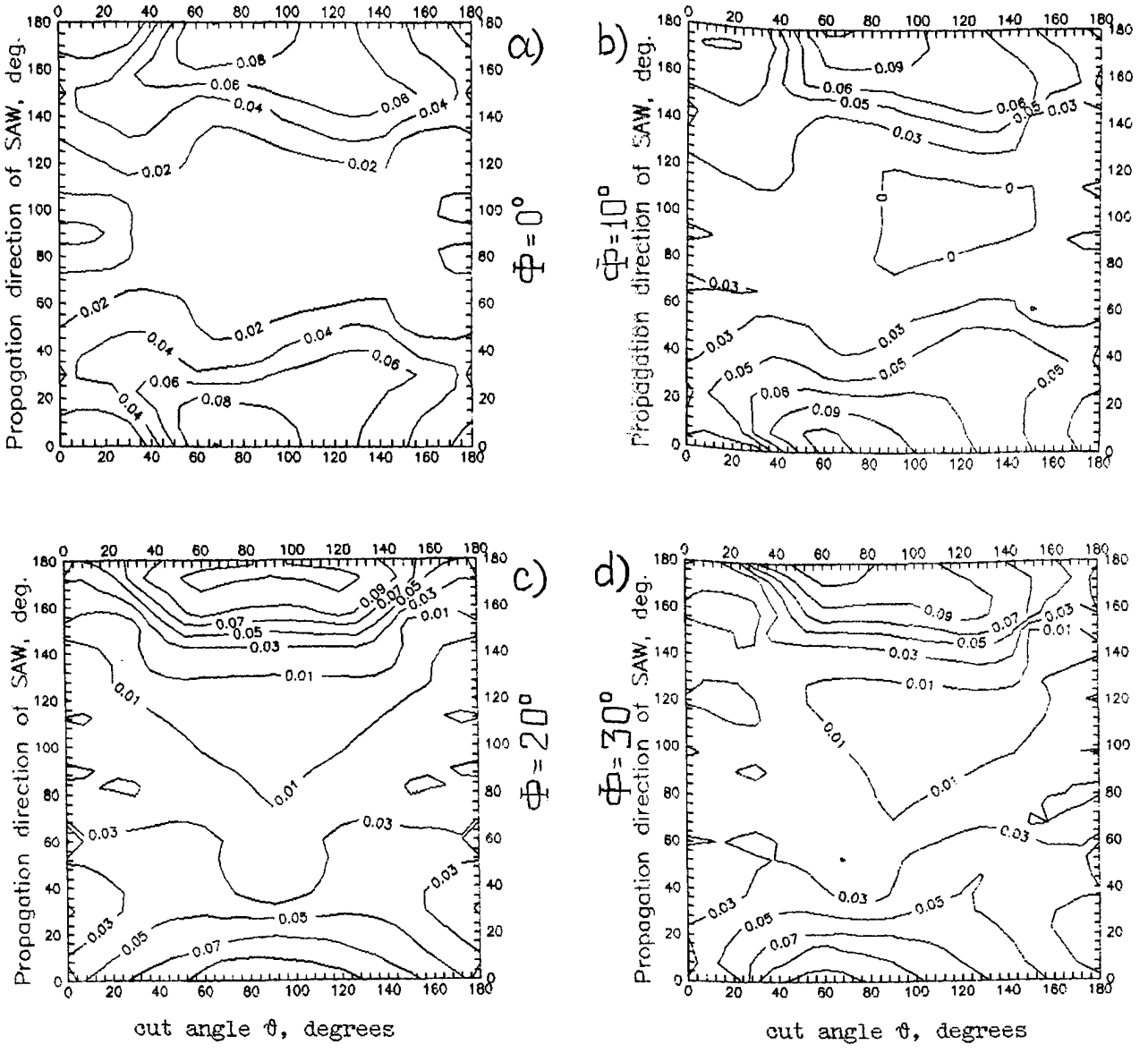


Fig.6a,b,c,d. A contour mapping of SAW electromechanical coefficient in α-quartz

angles Φ=20 , Ψ=127 , ϑ=55 . For angles Φ=20 ,Ψ=125 ,ϑ=36 SAW resonator coordinates are: (0.4R,-0.9R) and (0.9R,0.4R). It should be noted that in both cases SAW resonators are placed at the boundary contour line of membranes.

Since the relative change of SAW velocity at any point on the membrane surface due to pressure loading may be written also as:

$$\delta V/V = \alpha_{11}T_{11}(x,y)+\alpha_{22}T_{22}(x,y)+\alpha_{12}T_{12}(x,y)$$

For all-round in-plane radial force compression of plate with constant magnitude the similar relation is not coordinate dependent and gives

$$\delta V/V = \alpha_{11}T_{11}+\alpha_{22}T_{22} ,$$

here, α_{ij} are the coefficients, depending on material constants of plate. Unfortunately, comparing both above relations, we may expect that there no cuts and SAW propagation directions where would be both high pressure and extreme low all-round isotropic force compression sensitivity of SAW resonator. Sooner, there are cuts and propagation directions with both low pressure loading and low isotropic radial force compression sensitivity of SAW resonator. Really, our computation results presented in Fig.4 a,b,c,d show that if SAW resonator placed at the center of disc insensitive to pressure loading then also it as a rule insensitive to all-round isotropic force compression of disc. The points denoted by symbols "TG" and "STC" are the orientations of well-known "TG-cut" and "STC-cut" of quartz which widely utilizing for SAW resonator applications with extremely low sensitivity to pressure loading and all-round in-plane contour force loading of disc.

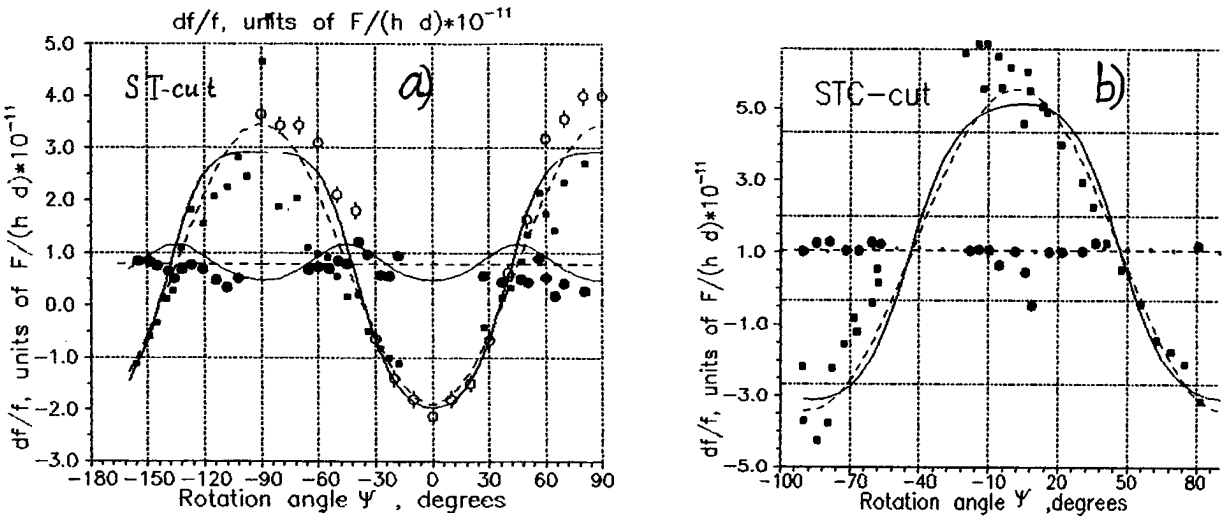


Fig.7. Force-frequency coefficient of SAW device placed at the center of ST-cut and STC-cut discs as a function of azimuthal angle ψ.
O- Sinha's experimental data [1`4],
■,● - Bigler's experimental data [7] for single pair and two orthogonal pair of diametrical force compession, respectively.

Fig.7 a,b show force-frequency coefficients of SAW device on ST-cut and STC-cut as a function of azimuthal angle of force loading. Curves 1 and 2 correspond for a pair diametrical force compression with the magnitude F and two pairs of orthogonal force compression with the magnitude F/2, respectively. Solid lines correspond to anisotropic model of stress tensor Tij . Dashed lines correspond to isotropic model of stress tensor Tij. One can see that there is an excellent agreement between results with these two models for SAW resonator placed at the center of disc. Moreover, there are a good agreement between experimental measurements by authors [7,14] and our computational results.

Conclusion

A simple semi-analytical model is presented to predict a pressure and radial in-plane symmetric force compression sensitivity of SAW resonator located at any point of surface of quartz circular plates. This allows one to obtain an extreme pressure sensitivity with corresponding location coordinates of two SAW resonators on the surface of membrane for dual channel differential sensor. Theoretical results are verified by experimental data obtained for different cuts of α-quartz. It is found that there are some cuts in α-quartz which has high pressure sensitivity and low sensitivity to all-round force compression of disc and zero value of temperature coefficient of SAW delay, simultaneously.

Acknowledgments.

The authors would like to thank Dr. Igor Yakovkin for his helpful discussions and suggestions.

References

[1] B.K.Sinha, W.J.Tanski, T.Lukaszek, and A.Ballato," Influence of biasing stresses on the propagation of surface waves," J. Appl. Phys.,vol. 57, pp. 767-776, 1985.

[2] G.Theobald, D.Hauden," Theoretical analysis of dynamic thermal effects in SAW devices," in Proc.38th Ann.Freq. Contr. Symp. IEEE, cat.84CH2068-8, 1984, pp. 294-299.

[3] S.Ballandras, E.Bigler," Experimental study of SAW quartz cuts with low sensitivity to dynamical thermal effects," Electron. lett., vol.27, pp. 1343-1345, 1991.

[4] B.K.Sinha," A stress and temperature compensated orientation and propagation direction for surface acoustic wave devices," IEEE Trans.Ultrason. Ferroelect. Freq.Contr. vol. UFFC-34, pp. 64-74, 1987.

[5] E.Bigler,G.Theobald, and D.Hauden," SAW quartz cuts with low stress and temperature sensitivity," Elect.lett., vol.23, pp.514-516, 1987.

[6] B.K.Sinha and S.Locke," Acceleration and vibration sensitivity of SAW devices," IEEE Trans.Ultrason. Ferroelect. Freq. Contr., vol. UFFC-34,pp. 29-38, 1987.

[7] S.Ballandras and E.Bigler," New results for the characterization of SAW sensitivity to mechanical stresses using Finite element analysis," in 1991 Ultrason. Symp. Proc., 1991,vol.1, pp.425-429.

[8] H.F.Tiersten and B.K.Sinha," A perturbation analysis of the attenuation and dispersion of surface waves," J.Appl.Phys., vol. 49, pp. 87-95, 1978.

[9] H.F.Tiersten and J.C.Baumhauer," Nonlinear electroelastic equations for small fields superposed on a bias," J. Acoust. Soc. Amer., vol. 54, pp. 1017-1034, 1973.

[10] S.G.Lehnitsky," A solution of a plane-stress problem for anisotropic elastic ellipsoidal plate," Reports of Academy of Sciences of the USSR, vol. 15, pp. 527-530, 1937.

[11] R.Bechmann,A.D.Ballato,and T.J.Lukaszek," Higher order temperature coefficients of the elastic stiffness and compliances of alpha-quartz," Proc. IRE, vol.50, pp.1812-1822, 1962.

[12] R.M.Taziev,E.A.Kolosovsky and A.S.Kozlov," Deformation-sensitive cuts for surface acoustic waves in a-quartz", in 1993 IEEE Inter. Freq. Contr. Symp. Proc., 1993, pp.660-664.

[13] G.N.Savin," Some problems in theory of anisotropic elastic media", Reports of Academy of Sciences of the USSR, vol.23,N3, pp.217-220,1939.

[14] B.K.Sinha and W.J.Tanski, "Influence of biasing stresses on the propagation of surface waves", J. Appl. Phys. vol.57, N3, pp.767-776, 1985.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DESIGNING SMALLER SAW OSCILLATORS FOR LOW VIBRATION SENSITIVITY

Donald Andres and Thomas E. Parker

Raytheon Company
Research Division
131 Spring Street
Lexington, Massachusetts 02173
USA

ABSTRACT

All quartz package (AQP) Surface Acoustic Wave (SAW) hybrid circuit oscillators with low vibration sensitivity, typically 3 to $4x10^{-10}/g$, have been routinely achieved, provided that the oscillator package is sufficiently rigid to minimize bending stresses caused by vibration loading. This has been accomplished by bonding a suitably thick ceramic stiffener to the base of the oscillator housing. However, because of size constraints, some applications cannot accommodate a sufficiently thick stiffener. One approach toward overcoming this potential limitation on performance is to use finite element analysis to determine the stresses in the base of an oscillator package that would otherwise have an insufficiently thick stiffener, and then arrange to mount the AQP SAW device in a region of minimal stress. This approach was verified experimentally by comparing the vibration sensitivity of the same oscillator with the AQP SAW device mounted in both high- and low-stress regions. The result of the experiment was a factor of two reduction in the oscillator's vibration sensitivity. An alternative approach is to apply mass loading directly to the AQP SAW device. This is accomplished by bonding two small weights, approximately 60 mg each, to the backside of the AQP SAW substrate. Under vibration loading, these weights alter the stress gradients in the active acoustic area of the SAW substrate which reduces $\overline{\Gamma}$, the magnitude of the vibration sensitivity vector. The same procedure was unsuccessfully applied to AQP SAW oscillators that had sufficiently thick stiffeners. Thus, mass loading is only effective in decreasing $\overline{\Gamma}$ for an oscillator with an insufficiently thick stiffener.

I. INTRODUCTION

Low vibration sensitivity in AQP SAW oscillators has previously been demonstrated by the use of an optimized frit geometry for the vibration sensitivity normal to the AQP SAW substrate, γ_1, and for the two orthogonal directions in the plane of the substrate, γ_2 and γ_3 [1]. This has been confirmed analytically for the case of acceleration normal to the SAW substrate [2], and also for accelerations applied in the plane of the SAW substrate [3]. These analytical approaches rely solely on the use of optimized frit geometries to minimize γ_1. But the use of optimized frit geometries is severely constrained by recent trends toward larger acoustic apertures and smaller outside AQP dimensions. Also, the optimized frit geometry is effective only when the AQP SAW device is mounted in a housing that is sufficiently rigid. This has been accomplished by attaching a thick ceramic stiffener to the oscillator housing [4]. The use of ceramic stiffeners is a practical solution, effectively minimizing deflections of the oscillator package. But in some applications, the use of thick ceramic stiffeners is unacceptable. This work addresses the effects of package stresses upon vibration sensitivity and how to minimize the resulting vibration sensitivity even when the use of sufficiently thick package stiffeners and optimized frit geometries are precluded for other reasons.

The most effective method for improving oscillator vibration sensitivity is to minimize the stress levels in the oscillator package beneath the AQP SAW device. This can be accomplished with careful design of the oscillator package. Also, with some *a priori* knowledge of the stress gradients in the oscillator package, the AQP SAW device can be located in a region of minimal stress. Vibration isolation is another technique frequently exploited to achieve reduced vibration sensitivity. A typical oscillator mounting scheme is shown in Fig. 1. The oscillator assembly consists of an AQP SAW oscillator which is bonded to a ceramic stiffener, which is also bonded to an adapter plate. A thinner alumina piece is bonded to the cover of the oscillator housing to eliminate the package's cover resonance. The complete

0-7803-1945-1/94 $3.00 © 1994 IEEE

assembly, or *stack assembly,* is mounted using silicone bushings, which provide vibration isolation at frequencies above 500 Hz. Due to the nature of the *stack assembly* the stress gradients vary widely through the assembly. Careful consideration of the location and type of mounting hardware can have a dramatic affect on the resultant package stresses. Material thicknesses and geometric stress concentrators will also affect the stresses in the AQP SAW substrate. These details are best addressed through the use of finite element analysis.

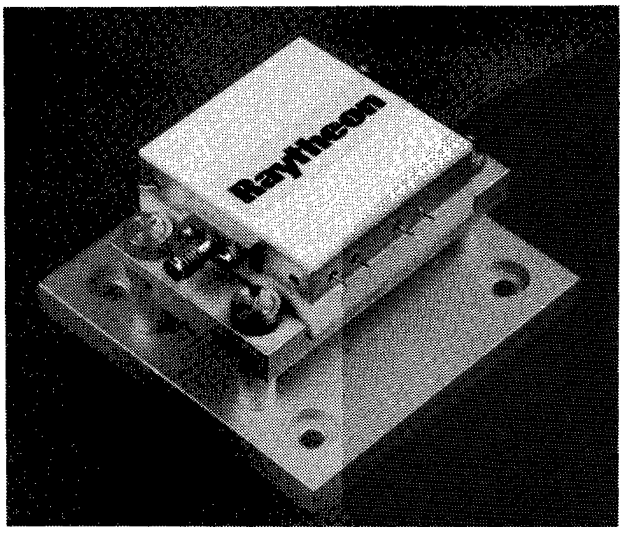


Figure 1. Photograph of the complete stack assembly mounted on vibration isolation bushings.

An alternative method to reduce the vibration sensitivity would be to compensate for the stresses which are generated by the flexure of the oscillator package and transferred through the material used to mount the AQP SAW device. The AQP SAW device is attached to the oscillator package with a low modulus silicone adhesive, as shown in Fig. 2. This compliant material minimizes stresses due to thermal mismatch of the AQP quartz cover and Kovar package. But the soft material is constrained between two stiff substrates, so that the silicone is effectively stiffened by the constraining substrates. The result is that the package bending stresses are still fairly well coupled to the SAW substrate. To compensate for these stresses we have attached weights to the backside of the AQP SAW substrate, which are positioned above the active area of the SAW device. The location of these weights can change the vibration sensitivity and phase angle between the FM sideband and the output of the accelerometer. We have demonstrated that this method can be used to lower the vibration sensitivity magnitude of an insufficiently stiffened AQP SAW oscillator.

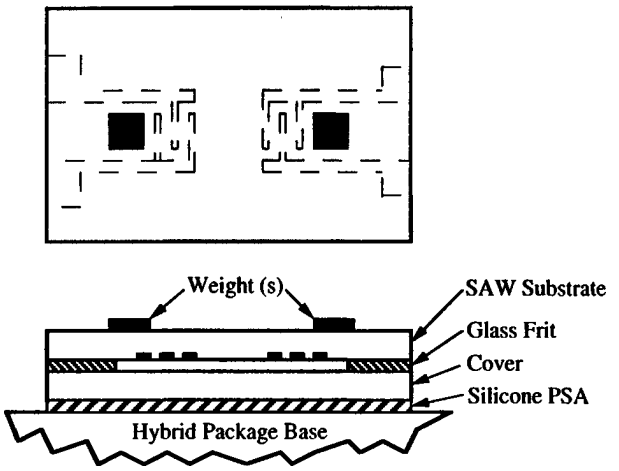


Figure 2. Schematic diagram of the AQP SAW device construction and mounting details. Also shown is the approximate position of the weights.

II. OSCILLATOR PACKAGE STRESSES

The problem of package stresses was first addressed in an existing design, which was insufficiently stiffened because system requirements would allow only enough room for a 0.25 inch thick ceramic stiffener. A stiffener approximately twice this thickness would have been necessary to sufficiently stiffen the base of the oscillator package. The stresses in the base of the oscillator package were determined using finite element analysis. The finite element code used to perform the stress analysis was COSMOS/M [5]. The symmetrical nature of the oscillator *stack assembly* was considered during the modeling process. The result was that only one quarter of the stack assembly was modeled to reduce the solution time of the analysis. Another assumption made during model construction was the elimination of the thermoplastic adhesive bond from the *stack assembly* components. Instead, the nodes of the elements representing the individual components, i.e., the oscillator housing, alumina stiffener, and adapter plate, were directly merged at their respective interfaces. This assumption was justified since the components are very rigid and the actual bonding layer thickness is small. Other assumptions include the simplification of geometric features in the adapter plate and isolation bushings. The loading conditions consisted of a 3g static load applied in the direction normal to the plane of the AQP SAW device. The von Mises stresses in the base of the oscillator housing directly below the AQP SAW device are plotted in Fig. 3. The curves represent two

different AQP SAW device locations relative to the isolation mounts, as shown in Fig. 4. The von Mises stress at the nodes along the interface of the hybrid housing and AQP SAW device are plotted as a function of distance along the propagation direction (X-axis) of the AQP SAW device. The data origin for the X-axis is located at the center of the resonant cavity and continues to the edge of the AQP SAW device. The average stress level for the non-rotated case is roughly twice that of the rotated case. Depending on precisely what percentage of these stress levels are actually coupled to the AQP SAW device, the rotated case should have a lower γ_1 sensitivity than the non-rotated case.

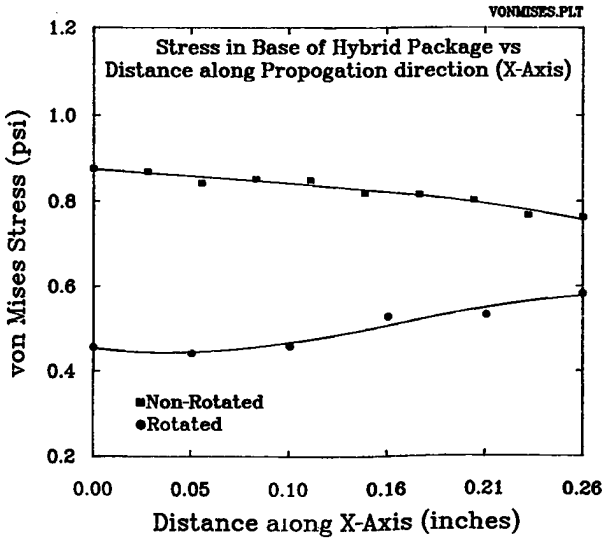


Figure 3. Package stresses from finite element analysis as a function of position beneath the AQP SAW device.

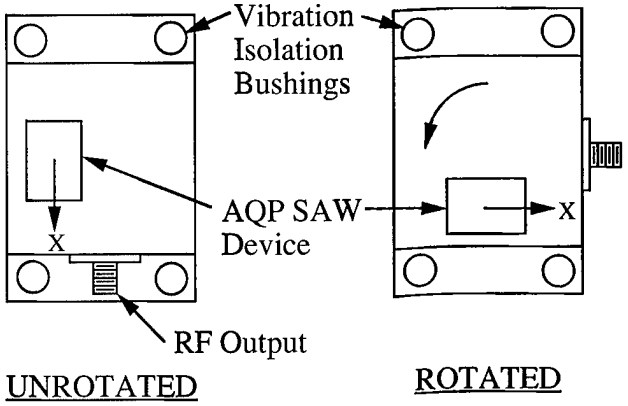


Figure 4. The position of the AQP SAW device relative to the bushing mounts for the unrotated and rotated conditions.

The analytical results were verified by measuring the vibration sensitivity of the same SAW oscillator with the AQP SAW device in both locations. An automated measurement system was used to determine the vibration sensitivity [4]. The discriminator technique is used by the system, which has an adequate measurement system floor for our purposes. A 0.25 peak g load was applied to the oscillator by the shaker table. The AQP SAW device was repositioned by rotating the entire hybrid circuit oscillator package 90 degrees relative to the stack assembly (see Fig. 4). This was done, rather than repositioning the AQP SAW device, to ensure that the results from the test would not be influenced by the remounting of the AQP SAW device. The magnitudes of the vibration sensitivity vector, $\overline{\Gamma}$, for the non-rotated and rotated cases are plotted in Fig. 5. The reduction in $\overline{\Gamma}$ is clearly evident at the resonance peak (~300 Hz) and below. Above 500 Hz, the vibration isolation takes effect and the two $\overline{\Gamma}$'s are indistinguishable from each other due to the measurement system floor. Unfortunately, the next higher level of assembly could not accommodate the rotated configuration because of mechanical interferences. Also, the AQP SAW device could not be repositioned within the oscillator housing without a complete redesign of the oscillator circuit.

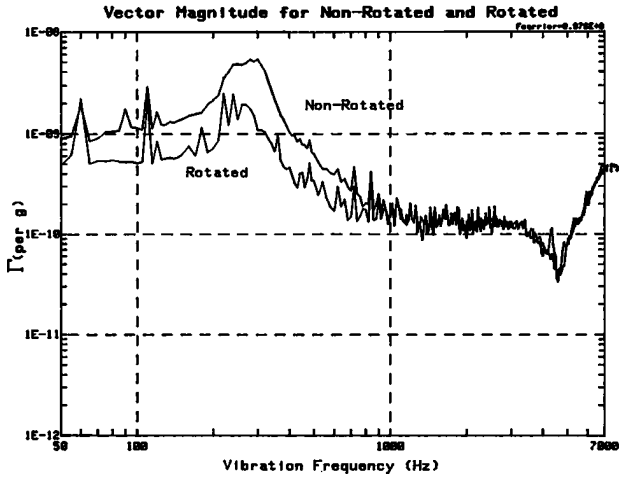


Figure 5. The measured vector magnitude as a function of vibration frequency for the non-rotated and rotated cases.

III. MASS LOADING

Another method of reducing the vibration sensitivity of an insufficiently stiffened oscillator is to apply mass loading directly to the backside of the AQP SAW substrate. This is easily accomplished, since the AQP SAW device is mounted cover side down, as noted in

Fig. 2. Another advantage is that since the weights are attached after the AQP SAW device is permanently mounted into the oscillator housing, this approach compensates for the differences in $\overline{\Gamma}$ which are introduced by variations in the mounting material. Experimental data taken on several sufficiently stiffened oscillators have shown substantial variation in vibration sensitivity. Subsequent remounting of the AQP SAW device in the same oscillator circuit clearly demonstrates that the variation in $\overline{\Gamma}$ is due to inconsistencies in the uniformity of the mounting support. Our experience with mounting materials indicates that these inconsistencies are too small to be practically controlled during the assembly process.

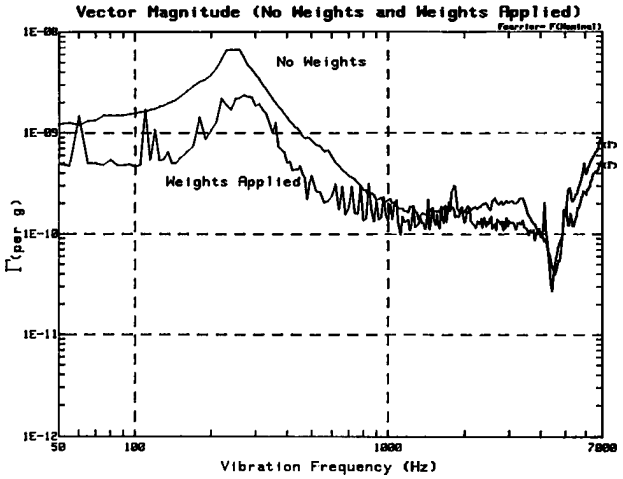


Figure 6. Magnitude of the vibration sensitivity vector as a function of vibration frequency for the same oscillator without weights and with weights applied.

The vibration sensitivity magnitude was measured for two oscillators without any mass loading. Next, the weights were temporarily attached to the SAW substrate with a double-sided pressure sensitive adhesive tape. This was done without removing the SAW oscillator from the shaker table. The weights were fabricated from a martensitic stainless steel, which closely matched the thermal expansion coefficient of the quartz substrate. This material was chosen to minimize the effect of the weights on the frequency vs temperature characteristics of the oscillator. The temperature characteristics were measured before and after the weights were attached. The weights had a negligible affect on the frequency vs temperature characteristics of both oscillators. Next, the normal vibration sensitivity, γ_1, was measured and the weights were repositioned until a minimum value of γ_1 was obtained. The weights were then permanently attached to the SAW substrate with an RTV adhesive. After the adhesive cured, $\overline{\Gamma}$ was measured on both devices. The magnitudes of $\overline{\Gamma}$ for the weighted and unweighted conditions of one oscillator are shown in Fig. 6. The weighted $\overline{\Gamma}$ magnitude of the second oscillator was approximately the same as the result plotted in Fig. 6, but did not show the same net reduction since the unweighted $\overline{\Gamma}$ was lower than the first device's.

The effect of weight position on γ_1 is plotted in Fig. 7a for an insufficiently stiffened oscillator at a vibration frequency of 150 Hz. The weights were repositioned, as illustrated in Fig.7b, and the γ_1 sensitivity was measured at each position. The effect on γ_1 is greatest when the weights are positioned near the center of the acoustic cavity.

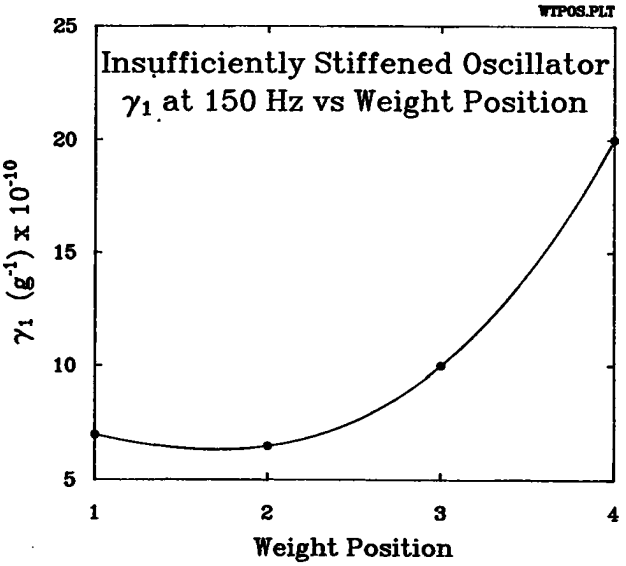


Figure 7a. γ_1 as a function of weight position for an insufficiently stiffened oscillator at a vibration frequency of 150 Hz.

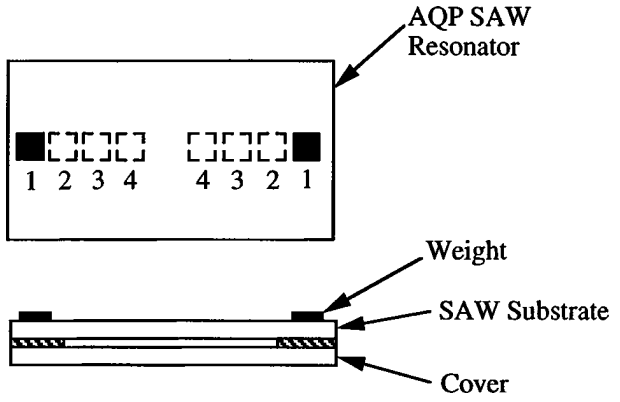


Figure 7b. The relative position of the weights for the data presented in Figure 7a.

This result is expected since it is reasonable to assume that the weights would have their maximum effect near the center of the acoustic cavity. As the weights were repositioned further from the center of the acoustic cavity (position no. 4), a decrease in the vibration sensitivity was measured. A minimum value of γ_1 is achieved when the weights are near position no. 2.

Mass loading was also applied to a SAW oscillator that had been sufficiently stiffened to eliminate package stresses. This was done in order to determine if mass loading would reduce the $\overline{\Gamma}$ of an oscillator that had been sufficiently stiffened. Figure 8 shows a reduction of γ_1 when the weights are applied. But the vector magnitude, $\overline{\Gamma}$ was unchanged for both the unweighted and weighted cases, Fig. 9a and Fig. 9b, respectively. For this particular device, the reduction in γ_1 for the weighted case was offset by an increase in the γ_3 sensitivity.

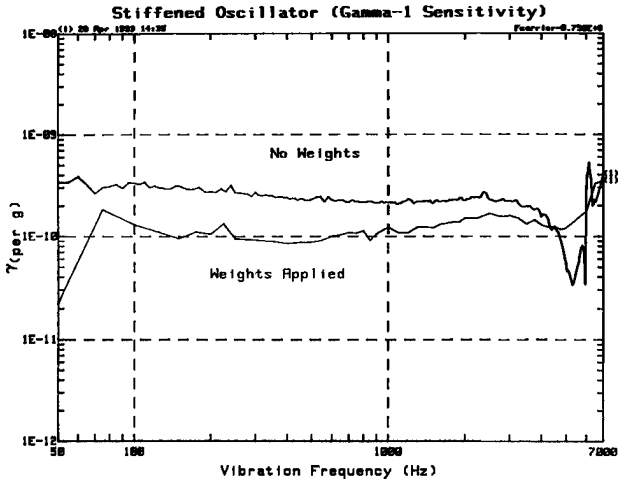


Figure 8. γ_1 as a function of vibration frequency for a sufficiently stiffened oscillator without weights and with weights applied.

IV. DISCUSSION

The most straightforward method to reduce the vibration sensitivity of an insufficiently stiffened SAW oscillator is to minimize the stress level at the AQP SAW device. One approach is to place the AQP SAW device in a region of minimal stress. We have demonstrated that the package stresses are transferred to the AQP SAW device through the mounting material. This is evident from the reduction in $\overline{\Gamma}$ which occurred when the AQP SAW device was repositioned. It is not clear from the basic stress analysis performed which stress component (or combination of stress components) in the oscillator package is transferred to the AQP SAW

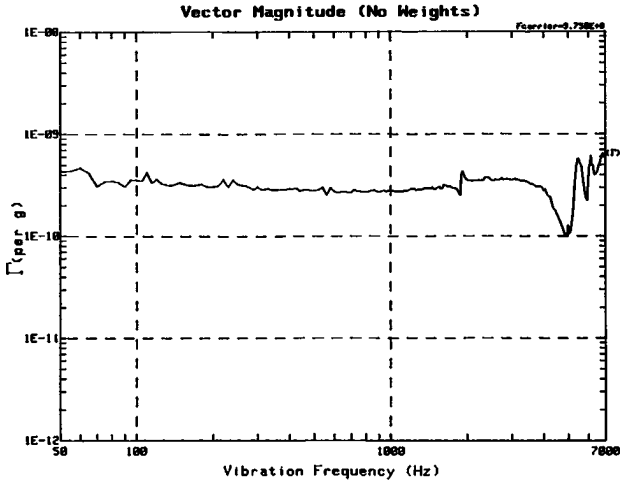


Figure 9a. Magnitude of the vibration sensitivity vector as a function of vibration frequency for a sufficiently stiffened oscillator without weight applied.

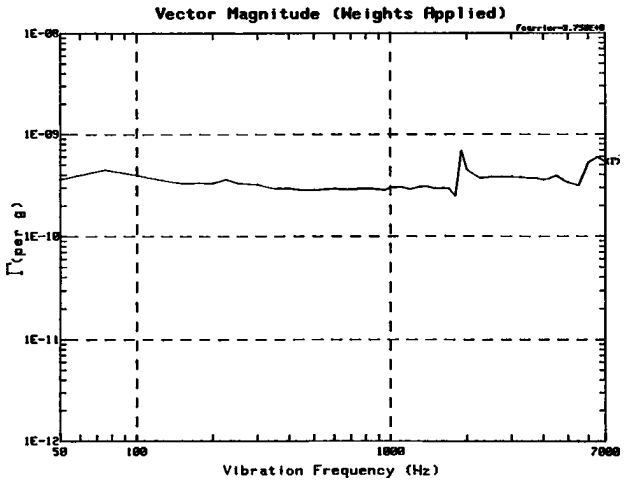


Figure 9b. Magnitude of the vibration sensitivity vector as a function of vibration frequency for a sufficiently stiffened oscillator with weights applied.

device. Previous analytical work has determined that the stress induced frequency shifts are directly proportional to the two biaxial stress components in the AQP SAW substrate [2]. These frequency shift coefficients are unequal and opposite in sign. The perturbation method was used to determine the normal stress sensitivities which are used along with the biaxial components to determine the acceleration sensitivity of the SAW device [6]. Although the biaxial stress components from the SAW substrate are necessary to calculate the acceleration sensitivity, a comparison of the biaxial stress components in the base of the hybrid package reveals a different ratio between the unrotated and

rotated cases. The ratio of the biaxial components in the rotated case, if it is assumed to reflect the stress gradients in the SAW substrate, would cause a smaller frequency perturbation than the unrotated case. This observation, along with the lower stress levels present in the rotated case, may also explain the reduction in normal vibration sensitivity. Based on these results, it is possible to use a very simple finite element model to make meaningful predictions about the vibration performance of SAW oscillators.

Recently, it has been determined that the cover and substrate thicknesses are also important parameters of vibration sensitivity. Finite element calculations on a 900 MHz AQP SAW resonator were performed. The AQP measured 0.4 inches by 0.5 inches. The calculations were performed with the AQP cover uniformly supported on a rigid substrate. The sensitivity normal to the SAW substrate is plotted versus the AQP SAW device's substrate thickness in Fig. 10. The model predicts an increase in γ_1 with increasing substrate thickness. This result is counterintuitive, since one would expect the thicker substrate to reduce the stresses and ultimately γ_1. However, this result is consistent with previous unpublished experimental observations. The same calculations were then repeated with a constant substrate thickness while varying the cover thickness. The results are plotted in Fig. 11, which show γ_1 versus the AQP SAW device's cover thickness. Increasing the cover thickness decreases γ_1 and eventually causes a relative phase change between the FM sideband of the carrier frequency and the output of the accelerometer which is monitoring the applied acceleration. The measured phases are nominally either 0° or 180°, which makes it easy to discern if the γ_1 component of the vibration sensitivity vector has changed sign (i.e., direction). Preliminary measurements on several sufficiently stiffened oscillators are consistent with these analytical results.

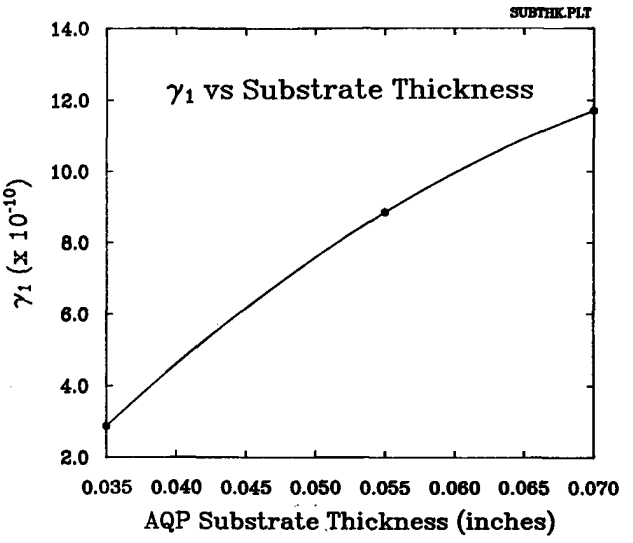


Figure 10. γ_1 as a function of AQP substrate thickness.

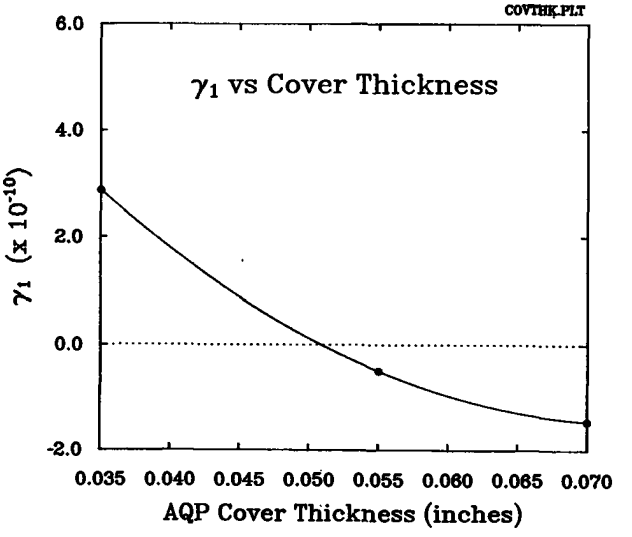


Figure 11. γ_1 as a function of AQP cover thickness.

V. SUMMARY AND CONCLUSIONS

Relatively simple finite element modeling can yield useful predictions for improving the vibration sensitivity of AQP SAW hybrid circuit oscillators. When packaging constraints prohibit the use of a sufficiently thick stiffener, the finite element method can be used to determine the optimum location for the AQP SAW device. If the AQP SAW device cannot be placed in the optimum location, then mass loading directly on the AQP SAW device can be used to improve the vibration sensitivity. When an external stiffener is thick enough, then mass loading will not decrease the vector magnitude, instead it will only rotate the vector. This attribute could be exploited in the case of a sufficiently thick stiffener when only one particular axis has a low vibration sensitivity requirement.

VI. ACKNOWLEDGMENTS

The authors would like to recognize J. Columbus for fabricating the SAW oscillators and B. Howard for fabricating the SAW devices. The authors would also like to thank J. A. Kosinski of the U.S. Army Research Laboratory, Ft. Monmouth, NJ for the software which was used to determine the rotated second order elastic coefficients.

REFERENCES

[1] J. A. Greer and T. E. Parker, "Improved Vibration Sensitivity of the All Quartz Package Surface Acoustic Wave Resonator", in *Proc. of the 42nd Annual Symp. on Freq. Control*, 1988, pp. 239-251.

[2] H. F. Tiersten and D. V. Schick, "An Analysis of the Normal Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators Rigidly Supported Along the Edges", in *Proc. of the 41st Annual Symp. on Freq. Control*, 1987, pp. 282-288.

[3] H. F. Tiersten and D. V. Schick, "An Analysis of the In-Plane Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators with Interior Rectangular Supports", in *Proc. of the 43rd Annual Symp. on Freq. Control*, 1989, pp. 396-404.

[4] T. E. Parker and J. A. Greer, "SAW Oscillators with Low Vibration Sensitivity", in *Proc. of the 45th Annual Symp. on Freq. Control*, 1991, pp. 321-329.

[5] COSMOS/M: *A proprietary finite element analysis computer program*, Structural Research and Analysis Corporation, Santa Monica, CA 90405

[6] B. K. Sinha and S. Locke, "Acceleration and Vibration Sensitivity of SAW Devices", *IEEE Trans. on Ultra., Ferr., and Freq. Control*, 1987, Vol. 34, No. 1, pp. 29-38.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DYNAMIC AND STATIC PRESSURE SENSITIVITIES OF AQP SAW OSCILLATORS

Jason Zhu, Gary K. Montress, and Thomas E. Parker

Raytheon Company
Research Division
131 Spring Street
Lexington, Massachusetts 02173
USA

ABSTRACT

The dynamic and static pressure sensitivities of two "all quartz package" (AQP) Surface Acoustic Wave (SAW) hybrid circuit oscillators, one at 375 MHz and one at 935 MHz, have been evaluated. Dynamic pressure sensitivity measurements were made from 50 Hz to 7 kHz, with pressure levels ranging from 20 to 500 Pa. For the static pressure sensitivity measurements, the pressure varied over the range from zero to two atmospheres in one atmosphere steps. Considerable care was taken in the measurement procedures to identify and minimize the effects of other competing environmental factors such as vibration and temperature changes. The oscillators were tested both with and without alumina stiffeners on the hybrid package's base and cover. The observed dynamic pressure sensitivity at 50 Hz agreed with the static pressure sensitivity within approximately a factor of three or four. The measured dynamic pressure sensitivities were in the range of $1x10^{-13}$ to $6x10^{-12}$ fractional frequency change per Pa, while the measured static pressure sensitivities were on the order of $1x10^{-13}$ fractional frequency change per Pa. Details of the measurement techniques are included and the results described.

I. Introduction

Many aspects of the frequency stability of Surface Acoustic Wave (SAW) oscillators, such as temperature and vibration sensitivities, PM and AM noise, and long-term drift, have been evaluated and reported in the literature [1]. However, there has been little, if any, investigation of the dynamic (acoustic) and/or static pressure sensitivities of these oscillators, and to the best of our knowledge, no information has been published in the open literature. The dynamic pressure sensitivity of a low noise oscillator can be an important parameter if the oscillator is to operate in an environment with very high acoustic noise levels. Equation 1 maybe used to calculate the phase noise level induced by acoustic noise:

$$\mathcal{L}(f) = 10\log\left[\left(\frac{SF_o}{f}\right)^2 \frac{P_d}{2}\right] \quad (1)$$

where

S = acoustic pressure sensitivity

$= \frac{\Delta F}{F}$ per pascal (Pa)

F_o = oscillator frequency (Hz)

f = offset (or noise) frequency (Hz)

P_d = sound pressure power spectral density (Pa^2/Hz)

As can be seen from Eq. 1, acoustically induced phase noise could severely degrade the quiescent phase noise performance of an oscillator when either S or P_d is large. Therefore, the acoustic sensitivities for two SAW oscillator designs were investigated experimentally. The oscillators' frequency stabilities were evaluated under the influence of dynamic and static pressure changes.

II. Measurement Procedures

A. *Dynamic pressure sensitivity measurement*

The oscillator under test was suspended using rubber bands, as shown in Fig. 1. It was placed inside an acoustic pressure chamber driven by two speakers mounted on opposite side-walls of the chamber, as illustrated in Fig. 2. The induced vibration levels were minimized due to the symmetry of the two speakers' locations and the rubber bands' high elasticities. The rubber bands act as "soft springs" that decouple the oscillator from the cover and side-walls of the pressure chamber. To ensure that the vibration level is low enough to allow for the accurate measurement of the oscillator's dynamic pressure sensitivity, an accelerometer was used to determine the vibration level seen by the oscillator when it was under the influence of dynamic pressure changes. The resulting residual

0-7803-1945-1/94 $3.00 © 1994 IEEE

oscillator acceleration levels were less than 0.02 *g*'s for a sinusoidal pressure level as high as 700 Pa.

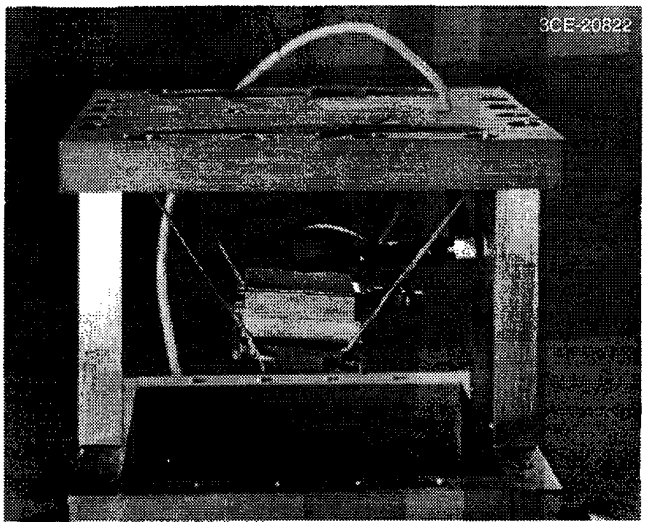


Figure 1. Oscillator suspended by rubber bands inside dynamic pressure chamber.

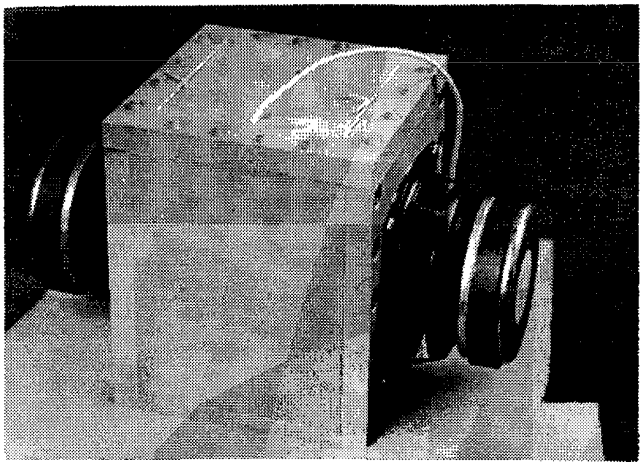


Figure 2. Outside view of the dynamic pressure chamber with two speakers symmetrically mounted.

The measurements of dynamic pressure sensitivity were made with sinusoidal pressure levels ranging from 20 to 500 Pa, in the frequency range of 50 Hz to 7 kHz. The data collection system was computer controlled and automatically made 50 or 261 measurements as the acoustic pressure varied over the above mentioned frequency range. The induced FM sidebands in the oscillator's output phase noise spectrum were measured with custom software on a Hewlett-Packard 3047A Phase Noise Measurement System using a 500 nanosecond coaxial cable delay line serving as a frequency discriminator. A microphone was placed inside the pressure chamber. Its output passed through a signal conditioning amplifier and was measured using a digital voltmeter. This signal measured the pressure level generated by the speakers inside the chamber. The microphone had a flat response over the entire frequency region that the acoustic pressure was applied. The overall system is capable of measuring dynamic pressure sensitivity as low as 10^{-13} fractional frequency change per Pa at 50 Hz and 10^{-12} fractional frequency change per Pa at 7 kHz.

The pressure chamber and the measurement system were placed in two separate rooms to ensure that the performance of the 500 nanosecond coaxial delay line was not influenced by coupling of the acoustic wave generated by the speakers.

B. *Static pressure sensitivity measurement*

The oscillator undergoing testing was placed in a chamber where the pressure could be varied from zero to two atmospheres (one atmosphere is approximately $1x10^5$ Pa) in step functions. Figure 3 is a photograph showing the outside view of the chamber, while Fig. 4 is a photograph illustrating an oscillator mounted inside the chamber. Pressure intensity was divided into three levels: zero atmosphere, one atmosphere, and two atmospheres. The fractional changes of frequency were measured as pressure varied from one level to another in a step function. The pressure induced frequency changes were resolved by observing the step like frequency changes that occurred within approximately one second after the application of the pressure change. The oscillator was thermally stabilized with a heater to ensure that it operated very close to its turn-over temperature in order that frequency offsets induced by the temperature changes could be minimized.

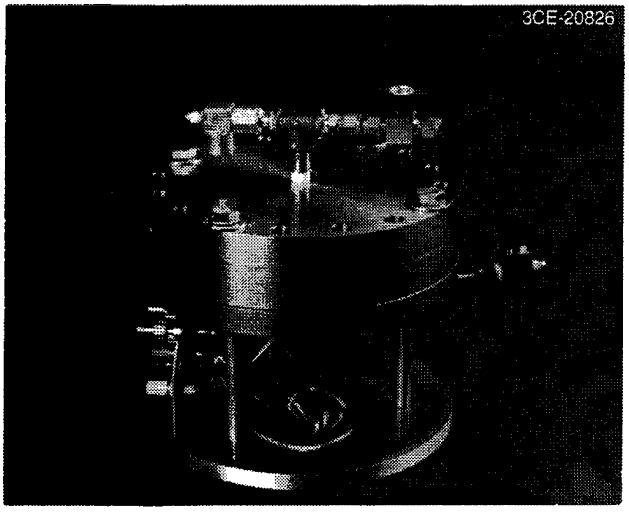


Figure 3. Outside view of the static pressure chamber.

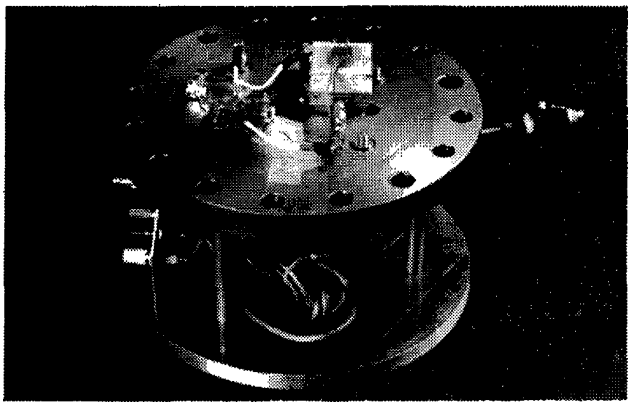


Figure 4. Oscillator inside the static pressure chamber with a heater attached.

III. Measurement Results and Discussion

The dynamic and static sensitivities of two AQP SAW oscillators, one at 935 MHz and the other at 375 MHz, were measured. They were tested both with and without ceramic stiffeners on the hybrid package's base and cover [2].

Figure 5 shows the results of dynamic pressure sensitivity measurements versus acoustic frequency for a 935 MHz SAW oscillator with both stiffeners in place. The lower curve in the figure is the measurement system floor. As can be seen from the figure, there is clearly a measurable dynamic pressure sensitivity level that is about $5x10^{-13}$ fractional frequency change per Pa at the lower end of the measurement frequency range and $5x10^{-12}$ fractional frequency change per Pa at the upper end of the measurement frequency range.

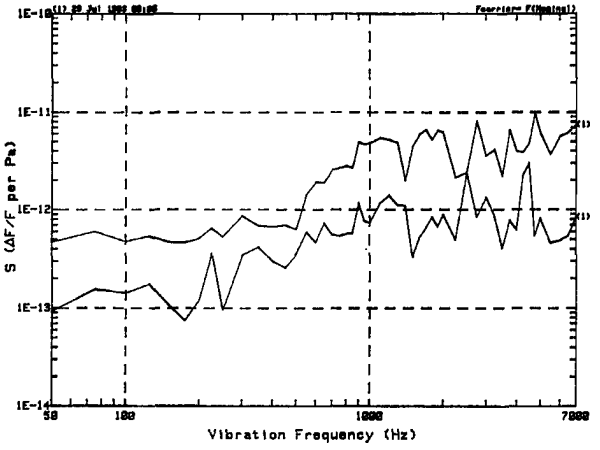


Figure 5. Dynamic sensitivity level of the 935 MHz oscillator with both stiffeners attached as the frequency goes from 50 Hz to 7 kHz. Lower curve is the measurement system floor.

The measured dynamic pressure sensitivity level for the same oscillator without stiffeners is shown in Fig. 6. The sensitivity level is clearly increased due to the fact that the stiffeners were removed. The sensitivity level is now in the range of $2x10^{-12}$ to $5x10^{-12}$ fractional frequency change per Pa.

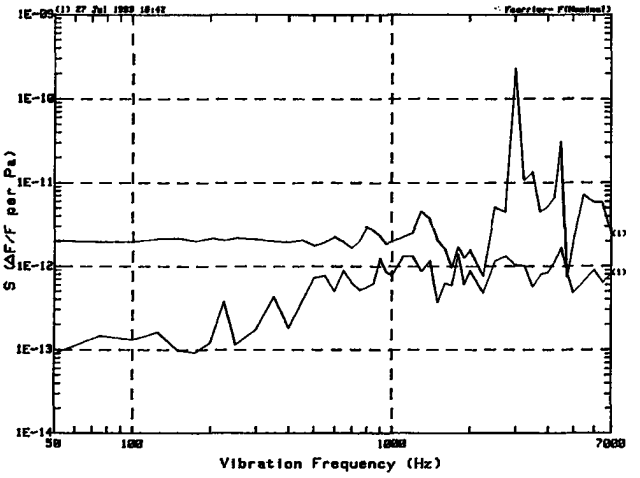


Figure 6. Dynamic sensitivity level of the 935 MHz oscillator with both stiffeners removed as the frequency goes from 50 Hz to 7 kHz. Lower curve is the measurement system floor.

Every effort was made to minimize other competing environmental factors such as temperature changes and vibration. The relatively short period of the sinusoidal dynamic pressure measurements and the large thermal time constants of the oscillators (~100 seconds) removed any problem with temperature fluctuations. In order to find out if vibration was a significant contributing factor in the measurements, the residual acceleration level seen by the oscillator was measured, and it came out to be less than 0.02 *g*'s. The vibration sensitivities of these oscillators were measured to be about 10^{-10} fractional frequency change per *g*. Thus, it was concluded that vibration effects also had minimal influence on the performance of the oscillators in our measurements.

After getting preliminary data for the dynamic (acoustic) sensitivities of the two SAW oscillators, the static sensitivities of the oscillators were measured to see if the dynamic and static sensitivities were in reasonable agreement. Figure 7 shows the result of a measurement of the 935 MHz oscillator's static pressure sensitivity with both stiffeners in place as the pressure went from one atmosphere up to two atmospheres at t=100 seconds, and then returned to one atmosphere at t=500 seconds. Thermal effects were found to be a major factor in the

static pressure sensitivity measurements. The small peak at t=120 seconds was caused by the oscillator's dynamic temperature coefficient and the large frequency offset following that was the effect of the static temperature coefficient. However, the pressure induced frequency changes could be resolved by observing the step like frequency changes that occurred within approximately one second after the application of the pressure change since the oscillators have thermal time constants on the order of 100 seconds. Thus, as long as the frequency change immediately after the application of the pressure change can be observed, it is possible to distinguish the pressure induced frequency changes from those caused by thermal effects.

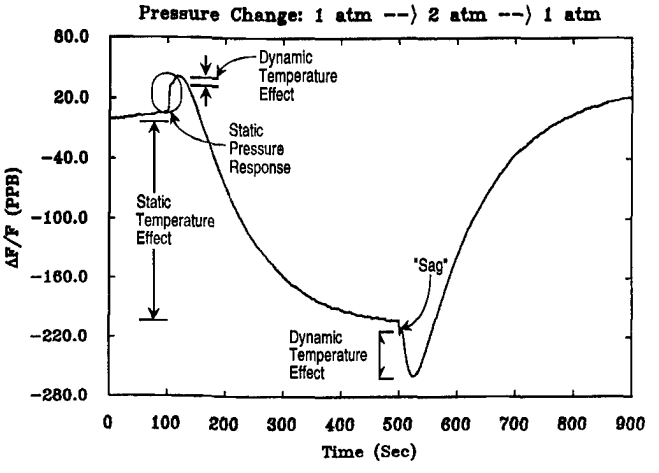


Figure 7. Fractional frequency response for the 935 MHz oscillator with both stiffeners attached as the pressure goes from 1 atm to 2 atm at t=100s and then comes back down to 1 atm at t=500s.

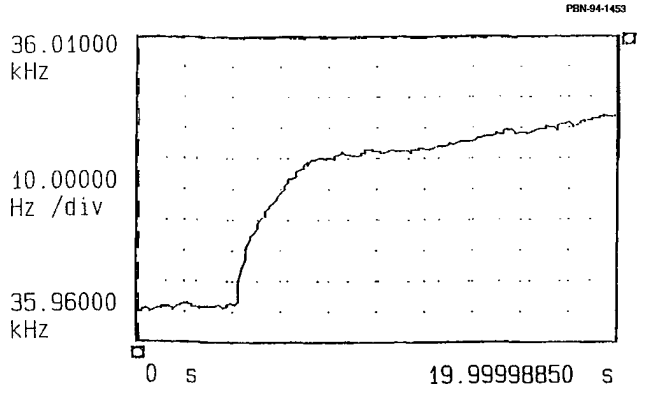


Figure 8. Close-up view of the frequency response for the 935 MHz oscillator with both stiffeners attached as the pressure goes from 1 atm to 2 atm.

Figure 8 is a close-up view of the oscillator's frequency change as the pressure varied from one atmosphere to two atmospheres at t=100 seconds. As shown in the figure, the frequency of the oscillator increased in a step like manner immediately after application of the pressure step. The static pressure sensitivity was calculated by first dividing the frequency change by the carrier frequency, which is 935 MHz in this case, and then dividing the result by the change of pressure that caused the frequency variation.

One puzzling result of the measurements is that the magnitudes of the fractional frequency changes per Pa as the pressure increased and decreased by one atmosphere did not exactly agree. Another phenomenon, as yet unexplained, is that as the pressure was decreased by one atmosphere there was a "sag" that did not appear when the pressure was increased by one atmosphere. A close-up view of this "sag" is shown in Fig. 9. These observations would need additional investigation in order to determine explanations for their origin.

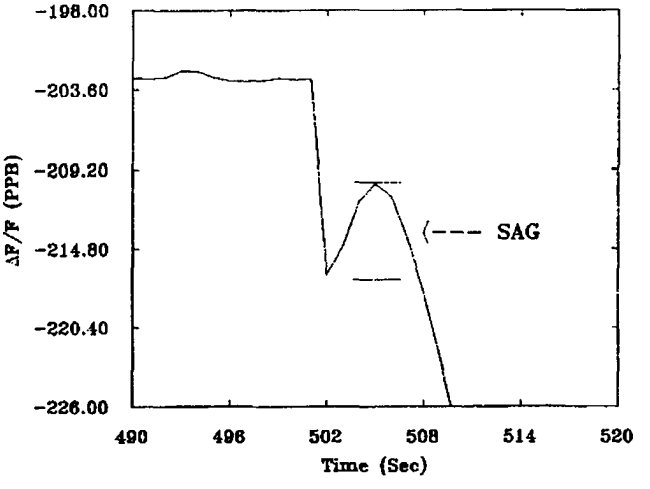


Figure 9. Close-up view of the fractional frequency response "sag" as the pressure goes from 2 atm to 1 atm.

The general ranges of acoustic sensitivity coefficients for the 935 MHz and 375 MHz oscillators are summarized in Table I. The measured dynamic (50 Hz) and static pressure sensitivities are not exactly equal, but agree within approximately a factor of three or four. One possible reason for the difference may be that the pressure levels were significantly different in the two types of measurements. The typical pressure level for the dynamic pressure sensitivity measurements was approximately 500 Pa, while it was about 10^5 Pa for the

Table I
Ranges of Acoustic Sensitivity Coefficients of 935 MHz and 375 MHz Oscillators

	Dynamic Pressure Sensitivity (ΔF/F per Pa)	Static Pressure Sensitivity (ΔF/F per Pa)
935 MHz Oscillator With Stiffeners	$5x10^{-13}$ to $5x10^{-12}$	~ $2x10^{-13}$
935 MHz Oscillator Without Stiffeners	$2x10^{-12}$ to $5x10^{-12}$	~ $7x10^{-13}$
375 MHz Oscillator With Stiffeners	$2x10^{-13}$ to $2x10^{-12}$	~ $8x10^{-14}$
375 MHz Oscillator Without Stiffeners	$2x10^{-12}$ to $1x10^{-11}$	~ $5x10^{-13}$

static pressure measurements. Another possible cause of the discrepancy may be that the time scales of the pressure changes were very different for the two measurements.

IV. Conclusions

Acoustic sensitivity is an important parameter related to a SAW oscillator's frequency stability. In an environment with high acoustic noise levels, the phase noise performance of an oscillator can be severely degraded. Preliminary information regarding the dynamic and static pressure sensitivities for two AQP SAW oscillators, one at 375 MHz and the other at 935 MHz, has been described. This is the first such published report, to the best of our knowledge, on this important topic.

An HP3047 Phase Noise Measurement System with a 500 nanosecond coaxial delay line serving as a frequency discriminator was used to measure the induced FM sidebands in the oscillators during the dynamic sensitivity measurements. The dynamic and static pressure sensitivity measurements were both difficult to perform. The frequency fluctuation induced by the pressure changes had to be distinguished from those induced by other environmental factors, such as temperature fluctuations and vibration.

Temperature fluctuations were not a problem in the dynamic pressure sensitivity measurements since the oscillators have large thermal time constants and the acoustic wave applied in the measurements had a very short period which did not allow the temperature to vary. To minimize acceleration induced effects, the speakers were mounted symmetrically on the sound chamber and rubber bands were used to decouple the oscillators from the walls of the sound chamber. It turned out that the acceleration levels were less than 0.02 *g*'s for a sinusoidal pressure level as high as 700 Pa. Typical results for the dynamic sensitivity measurements were in the range of $1x10^{-13}$ to $6x10^{-12}$ fractional frequency change per Pa, depending on the physical configuration of the oscillator.

Temperature fluctuations presented a major problem for the static pressure sensitivity measurements. However, the pressure induced frequency changes could still be resolved by observing a step like frequency change immediately after the application of the step change in pressure. The pressure step applied in the static measurements was on the order of 10^5 Pa, much larger than that applied in the dynamic measurements. The typical results indicated that the static sensitivities were on the order of 10^{-13} fractional frequency change per Pa.

In general, the measured dynamic and static pressure sensitivities agreed within approximately an order of magnitude, suggesting that the results are reasonably accurate. In order to see how the measured sensitivity

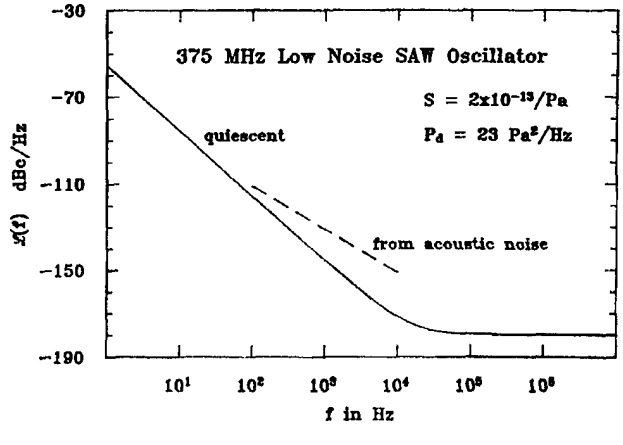


Figure 10. Predicted acoustically induced phase noise level in a typical missile environment.

would affect the performance of an oscillator in a typical missile environment, Fig. 10 illustrates the potential phase noise degradation for a low noise oscillator subjected to acoustic noise. As shown by the graph, when the acoustic level is high enough (P_d = 23 Pa^2/Hz), the acoustically induced phase noise seriously interferes with the performance of the oscillator. This reinforces the importance of the acoustic sensitivity as a parameter potentially affecting the spectral performance of AQP SAW oscillators.

References

[1] T.E. Parker and G.K. Montress, "Precision Surface-Acoustic-Wave (SAW) Oscillators," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. UFFC-35, No. 3, pp. 342-364, May 1988.

[2] T.E. Parker, J.A. Greer, and G.K. Montress, "SAW Oscillators With Low Vibration Sensitivity," in *Proceedings of the 45th Annual Symposium on Frequency Control*, 1991, pp. 320-329, IEEE Catalog Number 91CH2965-2.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DESIGN AND PERFORMANCE OF AN EXTREMELY LOW NOISE SURFACE ACOUSTIC WAVE OSCILLATOR

Gary K. Montress and Thomas E. Parker

Raytheon Company
Research Division
131 Spring Street
Lexington, Massachusetts 02173
USA

ABSTRACT

The design, component selection, and performance of an extremely low noise (PM and AM) all quartz package (AQP) surface acoustic wave (SAW) resonator hybrid circuit oscillator are described in detail. The prototype 400 MHz hybrid circuit oscillator's measured white phase noise floor is -185 dBc/Hz for carrier offset frequencies greater than 400 kHz. This is the best result yet reported for a SAW resonator oscillator's white phase noise floor, particularly for a hybrid circuit oscillator, and is almost 2 dB better than the previous best reported result for a discrete component, printed circuit board design. The oscillator's flicker FM ($S_\phi(f) \sim f^{-3}$) phase noise level at 10 Hz carrier offset is approximately -80 dBc/Hz. A UTO-1023 (Avantek) amplifier in substrate carrier form was used in the oscillator's feedback loop, and a buffer amplifier was not included in this prototype design.

I. INTRODUCTION

State-of-the-art phase noise performance has been demonstrated for both surface acoustic wave (SAW) resonator and delay line stabilized oscillators [1] - [7]. The same basic feedback loop oscillator design philosophy was applied in each case in order to achieve these results [8]. In this paper, the performance of a prototype 400 MHz AQP SAW resonator hybrid circuit oscillator is presented. The oscillator's measured white phase noise floor is -185 dBc/Hz for carrier offset frequencies greater than 400 kHz. This is almost 2 dB better than the previous best result for a discrete component, printed circuit board SAW resonator oscillator design [4], [5]. The oscillator's flicker FM noise level at 10 Hz carrier offset is approximately -80 dBc/Hz. Residual noise measurements on critical loop components played an important role in achieving these results, including residual PM and AM noise measurements on the loop amplifier. Details of these measurements are described, including an observed noise anomaly in the high power silicon bipolar transistor RF amplifier which was selected for this application.

Typically, the oscillator's vibration sensitivity ***magnitude*** was found to be in the 1 to $5 \times 10^{-10}/g$ range. This combination of excellent vibration sensitivity and phase noise performance in an AQP SAW resonator hybrid circuit oscillator, when coupled with the exceptionally good long-term (aging) frequency stability demonstrated by AQP sealed SAW devices, represents the current state-of-the-art in SAW resonator based hybrid circuit oscillator performance.

The oscillator's design and component level testing are described, and details of its performance are given, including: PM noise, AM noise, fractional frequency stability and output power variation versus temperature, load pulling, dc power supply voltage pushing, three-axis vibration sensitivity, electronic tuning sensitivity characteristic, and harmonic output levels. Design trade-offs are discussed in the context of the specific performance achieved.

II. OSCILLATOR DESIGN

Figure 1 shows a circuit schematic for the prototype, extremely low noise AQP SAW resonator hybrid circuit oscillator. A basic feedback loop oscillator architecture design

0-7803-1945-1/94 $3.00 © 1994 IEEE

approach was used, including: a single loop amplifier, an electronic phase shifter, a lumped element reactive power splitter (Wilkinson type), lumped element reactive phase adjusting circuitry, an attenuator for loop gain adjustment, and a low pass output filter for harmonic suppression. A buffer amplifier was not incorporated into the design in order that it not degrade the oscillator's white phase noise floor.

Figure 2 shows a photograph of the prototype hybrid circuit oscillator. The AQP SAW resonator device is a dual track design [7], with each track's acoustic aperture (finger overlap) approximately 125 wavelenths. The acoustic cavities are each overmoded (three resonances) as described previously, with effective cavity lengths of 365 wavelengths [7]. The transducers are apodized using a simple cosine weighting, and each transducer contains 77 fingers.

The test points labeled TP_1 and TP_2 in Fig. 1 are used to open the feedback loop, thereby permitting use of a network analyzer to establish the appropriate loop gain and phase conditions for oscillation, namely: approximately 3 dB (2 to 4 dB is typical) of excess small-signal gain (nominally 3 dB of loop gain compression when equilibrium conditions are reached after turn-on) and $2\pi N$ radians of net transmission phase shift through the loop, where N is an integer. An attenuator is used to adjust the amount of excess small-signal loop gain, as indicated, while the fixed component reactive phase adjust networks are used for coarse adjustment of the loop transmission phase shift [9], [10]. Minor adjustments to the meander lines are used to achieve accurate set-on of the oscillator's initial frequency. The electronic phase shifter is used to electronically tune the oscillator's frequency over the SAW resonator's 1 dB bandwidth. It is also useful during the course of phase noise measurements using the conventional two oscillator phase-locked technique.

A UTO-1023 (Avantek) type high power RF amplifier in substrate carrier form was selected for use as the loop amplifier. However, its nominal gain of 14 dB was found to be approximately 3 dB greater than would otherwise be optimum for use in these prototype oscillators. This is because any attenuation added to the loop in order to properly set-up loop oscillation conditions will degrade the oscillator's white phase noise floor [7]. In order to achieve the best possible white phase noise floor, the amplifier's small-signal gain was decreased by 3 dB. Figure 3 shows a simplified basic circuit schematic for the amplifier. Where multiple resistors are shown in the circuit schematic, wire jumpers could be used to either connect or short out individual resistors in the circuit in order to adjust the amplifier's DC and/or RF performance. The amplifier's small-signal gain was reduced 3 dB by simply decreasing the value of the total resistance in the amplifier's primary feedback path. This reduction was achieved at the expense of only a 0.5 dB degradation in the amplifier's small-signal noise figure. The amplifier's output power at 3 dB of gain compression was unchanged by this gain adjustment. As a direct consequence of this gain change, a resulting oscillator's white phase noise floor should be approximately 2.5 dB better (lower) than if the change had not been made. Typically, at most, a 0.25 dB or 0.50 dB attenuator was all that had to be used for loop gain adjustment in order to properly set-up the oscillator's feedback loop for operation.

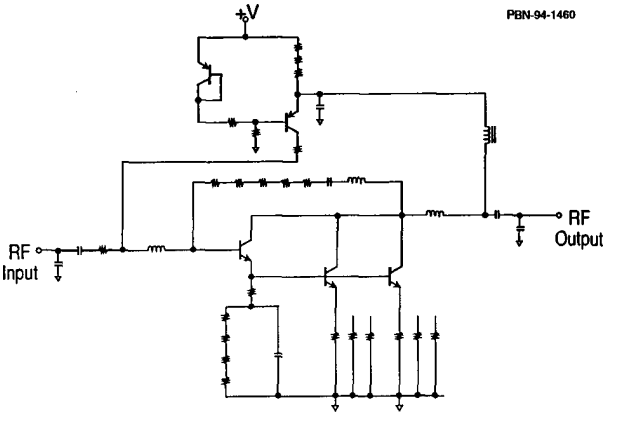


Figure 3. Basic circuit schematic for a high power, low noise silicon bipolar transistor RF amplifier.

III. OSCILLATOR PERFORMANCE

Figure 4(a) shows the measured phase noise spectrum for ***two*** oscillators. Also shown in Fig. 4(a) is the system noise floor for the measurement set-up. The total white phase noise measured for the pair of oscillators is

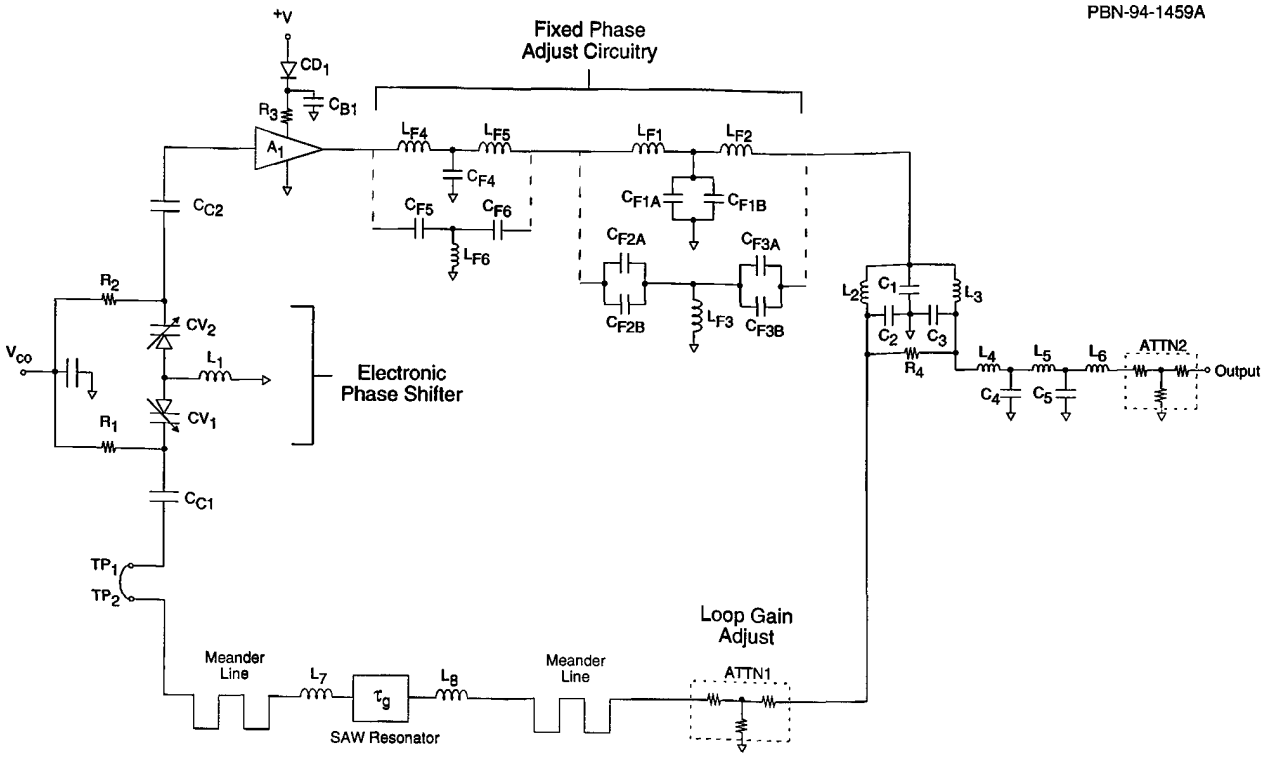


Figure 1. Circuit schematic for an extremely low noise AQP SAW resonator hybrid circuit oscillator.

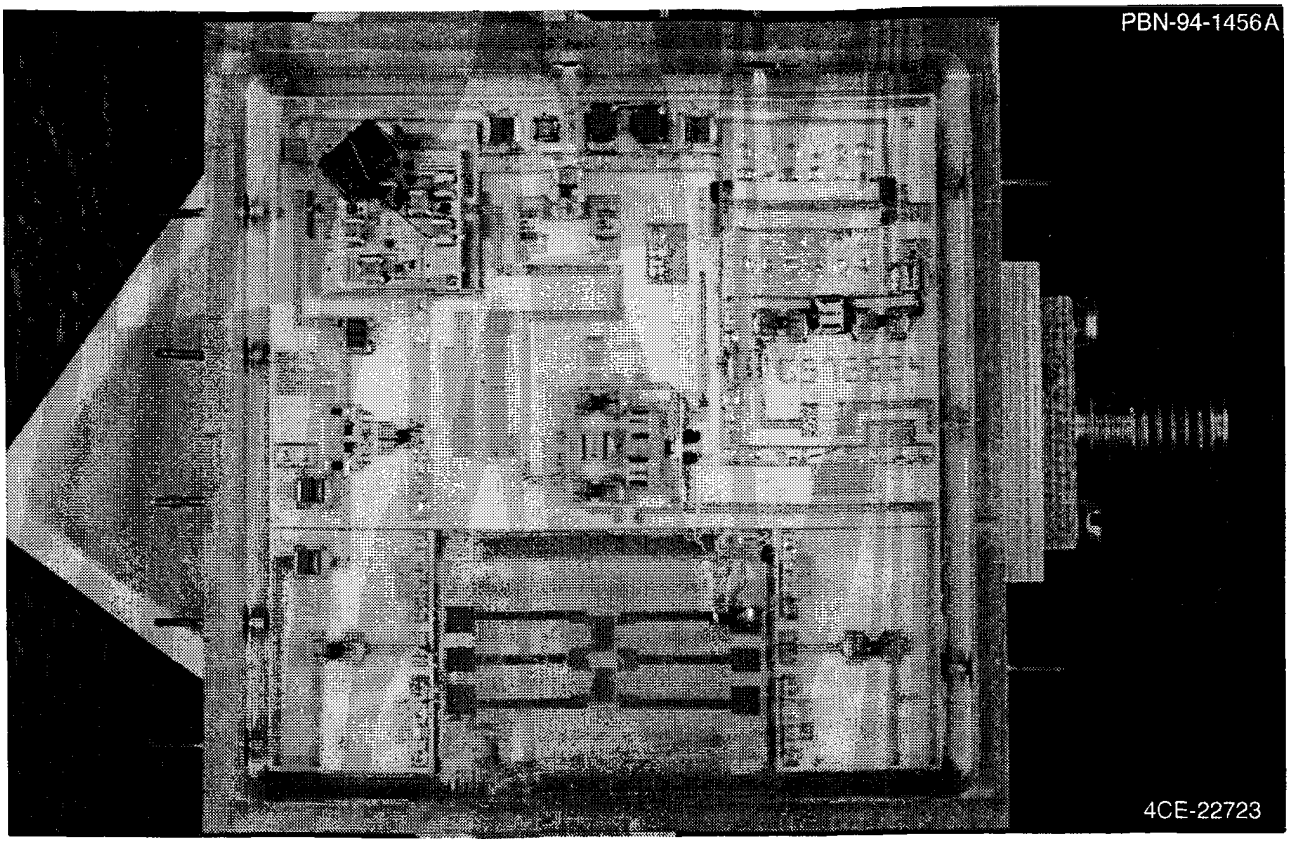


Figure 2. Photograph of an extremely low noise AQP SAW resonator hybrid circuit oscillator.

within 5 dB of the system floor, and therefore must be corrected to eliminate the system floor's contribution to the noise measured for the pair of oscillators. Corrections are accomplished by first subtracting the system floor from the two oscillator phase noise measurement, and then subtracting 3 dB from the result under the assumption that the two oscillators possess otherwise identical phase noise spectra. Figure 4(b) shows the result of this system floor correction having been carried out, along with a 3 dB correction to represent the phase noise spectrum of just one oscillator. It is evident that the oscillators' white phase noise floors are each approximately -185 dBc/Hz, for carrier offset frequencies greater than 400 kHz, under the assumption that the oscillators are contributing equally to the phase noise measurement. The flicker FM noise level for each oscillator is approximately -80 dBc/Hz at 10 Hz carrier offset. Both oscillators were independently verified to be very close to this flicker FM level as a result of separate phase noise measurements of each against a lower noise (for carrier offsets less than 300 Hz) frequency synthesizer reference oscillator (Hewlett-Packard 8662A).

The AM noise of each oscillator was also measured. Figure 5(a) illustrates the result for ***one*** of the oscillators. The system floor for the measurement (AM noise threshold sensitivity [11], [12]) is shown as the lower curve in the figure. Note that for carrier offset frequencies greater than approximately 200 kHz the measured AM noise spectrum for the oscillator is essentially equal to the system floor. For offsets greater than 200 kHz, all we can really say with any degree of certainty is that the oscillator's AM noise floor is at least 3 dB lower than the measurement system noise floor. However, the system floor may be subtracted from the measured AM noise spectrum for the oscillator in order to at least attempt to correct for the system floor's contribution to the oscillator's AM noise measurement. Figure 5(b) shows the result of this correction having been done. The oscillator's AM noise level for offsets greater than 200 kHz is likely even lower than the -185 dBc/Hz indicated. However, there is an interesting "bulge" in the oscillator's AM noise spectrum for carrier offset frequencies between 10 kHz and 100 kHz. A close examination of the oscillator's PM noise spectrum in Fig. 4(b) also shows indications of a slight "bulge" or "plateau" in the same carrier offset frequency region. Since the phase detector's AM suppression for the phase noise measurements was greater than 20 dB, the observed distortion in the oscillator's phase noise spectrum is not likely a measurement artifact attributable to the oscillator's AM noise "bulge" or "plateau", but rather it appears to be a real distortion in the oscillator's PM noise spectrum as well.

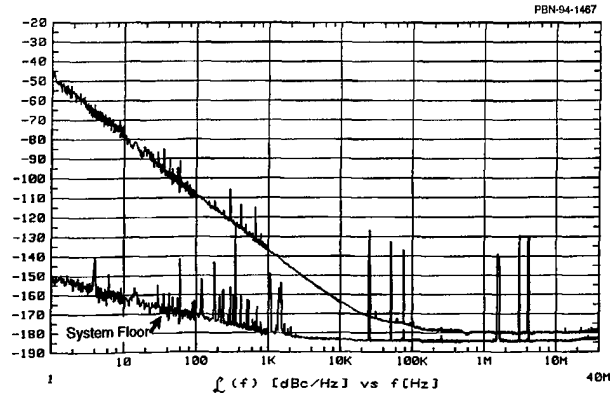


Figure 4(a). Measured phase noise for ***two*** prototype extremely low noise AQP SAW resonator hybrid circuit oscillators. Lower curve is the measurement system's noise floor.

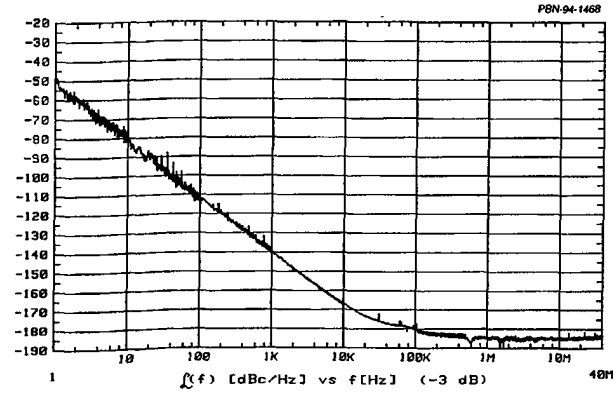


Figure 4(b). Measured phase noise for ***one*** prototype extremely low noise AQP SAW resonator hybrid circuit oscillator after system floor correction has been made, and 3 dB has been subtracted under the assumption that the two oscillators are identical.

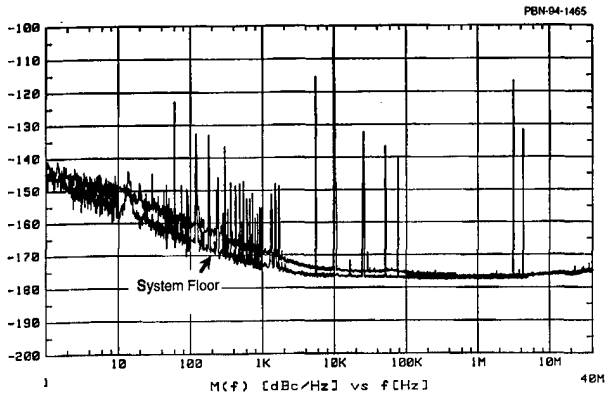


Figure 5(a). Measured AM noise spectrum for ***one*** prototype extremely low noise AQP SAW resonator hybrid circuit oscillator. Lower curve is the measurement system's noise floor.

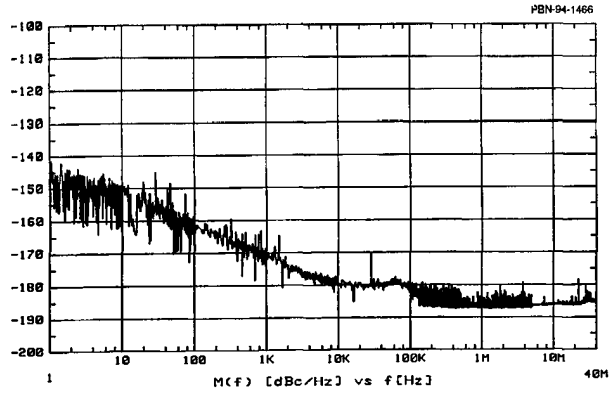


Figure 5(b). Measured AM noise for ***one*** prototype extremely low noise AQP SAW resonator hybrid circuit oscillator after system floor correction has been made.

In order to further investigate the source of these noise spectra distortions, residual PM and AM noise measurements were performed on a similar amplifier (not one of the two amplifiers used in the prototype oscillators). Figure 6 shows the result of a residual PM noise measurement performed on this similar amplifier using standard techniques [13]. The measurement system's noise floor is also shown as the lower curve in the figure. There is perhaps a slight hint of a "bulge" or "plateau" in the 10 kHz to 100 kHz offset frequency range. However, AM-to-PM conversion of the test source's AM noise by the amplifier under test precludes a really accurate assessment. Figure 7 shows the result of a residual AM noise measurement on the same similar amplifier. The system noise floor is shown in the figure. Note that the system noise floor crosses over the residual AM noise measurement on the amplifier at a carrier offset frequency of approximately 10 kHz. For this measurement the system noise floor is simply the AM noise spectrum of the test source used. When the device under test (i.e., the amplifier) is placed in the test set-up and evaluated with about 3 dB of gain compression, the amplifier actually strips off some of the test source's AM noise and reveals the amplifier's own residual AM noise level for offsets greater than 10 kHz to be at least 4 or 5 dB better than what would have otherwise been the limit set by the test source's own AM noise spectrum. Once again, there is perhaps a tantalizing hint of a "bulge" or "plateau" in the amplifier's residual AM noise spectrum. However, both residual noise measurements lack sufficient sensitivity to truly reveal accurate details of either the amplifier's residual PM or AM noise spectrum. Improved sensitivity is obviously required, such as would be provided by the cross-correlation technique [14], [15].

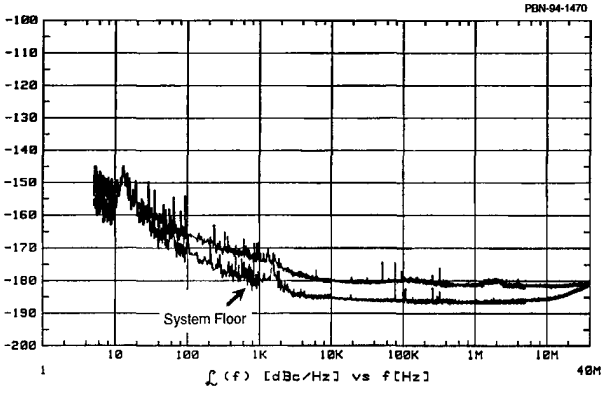


Figure 6. Measured residual PM noise spectrum for a UTO-1023 type amplifier. Measurement system noise floor is also shown.

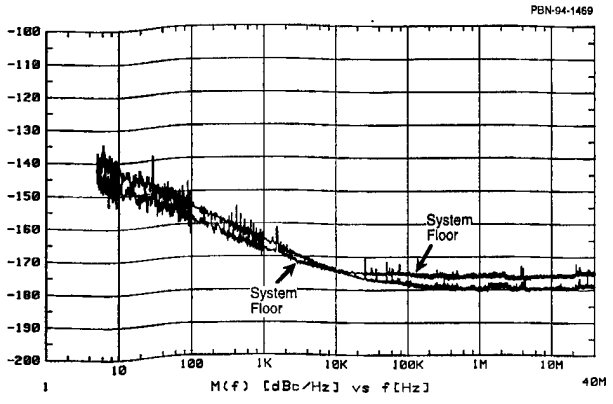


Figure 7. Measured residual AM noise spectrum for a UTO-1023 type amplifier. Measurement system noise floor is also shown.

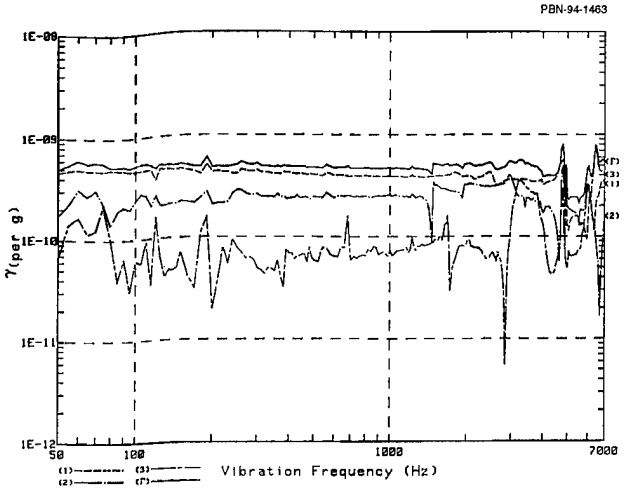


Figure 8(a). Vibration sensitivity performance for a prototype extremely low noise 400 MHz AQP SAW resonator hybrid circuit oscillator.

The prototype oscillators' vibration sensitivities were measured. Figure 8(a) shows the measured vibration sensitivity ***magnitude*** for one of the prototype oscillators (the solid curve in the figure). Also shown are the individual sensitivities for each of three mutually orthogonal axis. For this particular oscillator the vibration sensitivity was dominated by the γ_1 component for vibration applied in a direction normal to the plane of acoustic propagation [16], [17]. Figures 8(b) and 8(c) show typical vibration sensitivity results for several similarly designed oscillators. The measured vibration sensitivity ***magnitudes*** for typical oscillator designs of this type are usually in the 1 to $5x10^{-10}$/g range. These results are consistent with previously reported results for a variety of AQP SAW hybrid circuit oscillator designs [16]-[18].

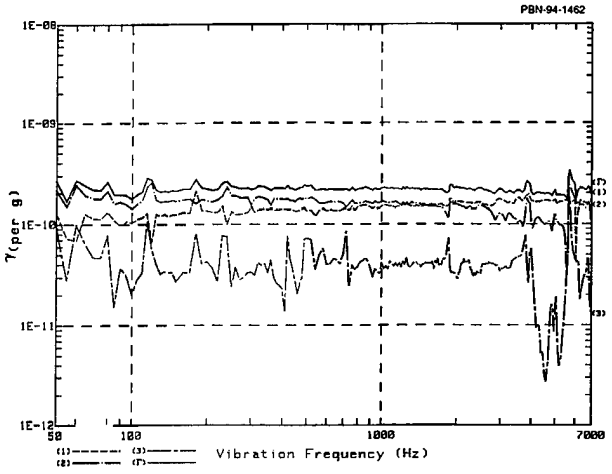


Figure 8(b). Vibration sensitivity performance for a similarly designed low noise 350 MHz AQP SAW resonator hybrid circuit oscillator design.

Figure 9 shows an oscillator's fractional frequency and output power variations with temperature. The oscillator's turn-over temperature is approximately 75°C. The oscillator's nominal output power is in excess of +22 dBm, and varies by less than a total of 0.5 dB over the -55°C to +100°C range. This is an exceptionally good result and is essentially a tribute to the gain compression stability of the loop amplifier.

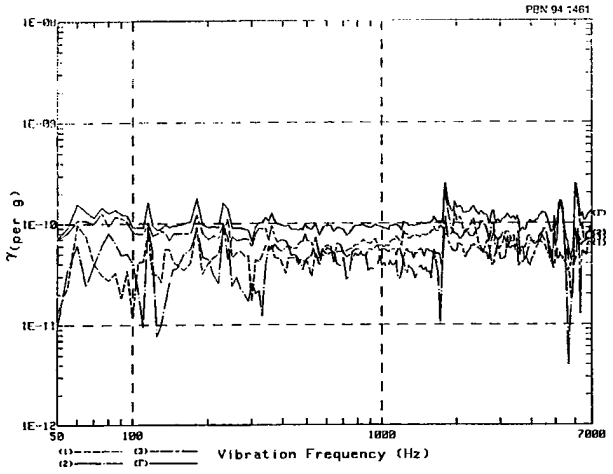


Figure 8(c). Vibration sensitivity performance for a similarly designed low noise 375 MHz AQP SAW resonator hybrid circuit oscillator design.

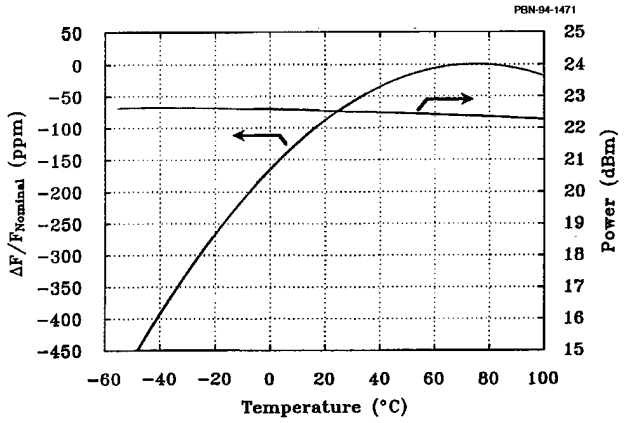


Figure 9. Fractional frequency and output power stabilities versus temperature for a prototype extremely low noise AQP SAW resonator hybrid circuit oscillator.

Figure 10 illustrates the typical electronic tuning sensitivity curve for one of the prototype oscillators. The nominal tuning slope is approximately +8 ppm/volt. The total usable fractional frequency tuning is greater than ±35 ppm for a tuning voltage range of 0v to +10v.

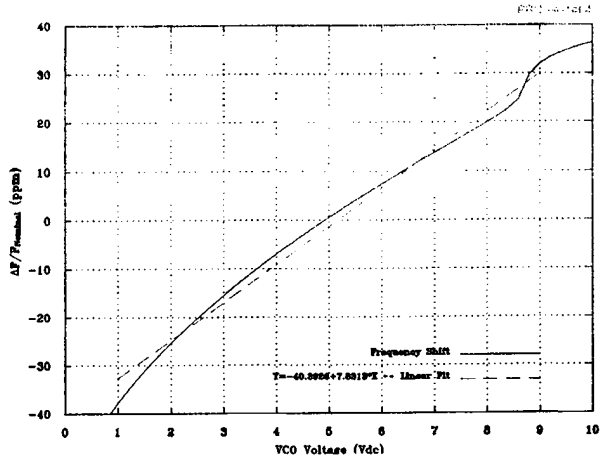


Figure 10. Fractional frequency versus tuning voltage curve for a prototype extremely low noise AQP SAW resonator hybrid circuit oscillator.

Figure 11 illustrates typical aging data taken on several AQP SAW resonator hybrid circuit oscillators where the dissipated power and peak stresses in the SAW resonators are comparable to the corresponding levels in the prototype hardware described herein. This level of performance, basically less than ±1 ppm/year, is perfectly well suited to a wide range of ground based and airborne radar system applications.

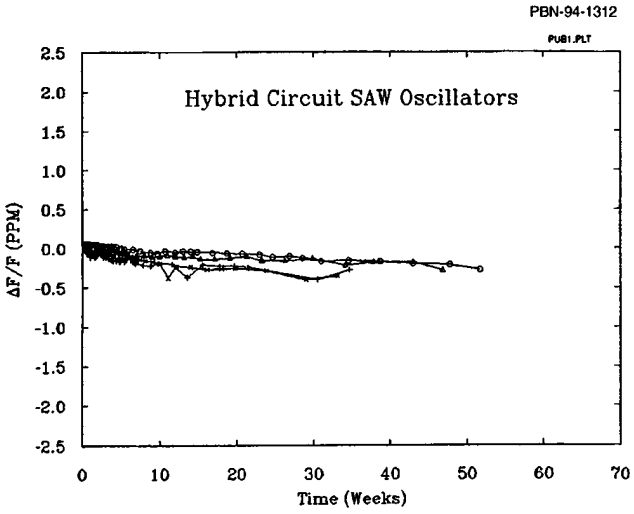


Figure 11. Typical long-term fractional frequency stability for AQP SAW hybrid circuit oscillators. The resonators' power dissipations and peak stresses are comparable to the operating conditions for the oscillators described herein.

Finally, Table I summarizes several other performance parameters for the two prototype oscillators, including harmonic output levels with and without filtering, load pulling, and voltage pushing. The load pulling is somewhat larger than might otherwise be expected. However, this is solely due to the fact that there is no buffer amplifier and/or coupling attenuator used in the prototype hardware that would help moderate the effect of load changes on the oscillator's operating frequency.

Table I

PBN-94-1457

- **Harmonics:**

	$2f_0$	$3f_0$	$4f_0$	$5f_0$	$6f_0$	$7f_0$	$8f_0$
w/filter (dBc)	-56	-65	-87	-87	-96	-102	-109
w/o filter (dBc)	-34	-39	-48	-55	-73	-70	-107

- **Load pulling: $< \pm 10$ ppm, 2:1 VSWR any phase**
- **Voltage pushing: ~ +4 ppm/volt**
- **Output power: ~ +22 dBm**
- **Tuning sensitivity: ~ +8 ppm/volt**

IV. SUMMARY

The performance of two prototype AQP SAW resonator hybrid circuit oscillators has been described. The phase noise performance achieved represents the current state-of-the-art for SAW resonator based oscillators operating in this frequency range. In addition, very good vibration sensitivity was also demonstrated. As noted, very careful attention must be paid to each component (e.g., amplifier, SAW resonator, varactors, etc.) if truly low noise levels (PM and AM) are to be achieved.

REFERENCES

[1] G. K. Montress, T. E. Parker, and M. J. Loboda, "Extremely Low Phase Noise SAW Resonator Oscillator Design and Performance", in *Proceedings of the IEEE Ultrasonics Symposium*, 1987, vol. 1, pp. 47-52.

[2] G. K. Montress, T. E. Parker, M. J. Loboda, and J. A. Greer, "Extremely Low-Phase-Noise SAW Resonators and Oscillators: Design and Performance", *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 35, no. 6, pp. 657-667, November 1988.

[3] T. E. Parker and G. K. Montress, "Low Noise SAW Resonator Oscillators", in *Proceedings of the 43rd Annual Symposium on Frequency Control*, 1989, pp. 588-595.

[4] T. E. Parker and G. K. Montress, "Frequency Stability of High Performance SAW Oscillators", in *Proceedings of the IEEE Ultrasonics Symposium*, 1989, vol. 1, pp. 37-45.

[5] G. K. Montress and T. E. Parker, "Design Techniques for Achieving State-of-the-Art Oscillator Performance", in *Proceedings of the 44th Annual Symposium on Frequency Control*, 1990, pp. 522-535.

[6] G. K. Montress, T. E. Parker, D. W. Kress, and J. A. Kosinski, "Design and Performance of a Low Noise, Wide Tuning Range AQP SAW Delay Line VCO", in *Proceedings of the IEEE Frequency Control Symposium*, 1992, pp. 356-370.

[7] G. K. Montress and T. E. Parker, "State-of-the-Art SAW Oscillators: Design and Performance", in *Proceedings of the 7th European Frequency and Time Forum*, 1993, pp. 439-454.

[8] T. E. Parker and G. K. Montress, "Precision Surface-Acoustic-Wave (SAW) Oscillators", *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 35, no. 3, pp. 342-364, May 1988.

[9] G. K. Montress and T. E. Parker, "High Stability SAW Oscillators: Design and Performance", Part II, Short Course presented at the 1992 IEEE Ultrasonics Symposium, 20 October 1992, Tucson, AZ.

[10] G. K. Montress and T. E. Parker, "Introduction to High Stability SAW Oscillators: Design and Performance", Part II, Tutorial Short Course presented at the 1994 IEEE Frequency Control Symposium, 31 May 1994, Boston, MA.

[11] J. G. Ondria, "A Microwave System for Measurements of AM and FM Noise Spectra", *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-16, no. 9, pp. 767-781, September 1968.

[12] J. L. Fikart and P. A. Goud, "The Direct-Detection Noise-Measuring System and Its Threshold," *IEEE Transactions on Instrumentation and Measurement*, vol. IM-21, no. 3, pp. 219-224, August 1972.

[13] G. K. Montress, T. E. Parker, and M. J. Loboda, "Residual Phase Noise Measurements of VHF, UHF, and Microwave Components", in *Proceedings of the 43rd Annual Symposium on Frequency Control*, 1989, pp. 349-359.

[14] W. F. Walls, "Cross-Correlation Phase Noise Measurements", in *Proceedings of the IEEE Frequency Control Symposium*, 1992, pp. 257-261.

[15] F. G. Ascarrunz, E. S. Ferre, and F. L. Walls, "Investigations of AM and PM Noise in X-Band Devices", in *Proceedings of the IEEE Frequency Control Symposium*, 1993, pp. 303-311.

[16] 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 Symposium on Frequency Control*, 1988, pp. 239-251.

[17] T. E. Parker, J. A. Greer, and G. K. Montress, "SAW Oscillators with Low Vibration Sensitivity", in *Proceedings of the 45th Annual Symposium on Frequency Control*, 1991, pp. 321-329.

[18] D. Andres and T. E. Parker, "Designing Smaller SAW Oscillators for Low Vibration Sensitivity", to be published in *Proceedings of the IEEE Frequency Control Symposium*, 1994.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

IMPEDANCE ELEMENT SAW FILTERS

James Heighway, Sergei N. Kondratyev, and Victor P. Plessky

Advanced SAW products SA, CH-2022 Bevaix, Switzerland

1. ABSTRACT

*Over the last few years a new generation of Surface Acoustic Wave (SAW) devices with low loss and operating at high frequency have been invented to meet the demand of mobile communication systems. Besides outstanding electrical characteristics, these filters need to have small size and weight, must be highly reproducible, need no tuning, must be very cheap and if necessary, able to handle power levels up to a few Watts. In this work we introduce a new class of SAW filters - **Balanced Bridge Impedance Element Filters**, - which have all these advantages: low insertion loss, extremely small size and weight, a wide variety of centre frequencies and passbands, excellent suppression in the stopband and require no external matching circuits .*

2. BASIC CONCEPTS

2.1 Impedance characteristics of one port SAW devices

The impedance characteristics of different one port SAW devices (long interdigital transducer, single port resonators) have a resonance type of behavior - at the resonance frequency the admittance is high and at anti - resonance it is low. Figure 1 illustrates this kind of frequency dependence for a very long interdigital transducer of leaky waves on 64-LiNbO3 surface. The curves look very similar to that of classic bulk wave resonator. For a synchronous SAW resonator the admittance curves (see Fig. 2) are similar , but oscillations due to the finite length of the reflectors slightly complicate what is basically the same resonance-anti-resonance (R-a-R) behavior.

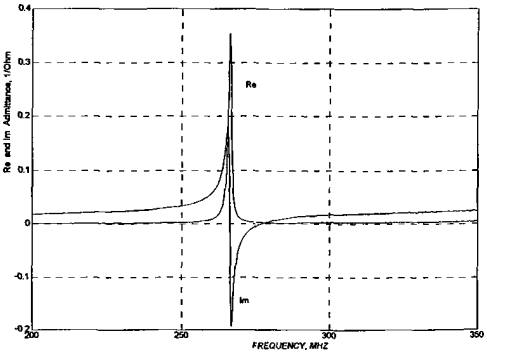


***Fig. 1** An experimental admittance frequency dependence for a long transducer (Nt=195 pairs of fingers, period λ = 16.42 μm, aperture W = 7.31λ) of leaky waves on 64-LiNbO3.*

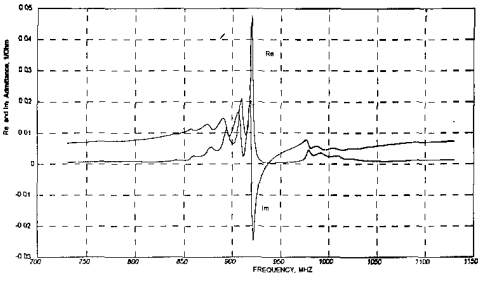


***Fig. 2** Admittance of a synchronous resonator*

*(This synchronous resonator test structure had 8 pairs of fingers in the interdigital transducer (IDT) and 42 electrodes in each of the reflectors placed on both sides of the IDT. The period of fingers was p=2.353 μm, the aperture W=51*2*p in both the transducer and the reflector gratings and there was no break in periodicity of the finger structure. The substrate was 64-LiNbO3 leaky wave cut, the thickness of Al electrodes being h/2p ≈ 3.7% .)*

Figure 3 shows the absolute value of the impedance dependence on frequency for a long interdigital transducer on a 64-LiNbO3 substrate (simulated

0-7803-1945-1/94 $3.00 © 1994 IEEE

curve, the structure contains 100 pairs of electrodes with period p≈2.35μ and aperture W=65μ). The frequency is in MHz, the impedance magnitudes in Ohms. The frequency dependence of the impedance is similar to that of classic L-C resonators with low resistivity at the resonance frequency and orders of magnitude higher resistivity at the anti-resonance frequency. The ratio of the max. to min. impedance can be increased by adding reflectors at both sides of the IDT, that is by creating a one port synchronous resonator. The solid curve of Fig. 3 illustrates this case. The resonator contains 70 pairs of electrodes in an IDT (p≈2.26 μ, and short reflectors on both sides of 16 shorted fingers). Both curves are simulated using the COM approach [1] . Generally speaking, this type of one port SAW device can be treated as ***a SAW impedance element*** **(IE)**, the frequency characteristic of which can be manipulated over a wide range using different substrate materials, types of waves, or designs of the IE itself.

For a *long transducer* type of device, the resonance - anti-resonance (R-a-R) frequency gap , important for IE filters design , is determined basically by the $\Delta V/V$ parameter and equals *0.91* (ΔV/V)* for single electrode transducer with 50% metallization ratio, the resonance being shifted to lower frequencies (to the left edge of the stopband) , the relative frequency shift being equal to $(-\kappa/2\pi)$, where κ if a reflection coefficient per wavelength.

2.2 A simple example of IE filter

The main idea of this work is to use these impedance elements as a building block for filters designed using classic network theory. Some lattice type filters have

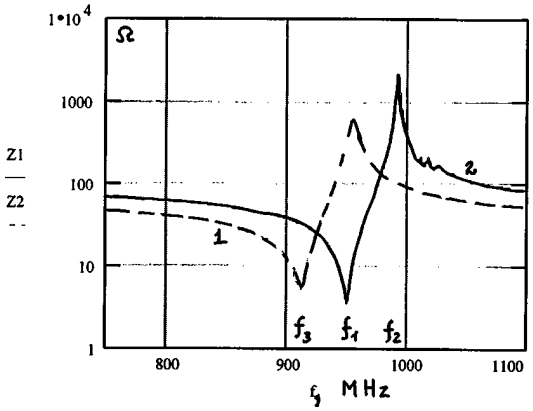


Fig. 3 Impedance of long IDT (dotted curve) and synchronous resonator (solid curve) on 64-LiNbO3 substrate.

been developed by Japanese authors [2]-[3] based on this approach, but the full potential of this idea and some significant improvements which can occur in SAW device design have not yet been wholly understood. At some stage in all SAW devices , a signal is transformed from electrical into acoustic form and later reconverted. This fundamental feature results in several well known difficulties: loss, low power handling ability, and sensitivity to the accuracy of positioning of electrodes. All SAW device design is a struggle to overcome these difficulties.

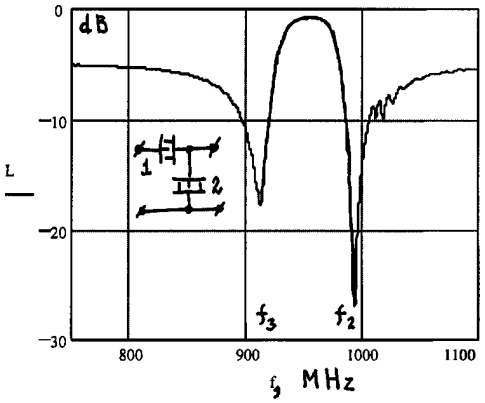


Fig. 4 Two impedance elements Γ-type filter.

With IE designs these difficulties can be overcome easily. Imagine that we have a simple Γ - type scheme including two IE's : the first (#1, solid curve on Fig. 3) in series between the input and the output (50/50 ohm) and the second (#2, dotted curve, Fig. 3) in parallel with the output. The first IE has low impedance , say 3 to 5 Ohms, at frequency $f1 \approx 950$ *MHz* and high impedance , more than 1000 Ohms at a higher frequency $f2 = f1 + \Delta f \approx 990$ *MHz* (numbers which can be realized without any difficulty). The parallel connected element can be chosen to have similar characteristics but shifted to a lower frequency: min. impedance at $f3 = f1 - \Delta f$ and maximum at $f = f1$ (dotted curve, Fig. 3) .

It is obvious that this device will work as a passband filter (See Fig. 4, simulated filter performance) with a passband around frequency $f1$ and deep notches at $f2$, $f3$. The low loss at frequency $f1$ is due to the fact that the series connected element has, at this frequency, a very low impedance, while the parallel connected one has at this frequency a very high impedance.

The out-of-band suppression of the filter is determined by the static capacitance ratio of the IDTs. In practice this out-of-band suppression is not adequate in the one section case discussed above and a few sections must be connected in series.

Note that there is no complete electric ↔ acoustic signal energy transformations. Only a small part of the voltage is applied to IE number 1 at the passband frequency. Since the distribution of voltages is approximately proportional to the impedances, practically all the input voltage in the passband frequency range is transferred to the output. Thus the device can handle high power levels with low insertion loss.

3. *LADDER TYPE IE-FILTERS*

Figures 5a and 5b show the simulated and measured results for 2- and 4-section ladder type filters wherein the elementary section of two IE's is exactly the same as in an above example.

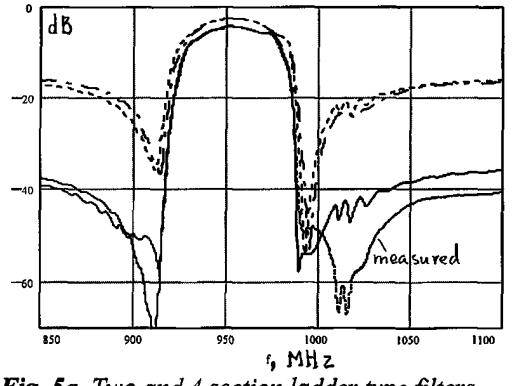


Fig. 5a *Two and 4 section ladder type filters*

It can be seen that the simulation gives excellent results in the passband and is quite accurate up to the 50 dB level where the feedthrough must be taken into account. These particular comparisons do not show the best achievable results and only serve to illustrate the principle. The insertion loss can be reduced to a value about 1 dB for 2-section devices and 2 dB for a 4-section ladder type filter.

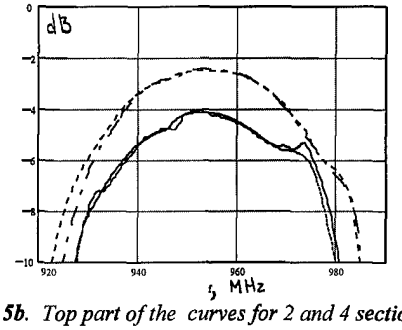


Fig. 5b. *Top part of the curves for 2 and 4 section filter.*

4. *BALANCED BRIDGE FILTERS*

The main drawback of ladder type IE filters is the relatively poor suppression in the stopband. At the stopband frequencies the IDTs do not excite the waves and act as simple capacitors. Part of the voltage from the input is inevitably transferred to the output through the ladder scheme of capacitors. This output signal level is basically determined by the capacitance ratios in the parallel and series arms of the scheme and is almost independent of frequency.

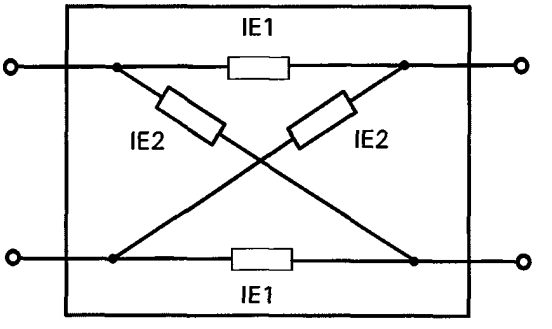


Fig. 6 *A balanced bridge network*

We propose a new ***balanced bridge SAW impedance element filter*** in which this static capacitance is compensated. This filter comprises a first pair and second pair of SAW impedance elements electrically coupled to form a bridge circuit, (see Fig. 6) wherein the centre frequency of each SAW impedance element of the first pair is different to the centre frequency of each SAW impedance element of the second pair , and the product of the static capacitances of the first pair is substantially equivalent to the product of the static capacitances of the second pair. In particular , all static capacitances can be made equal. The advantage of this bridge filter is that the voltages transmitted from the input to the output due to static capacitances are canceled, thereby reducing out of band transmission. For frequencies of an input signal in the passband range, the admittances of the IE in each of the before mentioned pairs are different (due to SAW generation and due to the shift in electrode structure periods), and thus the input signal sees a different admittance sequence in each arm of the bridge. The bridge is strongly non balanced and the signals pass from the input to the output through the low resistivity IE's. The frequency shift can be chosen so that the resonance frequencies of one pair of IE's (where the impedance is low) approximately coincides with the anti-resonance of the second pair of impedance elements (See Fig. 7) . The impedance of these second pair of IE's is high and they do not have important influence in the passband. But in the stopband

all impedances become progressively identical , the bridge scheme balanced and the output signals suppressed.

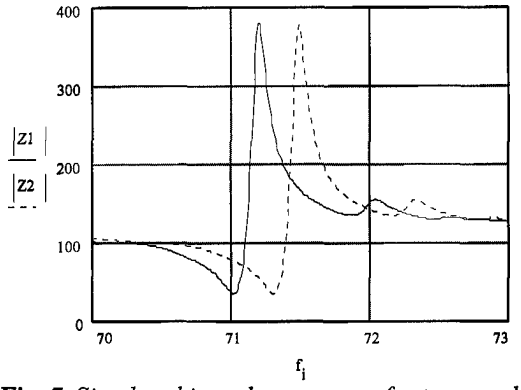


Fig. 7 Simulated impedance curves for two synchronous resonators on 112-LiTaO3 (frequency in MHz, impedance in Ohms).

Fig. 8 shows an example of a filter performance designed using this approach with IE's the same as on Fig. 7.

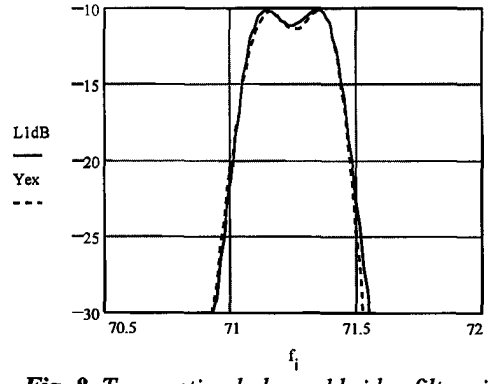


Fig. 8 Two section balanced bridge filter simulated (solid line) and measured (dotted line) frequency characteristics

Each resonator contained 90 pairs of electrodes in the IDT, 46 reflecting electrodes , the periods were approximately 23.0 and 22.9 microns, the apertures equal to 400 μ in both cases.

In the passband frequency range the first pair of IE practically connects the input to the output with minimal losses, while the signal cannot pass through the second pair of IE's which have high resistivity.

The frequency shift between the arms determines the passband of the filter and to provide low loss, must be chosen comparable to the electro-mechanical coupling coefficient of the substrate. Two or more sections can be connected in series to get better stopband suppression.

From a practical design point of view this new type of balanced bridge IE filter has a number of important advantages. The impedance elements are just simple long transducers (possibly with synchronous reflectors) having relatively large number of electrodes. The performance of these elements can be accurately modeled using existing models. The separate IE's are not interacting acoustically, they can be placed independently on the substrate. The large number of electrodes makes the resistivity of the electrodes relatively unimportant and parasitic capacitances are normally much smaller than the static capacitances of the transducers. The fact that SAW's do not have to be generated and received makes the overall length of the device shorter than in case of conventional SAW filters of the same passband.

Using different types of surface waves (Rayleigh waves, leaky SAW, STW) and different substrate materials, balanced bridge filters with passband ranges of 0.05% - 5% have been realized. At present, balanced bridge IEF's have been operated at frequencies up to 1.5 GHz [4]. Fig. 9 illustrates the characteristics of a filter for this frequency range.

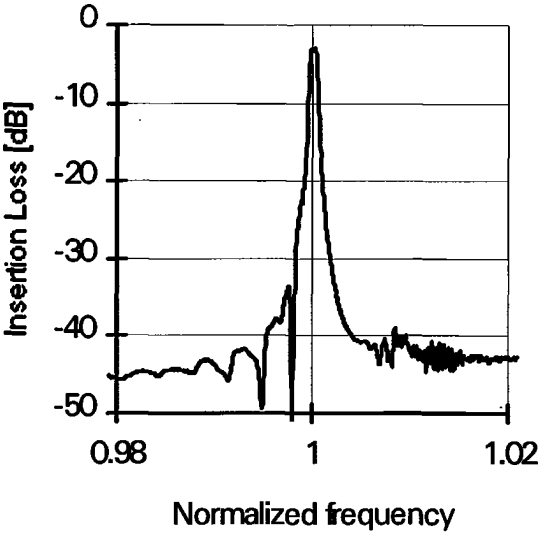


Fig. 9 Measured frequency response of a balanced bridge STW filter

In this one section 1.5 GHz balanced bridge device the insertion loss was about 3 dB, the passband close to 1 MHz, sidelobe suppression better than 40 dB. It is clear that device performances can be superior to conventional SAW device characteristics.

5. CONCLUSIONS

A new class of SAW devices has been proposed - *balanced bridge SAW impedance element filters* - which have significant advantages compared both with conventional SAW filters and with recently developed ladder type SAW impedance element filters.

6. ACKNOWLEDGMENTS

The authors are grateful to C. S. Hartmann, Prof. D. P. Chen, Dr. M. A. Sharif, Dr. T. Thorvaldsson , and P. Durrant for fruitful discussions.

7. REFERENCES

[1] V. P. Plessky, C. S. Hartmann, " Characteristics of leaky SAWs on 36-LiTaO3 in periodic structures of heavy electrodes", 1993 Ultrasonics Symposium, pp. 1239-1242.

[2] M. Hikita, N. Shibagaki, T. Akagi, K. Sakiyama, "Design methodology and synthesis techniques for ladder-type SAW resonator coupled filters", 1993 Ultrasonics Symp., pp. 15-24.

[3] O. Ikata, T. Miyashita, T. Matsuda, T. Nishihara, and Y. Satoh , "Development of low-loss band-pass filters using resonators for portable telephones", 1992 Ultrasonics Symposium, pp. 111-114

[4] V. P. Plessky, T. Thorvaldsson, "GHz rang STW resonators and narrow band filters", 1994 EFTF, Munchen, to be published.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SURFACE TRANSVERSE WAVE OSCILLATORS WITH EXTREMELY LOW THERMAL NOISE FLOORS

IVAN D. AVRAMOV
Institute of Solid State Physics
1784 Sofia, Bulgaria

F.L. WALLS, THOMAS E. PARKER
Time and Frequency Division
National Institute of Standards and Technology
325 Broadway
Boulder, Colorado 80303 U.S.A.

GARY K. MONTRESS
Raytheon Research Division
Lexington, MA 02173 U.S.A.

Abstract

This paper presents state-of-the-art results on 1 GHz surface transverse wave (STW) power oscillators running at extremely high loop power levels. High-Q single-mode STW resonators used in these designs have an insertion loss of 3.6 dB, an unloaded Q of 8000, a residual phase noise of -142 dBc/Hz at 1 Hz intercept and operate at an incident power of up to 31 dBm in the loop. Other low-Q STW resonators and coupled resonator filters (CRF) with an insertion loss in the 5-9 dB range can conveniently handle power levels in excess of 2 W. These devices were implemented in voltage controlled oscillators (VCO's) running from a 9.6 V source at an output power of 23 dBm and a RF/dc efficiency of 28%. Their tuning range was 750 kHz and the noise floor -180 dBc/Hz. The oscillators, stabilized with the high-Q devices, use specially designed AB-class power amplifiers, deliver an output power of 29 dBm and demonstrate a noise floor of -184 dBc/Hz and a 1 Hz intercept of -17 dBc/Hz. The 1 Hz intercept was improved to -33 dBc/Hz using the UTO-1023 as a loop amplifier. In this case the output power was 22 dBm. In all cases the loop amplifier was the limiting factor for the close-to-carrier oscillator phase noise performance.

1. Introduction

Surface acoustic wave (SAW) based oscillators are well known for their excellent phase noise performance in the frequency

0-7803-1945-1/94 $3.00 © 1994 IEEE

range 0.1-1 GHz [1] - [3]. Resonator stabilized oscillators, operating in the 400-500 MHz range and featuring a thermal noise floor of -184 to -185 dBc/Hz were demonstrated recently [3]. Along with a 1 Hz intercept of -48 to -55 dBc/Hz these oscillators are considered to represent the state-of-the-art phase noise performance in this frequency range. One limiting factor to further improvement of the oscillator noise floor is the power handling capability of the SAW device which limits the maximum drive power level to about 26 dBm even if large area multi-track designs are used. An example was presented in [4] where a very high power SAW resonator failed after 200 weeks of operation at a power dissipation level of 130 mW.

This limit can be extended if the acoustic resonator uses the STW mode. As shown in [5] and experimentally verified in [6] devices with relatively small acoustic area can conveniently handle orders of magnitude higher drive power levels without degradation in performance. About a year ago this unique feature was used to design a 1 GHz low noise voltage controlled oscillator (VCO) which was running at a loop power of 34 dBm and demonstrated a noise floor of -194 dBc/Hz [7]. This oscillator featured excellent tuning and wide band frequency modulation abilities. Its RF/dc efficiency was 19%.

We present results from a one year research effort on improved STW power oscillators. Different STW resonant devices and 1 GHz fixed frequency and voltage controlled oscillators were characterized. The phase noise performance of the STW devices and oscillators, measured with different methods, are presented and discussed.

2. STW Resonant Devices for Power Oscillator Applications

Compared to SAW, STW offer a greater flexibility in the design of metal strip resonators and narrow band filters for oscillator applications. This increased design flexibility comes from the fact that metallization allows an additional degree of freedom in controlling the resonant Q while keeping low device insertion loss even in simple resonator configurations [8], [9]. This unique feature has been used extensively in the design of different kinds of single and multimode resonators and coupled resonator filters with a loaded Q ranging from 500 to 8000 and an insertion loss well below 10 dB at 1 GHz. All these devices can conveniently stand drive power levels in excess of 2 W and were found to operate without measurable performance degradation for several months in different fixed frequency and voltage controlled power oscillators. The design details for such devices have been well documented in references [5], [6], [10] and [11]. Here we will characterize only some of the devices used in this study.

Figures 1 A), B), C) and D) present data on a 1 GHz single-mode high-Q resonator which was designed at the Institute of Solid State Physics in Sofia, Bulgaria and fabricated with all quartz technology [12].

This device has an insertion loss of 3.6 dB, a loaded Q of 2740 and an unloaded Q of 8000. It was intended for use in a fixed frequency power oscillator. However, if run at a loaded Q of about 3000 it would allow a tuning range of about 180 KHz over the 1 dB device bandwidth as evident from Fig. 1 B).

A much wider tuning range can be achieved with the low-Q resonator characterized in Fig. 2. It has an insertion

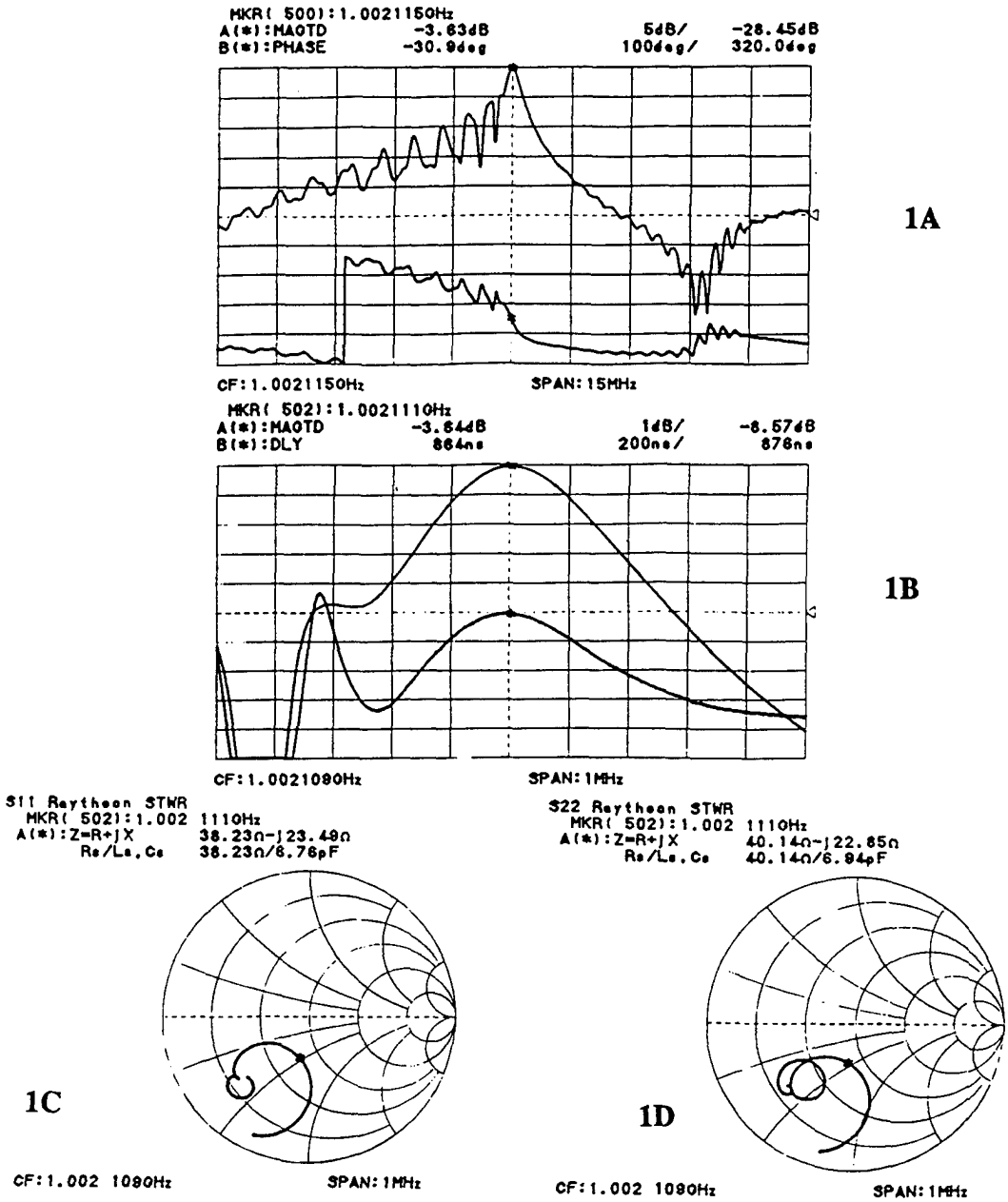


Fig. 1. Characteristics of a single mode high-Q device realized at Raytheon:
A) Frequency and phase responses,
B) Detailed frequency and group delay responses,
C) Input reflection coefficient, S11,
D) Output reflection coefficient, S22.

loss of 5.2 dB, a loaded Q of 1500 and allows a tuning range of about 700 KHz if the VCO is tuned over the 3 dB device bandwidth.

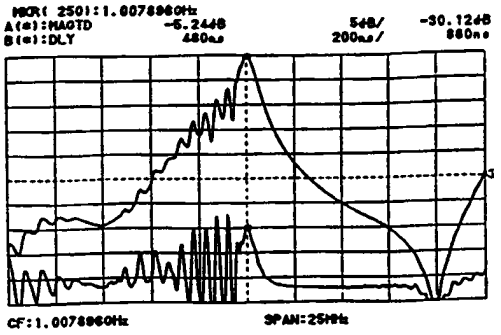


Fig. 2. Frequency and group delay responses of an STW resonator with Q_l = 1500 and 5.2 dB insertion loss.

If even wider tuning ranges are necessary, the 2-pole coupled resonator filter characterized in Fig. 3, can be used. This device has an insertion loss of 8.5 dB and a 1 dB bandwidth of 1.5 MHz. As evident from its phase response, a variable phase shift of 0-180° would be necessary for tuning over the entire 1 dB bandwidth. In this case cascading 3 C-L-C varactor tuned phase shifters would be necessary [4].

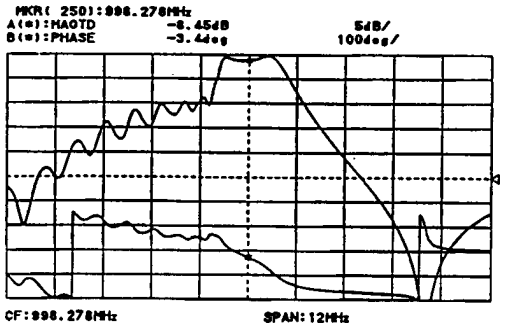


Fig. 3. Frequency and phase responses of a 2-pole coupled resonator filter.

All devices were fabricated in a standard single step photolithographic process with careful control over the metallization. The resolution required was about 1.3 μm which can readily be realized with almost all kinds of photolithographic equipment available to date.

2. STW High Loop Power Oscillator Designs

As demonstrated with the STW oscillator described in [7], a substantial improvement in the thermal noise floor can be achieved if the loop amplifier is capable of generating output power levels in excess of 2 W. At GHz frequencies these levels are very difficult to achieve with A-class amplifiers using bipolar transistors which are known for their low 1/f noise. Even if some power transistors can generate high output power in the A-class of operation and a 50 Ω environment, their efficiency rarely exceeds 10%. An elegant and inexpensive solution of the efficiency problem can be obtained if AB-, B- or C-class amplifiers are used. With one of the STW oscillators, using an AB-class loop amplifier [7], we were able to achieve an RF/dc efficiency of 36% with an oscillator output power of 28 dBm at 1 GHz. One major drawback of AB-class amplifiers is that they require careful reactive matching at their input and output in order to be able to work efficiently in a 50 Ω environment. This is because the input and output impedances of the power transistors in the B- or C-class of operation are up to an order of magnitude lower than 50 Ω. If such an amplifier is to be used in an STW power oscillator, the matching circuits have to be changed in order to achieve a low reflection coefficient at the

input and output of the STW device since its impedances also differ from 50 Ω (see Fig. 1 C and D). Under these circumstances measuring the oscillator loop power becomes a serious problem. Just breaking the loop and loading it on both sides with the 50 Ω impedances of the measurement system will not work because this will seriously deteriorate matching. Therefore the loop power has to be measured under closed loop conditions. We have solved this problem by means of the capacitive probe shown in Fig. 4. It consists of a piece of coaxial cable ending with a small series 0.47 pF capacitor soldered to the central line.

The ground skirt is split into two parts symmetrically bent on both sides of the cable in such a manner that the probe can conveniently touch any point of the loop strip lines and ground planes, surrounding them, as shown in Fig. 4. The reading is obtained by a spectrum analyzer or power meter connected to the other end of the cable. The probe is calibrated by touching the load at the oscillator output at which the power can be precisely measured with a power meter. This reading will give the attenuation of the probe (14 dB in our case). The loop power at any other point of the loop is obtained by adding the probe loss to the reading. We found that the probe did not deteriorate the matching conditions at the points of measurement. Only a slight frequency shift of up to 20 ppm was observed. This shift was well within the oscillator tuning range in which the output amplitude was constant.

We investigated three types of high loop power STW oscillators using highly efficient AB-class loop amplifiers. Since the operation principle of such oscillators was described in detail in [7], we present only the block circuits and loop level diagrams measured with the capacitive probe.

The simplest circuit which requires a minimum number of passive components and only one power transistor is shown in Fig. 5. It was designed to run with minimum loss around the loop.

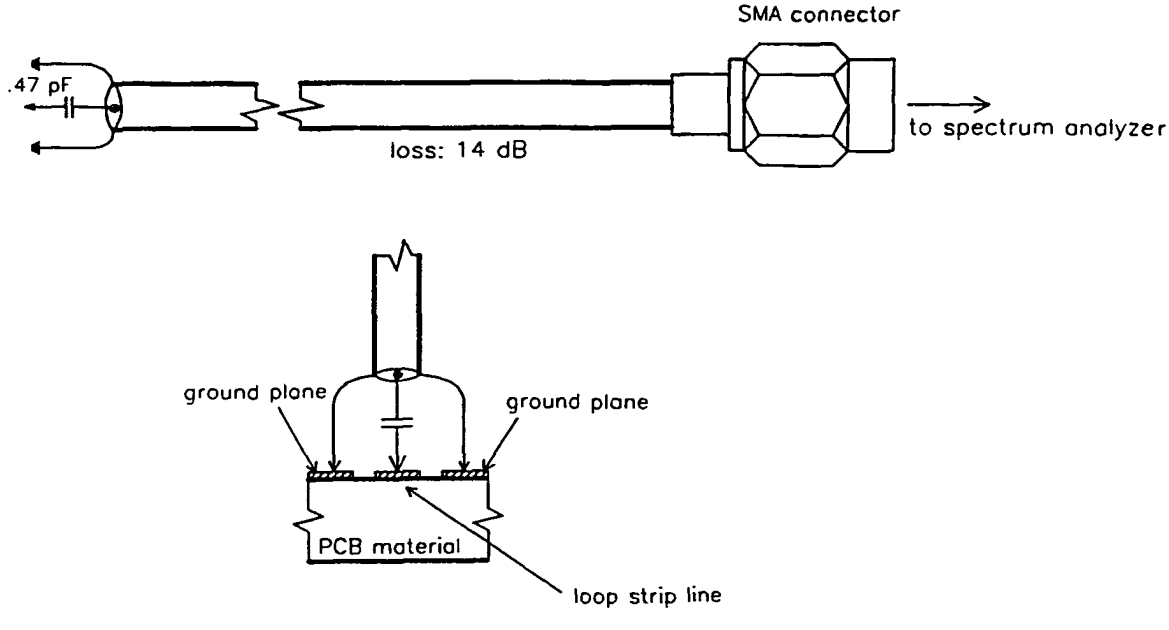


Fig. 4. A capacitive probe for evaluation of the oscillator loop power under closed loop conditions.

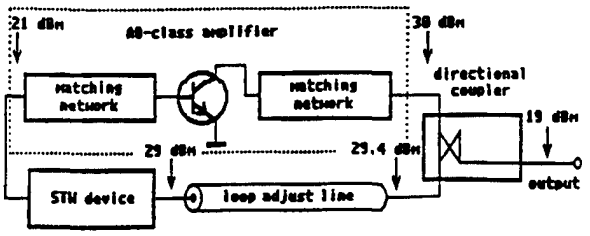


Fig. 5 Simple STW high loop power oscillator with directional coupling to the load.

This was achieved using a directional coupler instead of a 3 dB power splitter for load coupling. Thus, the loss (typically 3.5-4 dB) of a 3 dB power divider was reduced to 0.6 dB. With a loop power of 30 dBm and an output power of 19 dBm the oscillator was found to provide stable fixed frequency operation and was insensitive with respect to load changes. Unfortunately, we were unable to achieve usable frequency tuning with this design. The addition of a varactor tuned phase shifter in the loop was found to make the oscillator unstable with tuning due to a deterioration of the matching condition. Another drawback of this design was that the gain compression had to be kept very low (1-2 dB) which was necessary because an AB-class amplifier is a poor limiter. Increasing the loop power results in an increase of the collector current which can thermally overload the transistor.

The tuning and limiting problem could readily be solved with the circuit in Fig. 6. Here an A-class amplifier is incorporated between the variable phase shifter (VPS) and the AB-class power amplifier. Its function is two-fold. First, it provides sufficient isolation between the VPS and the power stage and second, it performs the limiting function providing a safe input power level to the power transistor. Thus excellent tuning over the 3 dB device bandwidth and

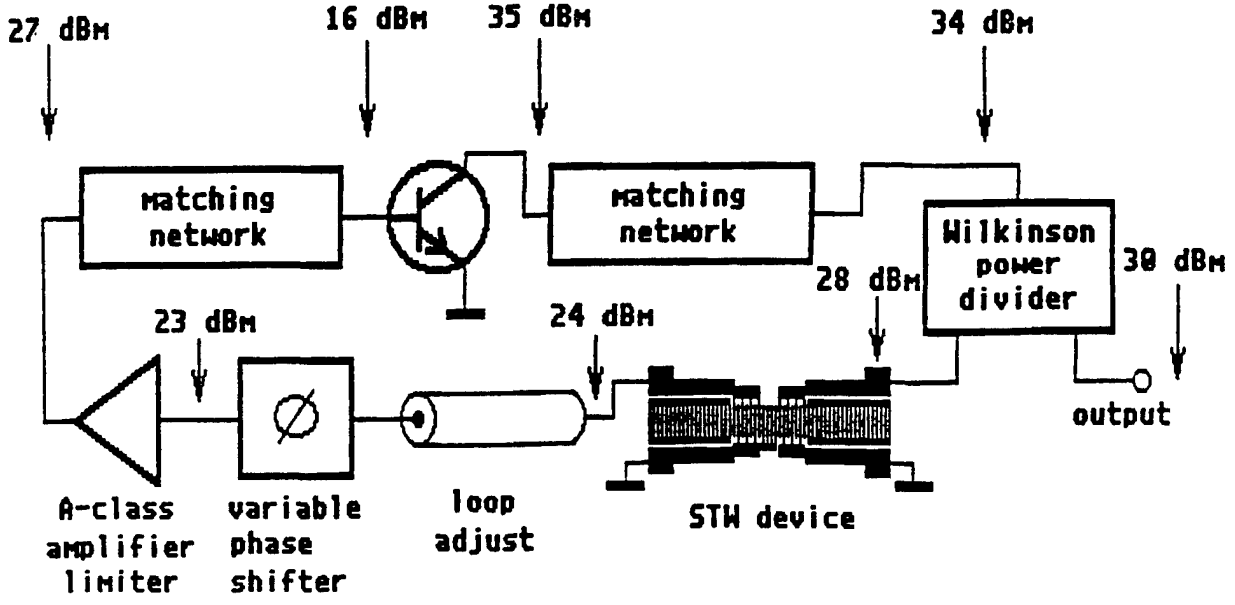


Fig. 6. Block and level diagram of a 1 W VCO stabilized with the high Q device from Fig. 1.

stable amplitude over the entire tuning range are guaranteed. However, one has to be careful when adjusting the gain compression. Too much gain compression will result in a deterioration of the overall phase noise performance as shown in Fig. 19 below. A trade-off between gain compression, loop loss, device Q and tuning range can be achieved by adding series capacitors to the STW device. In this case a readjustment of the matching circuits is necessary. Another 1-2 dB variation of the gain compression is also possible by unbalancing the Wilkinson power divider. This changes the output power accordingly.

The circuit in Fig. 6 was stabilized with the high-Q device from Fig. 1. The loop power was measured to be 35 dBm and the incident power on the STW device could be altered between 28 and 31 dBm by unbalancing the power divider. According to Leeson's model this oscillator should have a noise floor of -195 dBc/Hz [3].

Figure 7 shows a highly efficient power VCO which uses the same concept. Since a wide tuning range of 700 KHz was necessary, the oscillator was stabilized with the low-Q resonator from Fig. 2. It was designed to run from a 9.6 V rechargeable NiCd battery for portable applications. The output power is 23 dBm and the RF/dc efficiency is 28%. When run at a supply voltage of 16 V the output power increased to 28 dBm and the RF/dc efficiency decreased by only 3%.

Curves 8 A, B and C in Fig. 8 show the tuning characteristics of the power VCOs from fig.s 6 and 7. They all were obtained with a single C-L-C type VPS which was designed to deliver about 60° of variable phase shift [4]. Cascading two identical

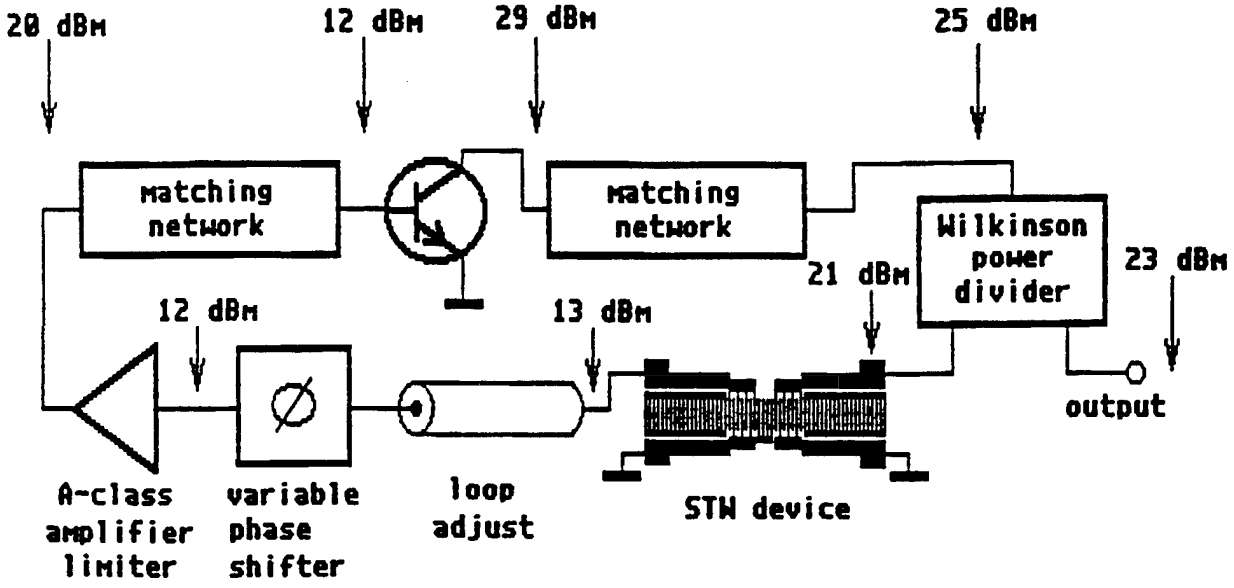


Fig. 7. 9.6 V VCO for portable applications with RF to dc efficiency of 28%.

phase shifters and using the CRF from Fig. 3 would increase the tuning range to 1.5 MHz with a tuning voltage of 0 to 9.5 V.

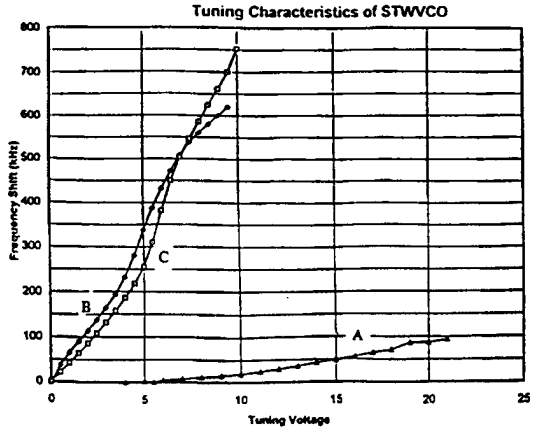


Fig. 8. Tuning characteristics of the power VCO:
A) 1W VCO using the high-Q device from Fig. 6,
B) 9.5 V wide tuning range VCO from Fig. 16 B,
C) Wide tuning range VCO at 16 V supply voltage from Fig. 16 C.

3. Phase noise measurements.

The major goal of this study was to evaluate the phase noise performance of STW power oscillators at 1 GHz since this is important to a variety of applications in this frequency range. Since it is difficult to obtain the overall oscillator phase noise data for Fourier frequencies ranging from 1 Hz to 10 MHz away from carrier with one single measurement, especially if two identical oscillators are not available, we set up different measurement systems and performed several measurements on the power oscillators to make sure that the systems delivered comparable results.

First we tried to measure the oscillator noise floor. The simplest and most forgiving system for noise floor evaluation is the single channel frequency discriminator with coaxial cable of delay τ_d shown in Fig. 9.

This system is not sufficiently sensitive for close to carrier measurements on stable oscillators but provides very good results for Fourier frequencies as high as 35% of the frequency at which the first null of its transfer function occurs $(f < \frac{0.35}{\tau_d})$ [13]. Moreover, it needs only one oscillator and adapts to small changes of the oscillator frequency during the measurement.

If the outputs of two identical channels of this system are cross correlated in a 2-channel FFT analyzer (Fig. 10) then the uncorrelated system noise cancels out and the system sensitivity can be improved by 20-25 dB [13] - [15].

This cross correlation concept can also be applied to a measurement system using two identical oscillators (Fig. 11). The system noise floor in this case can exceed -195 dBc/Hz even at very high Fourier frequencies [13] - [15].

This cross correlation concept can also be applied to a measurement system using two identical oscillators (Fig. 11). The system noise floor in this case can exceed -195 dBc/Hz even at very high Fourier frequencies [13] - [15].

Residual phase noise measurements on the high-Q STW device from Fig. 1 were

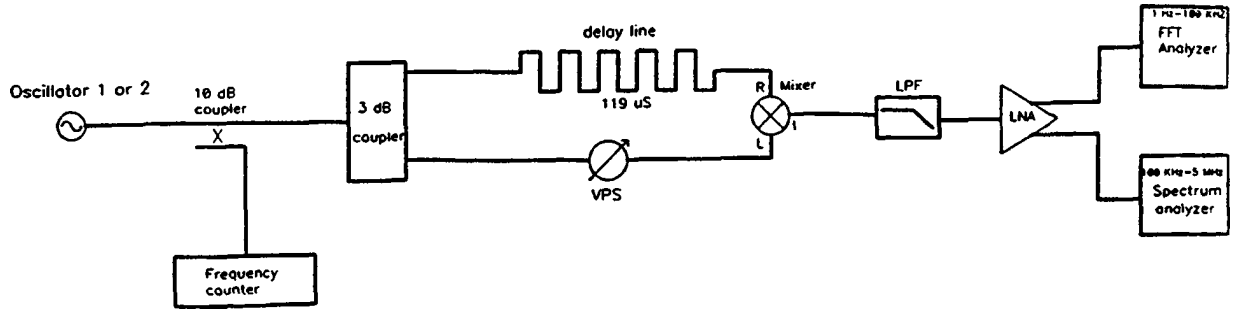


Fig. 9. Measurement setup using a single channel delay line frequency discriminator.

performed with the setup in Fig. 12, adapted from reference [4]. A low PM noise frequency synthesizer [16] was used to make phase noise measurements up to 1kHz from the carrier. To provide high enough mixer drive level and sufficient suppression of the synthesizer AM noise we amplified the source signal with two cascaded low PM noise amplifiers [17], the second of which was driven into 3 dB of gain compression.

Close-to-carrier phase noise measurements on very low-noise STW oscillators using the low PM noise amplifier [17] were performed with the system setup in Fig. 13. This setup assumes that the reference source (in this case a low-noise frequency synthesizer) is substantially quieter than the STW oscillator. In our case this condition breaks down for offsets greater than 1 KHz.

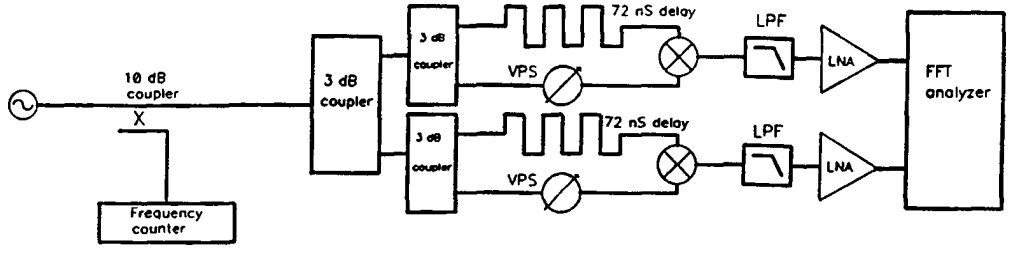


Fig. 10. Cross-correlation frequency discriminator measurement setup.

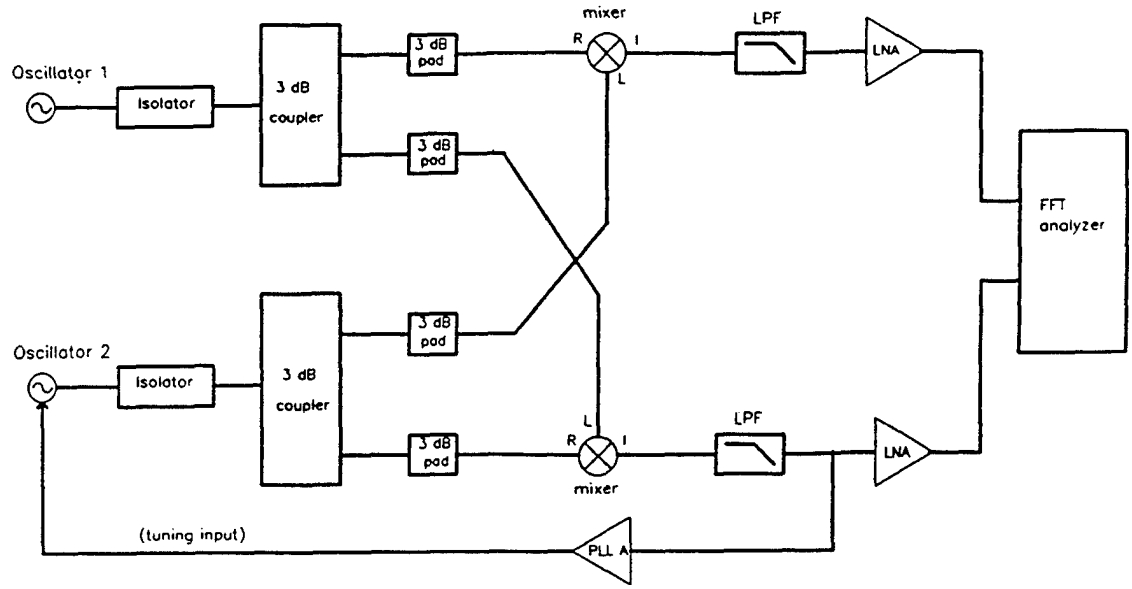


Fig. 11. Two oscillator cross-correlation measurement.

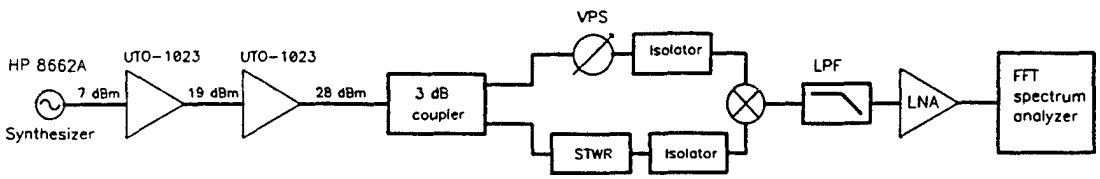


Fig. 12. Setup for residual phase noise measurements on high-Q STW devices.

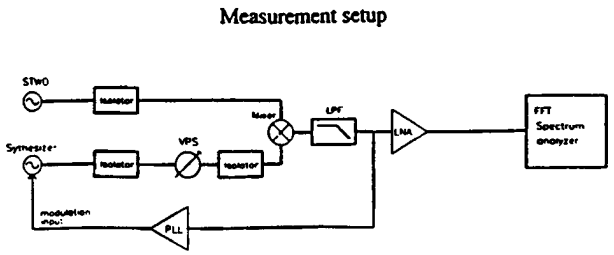


Fig. 13 System for measuring the close-to-carrier phase noise of fixed frequency STW oscillators.

4. Evaluation of the Phase Noise Data

Figure 14 shows the phase noise plot of two nearly identical power oscillators using the design in Fig. 6. The loaded Q of the STW devices was adjusted to a value of about 4000 which is half of the unloaded Q. The output power was in this case 29 dBm. A noise floor of -184 and -182 dBc/Hz was measured with the single channel frequency discriminator method for Oscillator 1 and Oscillator 2 respectively. Then both oscillators were measured against each other using the two oscillator cross correlation measurement configuration from Fig. 11. Assuming equal noise in each oscillator a noise floor of -181 dBc/Hz was obtained (Fig. 15).

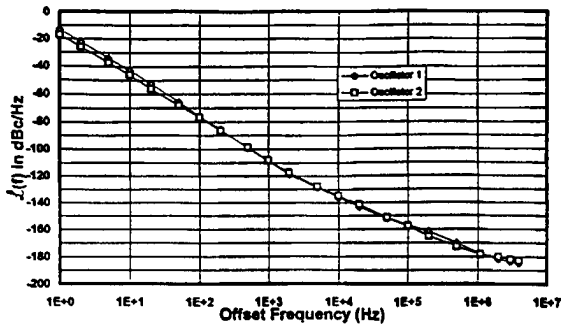


Fig. 14 Phase noise plots of two high-Q STW power oscillators measured with the single channel frequency discriminator method. Both oscillators have Q_l=4000, P_{out}=29 dBm.

Excellent results were obtained with the highly efficient wide tuning range VCO at supply voltages of 9.5 and 16 V (Curves 16 A and B in Figure 10). Noise floors of -180 and -185 dBc/Hz were obtained using the dual channel frequency discriminator cross correlation method (Fig. 10). The 1 Hz intercept points were measured as -4 and -1 dBc/Hz and the output power levels were 23 and 28 dBm at 9.5 and 16 V supply voltage respectively. The RF/dc efficiencies were accordingly 28% and 25% for both supply voltages.

A comparison of the phase noise plots for the high-Q and low-Q STW oscillators (Figs. 14 and 15 versus Fig.s 16 A and B) shows the trade-off between the tuning range (Curves A, B and C in Fig. 8) and the phase noise performance of the investigated STW power oscillators.

Figure 17 is the residual phase noise of one of the high-Q STW devices characterized in Fig. 1. The value of -142 dBc/Hz at 1 Hz intercept is a remarkable result for a 1 GHz metal strip device. The system floor, measured with a 4 dB pad instead of the STW device, was 2-4 dB lower than the residual noise of the STW device in the 1-100 Hz range. Both curves show a 10 dB/decade slope for offset frequencies below 100 Hz. The steeper slope below 2 Hz for the STW device and the 4 dB pad was caused by the measurement system. The actual STW device 1 Hz intercept is approximately -147 dBc/Hz.

This same STW device was implemented in an a simple oscillator loop using a power amplifier which is known for its very low residual noise [3], [17]. This test oscillator is shown in Fig. 18. First it was run at 7 dB gain compression for minimum loop loss. The output power was

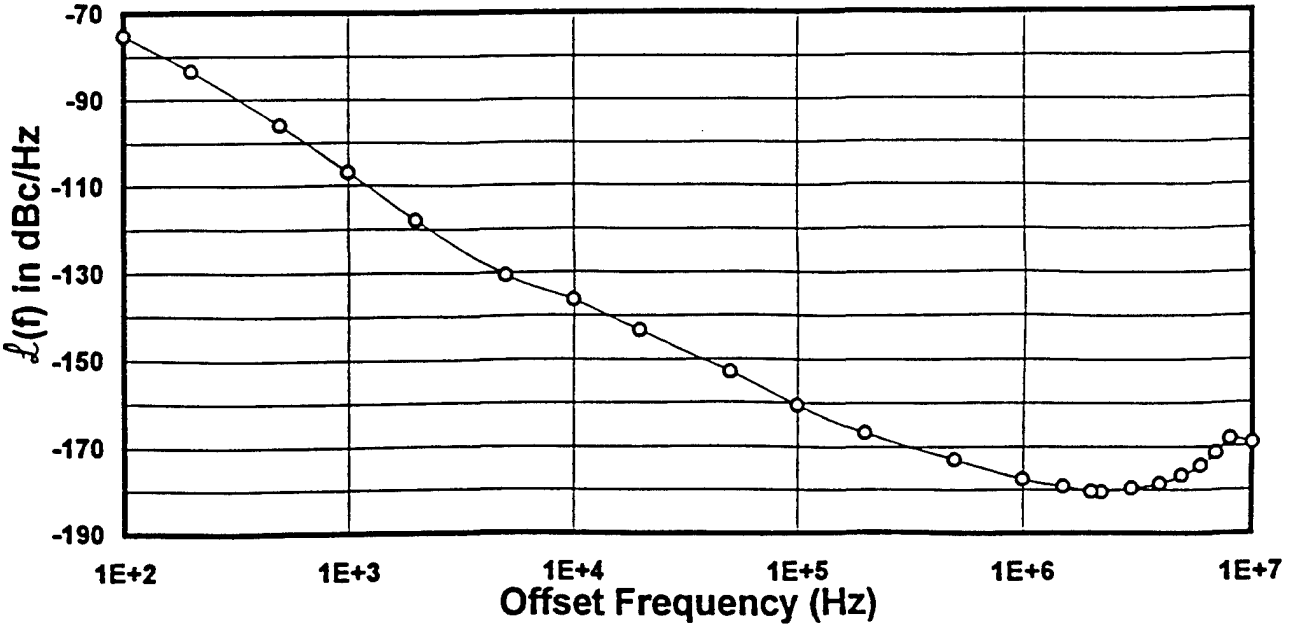


Fig. 15. Phase noise data of the high-Q oscillators obtained from the two-oscillator cross correlation measurement (Fig. 11) assuming equal noise.

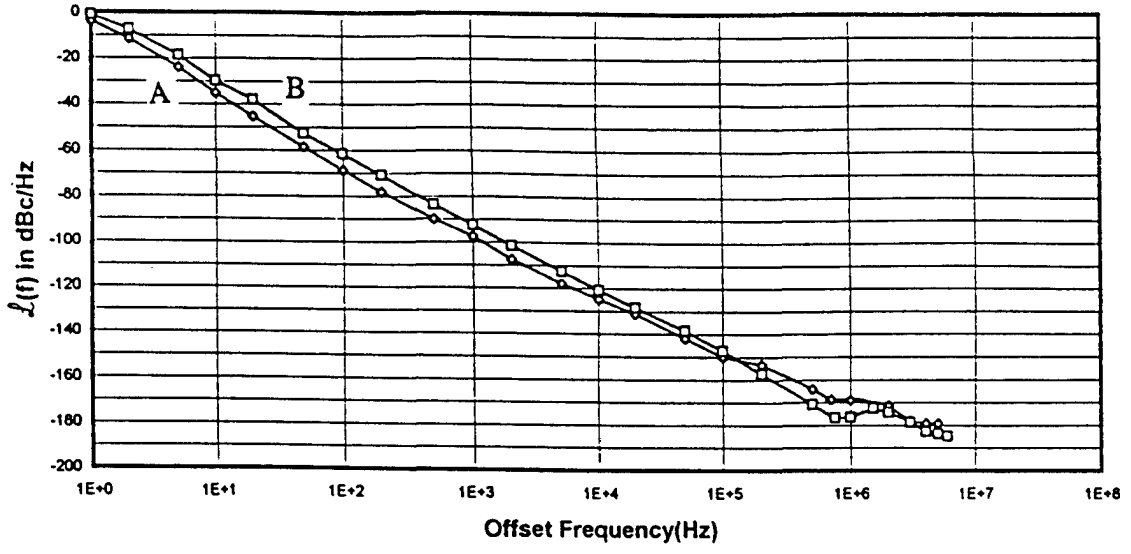


Fig. 16. Phase noise of the wide tuning range VCO at a supply voltage of:
A) 9.5 V (P_{out}=23 dBm, efficiency=28%) and
B) 16 V (P_{out}=28 dBm, efficiency=25%).

24 dBm. Figure 19 indicates a noise floor of -187 dBc/Hz for this case. The 1 Hz intercept of -21 dBc/Hz was 16 dB worse than what we expected from the residual noise measurement on the STW device (Fig. 17). Also the slope was -25 dB/decade

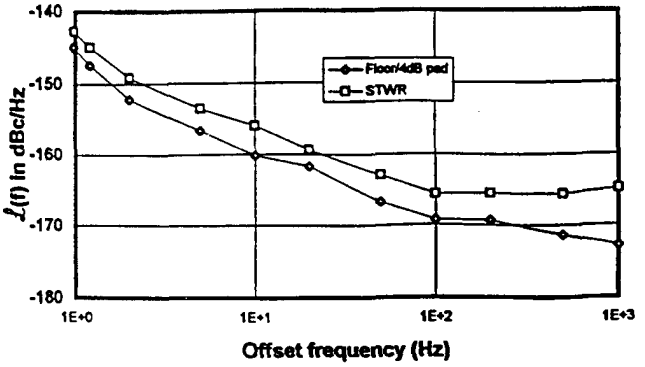


Fig. 17. Residual phase noise of one of the high-Q STW devices fabricated at Raytheon with Q_L=2737 and insertion loss of 3.6 dB.

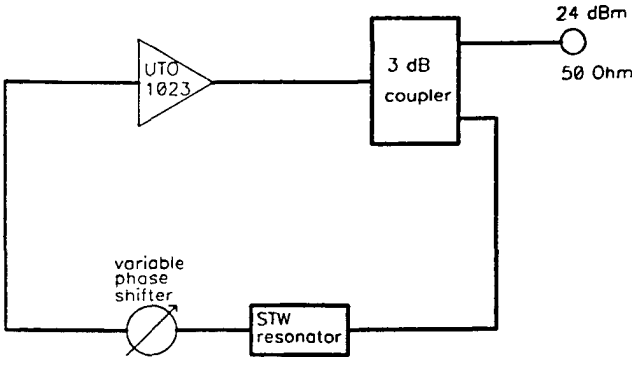


Fig. 18 Test oscillator using a low 1/f noise loop amplifier.

instead of 30 dB/decade. We found that the high gain compression was the reason for this poor performance. After we reduced the gain compression to 1 dB we obtained the results in Fig. 20. The 1 Hz intercept was -33 dBc/Hz, only 4 dB worse than what we would expect if the loop amplifier was ideally noiseless. A measurement with a second low PM amplifier from the same manufacturer indicated a 1 Hz intercept of -31 dBc/Hz with the same STW device [17].

5. Discussion.

The oscillators in Fig. 6 were designed for a noise floor of -195 dBc/Hz. We measured 11-14 dB less. The reason is that these circuits were designed for an STW device loss of 7-8 dB. The devices we used had a loss of about 3.5 dB. We increased this loss by increasing the loaded Q but this resulted in deterioration of the matching conditions and the amplifier noise figure. We believe that this problem can be overcome by more careful amplifier design which should start after one knows the impedances and insertion loss of the STW devices to be used.

The noise floor of the single transistor stage oscillator (Fig. 5) could not be measured correctly because the output power was insufficient for the measurement systems used and there were no identical oscillators available. We believe that this concept will also yield very low noise floors because it allows high loop power and extremely low loop loss.

We expected better than -181 dBc/Hz from the two oscillator cross correlation measurement (Fig. 5). Unfortunately, strong interference with local AM radio stations caused the bump around 10 MHz and affected the measurement. This bump was not observed with the single channel frequency discriminator measurement on the same oscillators (see Fig. 9 and 14). This is probably due to the fact that the single channel system was much simpler, easier to set up and less sensitive to parasitic interference.

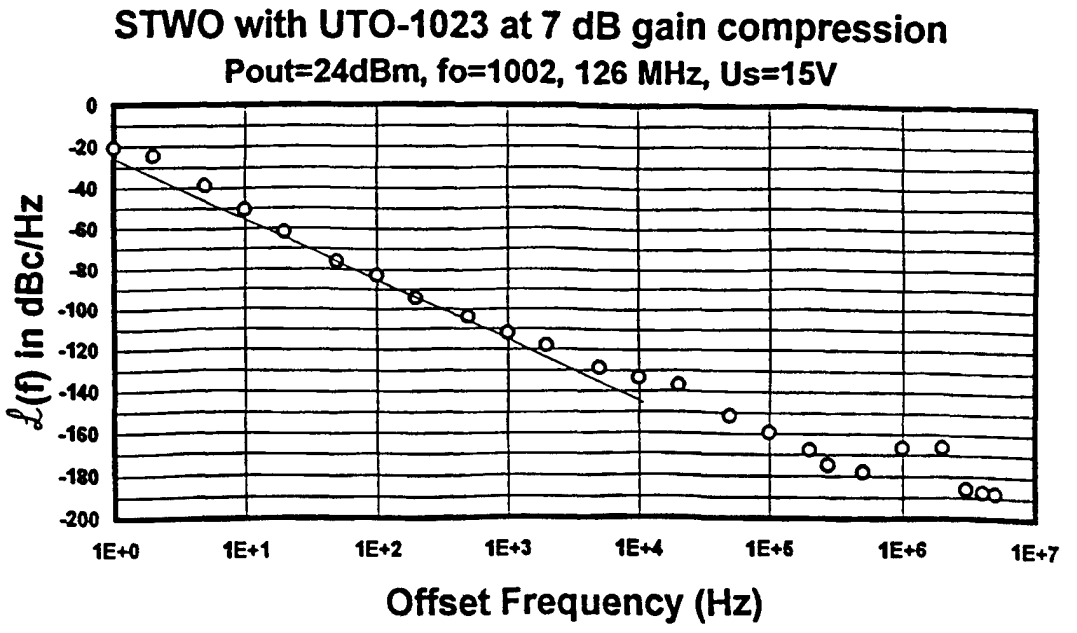
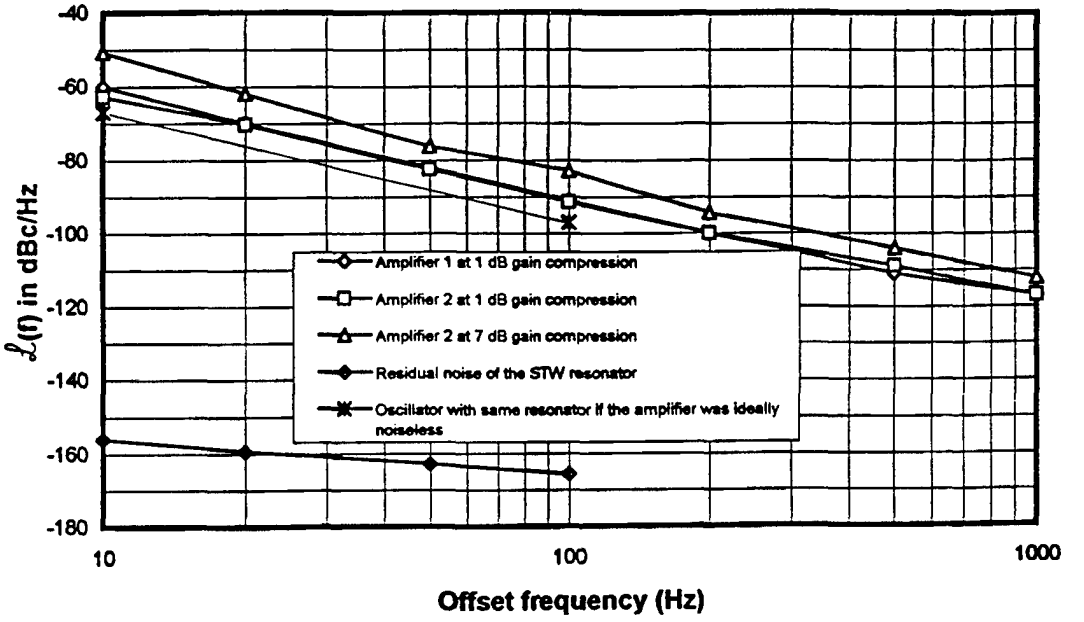


Fig. 19. Phase noise of the test oscillator of Fig. 18 at a gain compression of 7 dB, P_{out}=24 dBm and P_{loop}=28 dBm.

6. Summary and conclusions.

We have demonstrated state-of-the-art 1 GHz STW resonators for power oscillator applications featuring an insertion loss of 3.6 dB. a loaded Q of 2740, an unloaded Q of 8000 and a residual noise level of -142 dBc/Hz at 1 Hz intercept. Other low-Q resonators and two-pole coupled resonator filters with an insertion loss of 5-9 dB allow



Fig

Fig. 20. Phase noise data of STW oscillators and the residual noise of the STW device with low noise loop amplifier [18].

tuning ranges of 700-1500 ppm in highly efficient STW power oscillators which can run from supply and tuning voltage sources below 10 V and are suitable for portable applications. The output power and the noise floor of such VCO are typically 23 dBm and -180 dBc/Hz respectively. Higher output powers in the 28-33 dBm range and lower noise floors in the -185 to -194 dBc/Hz range can be achieved if higher supply voltages are used (see also Reference [7]). The RF/dc efficiency is typically 25-28% but a value of 36% was also observed.

1 W oscillators running at a loop power of up to 35 dBm feature a noise floor of -184 dBc/Hz and a 1 Hz intercept of -17 dBm. If a low 1/f noise loop amplifier is used the noise floor can be reduced to -187 dBc/Hz [17]. Such oscillators typically run at a loop power of 28 dBm and an output power of 24 dBm and demonstrate a 1 Hz intercept of -21 dBc/Hz. A reduction of the gain compression to 1 dB greatly improves the close-to-carrier phase noise behavior resulting in a 1 Hz intercept of -33 dBc/Hz and a slightly decreased output power of 22 dBm.

The results in Fig. 20 represent the state-of-the-art close-to-carrier phase noise performance of 1 GHz STW power oscillators. They clearly indicate that the loop amplifier and not the STW device is the major source of 1/f noise even if the best loop amplifiers currently available in the market are used. Therefore, further care has to be taken in designing high power loop amplifiers with improved residual phase noise performance.

We believe that STW resonant devices have a strong potential for use in designing extremely low noise power oscillators in the lower GHz range. Increasing the active acoustic area could improve the device residual phase noise to values lower than -150 dBc/Hz at 1 Hz intercept. Using careful loop amplifier design and taking special care of low gain compression, matching and noise figure will result in noise floor values below -195 dBc/Hz and a 1 Hz intercept of -45 dBc/Hz in the lower GHz range.

Acknowledgements

The authors gratefully acknowledge the Institute of Semiconductor Physics in Novosibirsk, Russia for the opportunity to have the low-Q STW devices fabricated at those laboratories. We thank also Alan Doak for his help in preparing the figures for this paper.

This study was in part supported by the National Research Fund of Bulgaria under contract No. F212 of September 30, 1992.

References

[1]. T. E. Parker, G. K. Montress, "Precision SAW Oscillators", IEEE Transactions on Ultrason., Ferroelec. and Freq. Contr., Vol. (UFFC-35), No. 3, May 1988, pp. 342-364.

[2]. G. K. Montress, T. E. Parker, D. W. Kress and J. Kosinski, "Design and Performance of a Low Noise, Wide Tuning Range AQP SAW Delay Line VCO", in Proc. IEEE 1992 Frequency Control Symposium, pp. 356-370.

[3]. G. K. Montress, T. E. Parker, "Design and Performance of an Extremely Low Noise Surface Acoustic Wave Oscillator", in Proc. IEEE 1994 Int. Freq. Contr. Symp.

[4]. T. E. Parker, G. K. Montress, "Introduction to High Stability SAW Oscillators: Design and Performance" Tutorial presented at the 1994 IEEE Int.

Freq. Contr. Symp.

[5]. T. L. Bagwell, R. C. Bray, "Novel Surface Transverse Wave Resonators with Low Loss and High Q", IEEE 1987 Proc. Ultrason. Symp., pp. 319-324.

[6]. I. D. Avramov, "1 GHz Voltage Controlled Oscillator Using a SSBW[BGW Combined Mode Resonator Filter", Proc. 45th Annu. Symp. on Freq. Contr., 1991, pp. 230-238.

[7]. I. D. Avramov, "Extremely High Loop Power GHz Range Surface Transverse Wave Oscillators Using AB-class Amplifiers", in Proc. 1993 IEEE Int.Freq. Contr. Symp., pp. 728-732.

[8]. I. D. Avramov, "High Q Metal Strip Resonators Using a SAW Design", IEEE Trans. on Ultrason., Ferroelec. and Freq. Contr., Vol. (UFFC-37), No. 6, Nov. 1990, pp. 530-534.

[9]. I. D. Avramov, D. F. Thompson, "Influence of the Metallization on the Electrical Performance of STW Two-port Resonators, Proc. 6th Conf. "Acoustoelectronics'93", Sept. 19- 25, Varna, Bulgaria, pp. 193-198.

[10]. I. D. Avramov, A. S. Kozlov, "Design of Combined Mode Resonator Filters Based on Surface Transverse and Surface Skimming Bulk Waves", in Proc. 10th Eur. Conf. on Circuit Theory and Design, ECCTD-91, Sept. 3-6, 1991, Lyngby, Denmark, pp. 729-738.

[11]. I. D. Avramov, "GHz Range Resonant Devices for Oscillator Applications Using Shear Horizontal Acoustic Waves", IEEE Trans. Ultrason., Ferroelec. and Freq. Contr., Vol. (UFFC-40), No. 5, Sept. 1993, pp. 459-468.

[12]. STW devices fabricated at Raytheon Co. in an all Quartz Package.

[13]. Algie Lance, Wendel Seal, and Fredrick Labaar, "Phase Noise and AM Noise Measurements in the Frequency Domain" in Infrared and Millimeter Waves, Vol 11, Academic Press, PP 239-289, 1984.

[14]. F.L. Walls, A.J.D. Clements, C.M. Felton, M.A. Lombardi, and M.D. Vanek, "Extending the Range and Accuracy of Phase Noise Measurements", in Proc. of 42nd Frequency Control Symposium, Baltimore, MD, June 1-3, 1988, 432-441.

[15]. Warren F. Walls, "Cross-Correlation Phase Noise Measurements", in Proc. 1992 IEEE Int. Freq. Contr. Symp., pp. 257-261.

[16] The low PM noise synthesizer was a HP 8662A.

[17]. The low PM noise amplifier was a UTO 1023 from Avantek.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

HIGHLY SENSITIVE SAW SENSORS

I.B.YAKOVKIN ,YU.N.VLASSOV, A.S.KOZLOV, S.M.REPINSKY

Institute of Semiconductor Physics ,
Siberian Branch of Russian Academy of Sciences
Lavrentyev av. 13, Novosibirsk, 630090, Russia.

Abstract

This paper presents experimental data on the performance of some types of SAW sensors like, biosensor, gas sensor and pressure sensors. It is shown that such parameter, as sensitivity is determined by kind of piezocrystal cut and direction of SAW propagation as well as by the construction of sensor, method of treatment of signal, properties of covering film adsorption.

It is shown that such devices have a high sensitivity to the liquid temperature, gas concentration and pressure, respectively.

Typical data of sensitivity of our sensor are as follows. Biosensor has possibility to measure relative velocity change of about $5.7 \cdot 10^{-8}$, gas sensor has the sensitivity of about 100÷1000 Hz/ppm, at last, different pressure sensors have sensitivity of the order of 4.0 Hz/Pa for low pressure and 0.02 Hz/Pa for the high one.

The general properties of measurement are determined on the features of frequency control element.

Introduction

The last decade is characterized by increasing interest to working out different sensors not only physical magnitudes but chemical, biological parameters of the gaseous and liquid media [1-3]. Investigation of operating principles of sensors should be related with sensitivity, accuracy, selectivity, quickness, reliability, dynamic region, etc. Namely, these properties are important in solving environment monitoring problems, for the control of technological processes, medical diagnostics, etc. Improvment of monitoring quality is directly related with the quality of life improving.

Physical principles of operating acoustoelectronic of these sensors are extremely simple. The main part of SAW energy in delay line is concentrated in surface layer at the depth near the wave length (on the 1 GHz it is equal to 3 microns). Change of delay time is due to external impact on the surface with SAW, namely: change of the distance between input and output transducers, change the of SAW velocity because of change of physical constants, change of boundary condition on the surface of solid, or due to parameters of layered structures.

Typical data of sensitivity of the different microacoustics sensor are in Table 1.

Table 1.

Measured parameters	Sensitivity	Resolution
Temperature	2800 Hz/grad C	$3.3\ 10^{-4}$ grad C
Pressure	0.36 Hz/Pa	2.8 Pa
Acceleration (force)	9200 Hz/G	$1.1\ 10^{-4}$ G
Electric field	6.9 Hz/V/mm	0.14 V/mm
Displacement	300 Hz/μkm	0.003 μkm
Flow	11 Hz/cm^3/sec	0.09 cm^3/sec
Gas concentration	150 Hz/ppm	

The problem of sensitivity increase of SAW sensors has not been solved till now. However there are the prerequisites for successful solution of such problem.

In this paper we consider three kinds of SAW sensors.

1. The liquid sensor (biosensor).
2. The gas sensors (detectors).
3. The pressure sensors.

1. The liquid sensor (biosensor).

Developed methods for acoustic microanalysis of liquids utilize the distinctive attributes of the wave propagation in solid-liquid-solid (SLS) structures [4]. Future reduction in the volume of the investigated liquid is of main interest for biological and medical applications. The normal-mode formalism is used to describe the acoustic process in the structure in the case of a layer whose thickness-H is much smaller than its length. Such parameters of normal modes as the phase and group velocity are determined by the properties of a liquid. To relate them we used a representation on leaky waves [5]

$$v = v_r \cdot \left(1 - \rho_f \cdot v_r \cdot \beta \cdot \kappa \cdot H/2\right)$$

$$v_g = v_r \cdot \left(1 - \rho_f \cdot v_r \cdot \beta \cdot \kappa \cdot H\right)$$

where β is the combination of solid parameters,
v_r is the Rayleigh velocity,
ρ_f is the density of liquid.

Leaky SAW propagation along SLS structure in two SAW delay lines was employed to detect only amplitude and phase variations. The phase differences between the reference and sensing signals were monitored by method that has been developed [6]. We have worked out the LSAW oscillator using pulse circulation and operating in the dynamic range of 30 dB and frequency region about 10 to 100 MHz. The study of the stability has been carried out on the delay line of 7.04 μksec, fundamental frequency of 43.6 MHz. A computer based measurement system is intended for determination of acoustic properties, of liquids of biological nature in the volume of 50 microliter, temperature range of 15÷60 grad C and sensitivity to the relative velocity change about $5.7 \cdot 10^{-8}$.

Fig. 1 shows the set-up of a differential unit (biosensor) electron system for diagnostic biological liquids.

Fig. 2 shows the temperature dependence of the sound propagation time in the solution of erythrocyte shades in the human blood. It is clearly seen that the curve has two peaks. That corresponds to two phase transformations at 36.5 and 49 grad C.

0-7803-1945-1/94 $3.00 © 1994 IEEE

2. The gas sensors (detectors).

We used equation that described frequency shift of SAW oscillator δf due to influence of the layer on semispace [7]. If the acoustic property of layer (ρ, λ, μ) essentially differs from that of semispace (ρ', λ', μ'), frequency shift is as follows:

$$\delta f = (k_1 + k_2) \cdot f_0^2 \cdot H \cdot \rho - k_2 \cdot f_0^2 \cdot H \cdot \left[4 \cdot \mu / v_r^2 \cdot (\lambda + \mu) / (\lambda + 2 \cdot \mu) \right]$$

where k_1 and k_2 are the constant for the substrate,

H is the thickness of layer,

f_0 is the central frequency of the SAW,

ρ, λ, μ and ρ', λ', μ' are the density, Lame coefficients for layer and semispace, respectively.

It is seen that first term describes shift of frequency due to massloading and the second one depends only on elasticity of layer structure.

At the frequency about one GHz the mass that may be measured by means of SAW is equal to 10^{-14} gram per millimeter square.

The sensitivity of gas sensor depends on frequency as a square. The limit of sensitivity or minimum massloading on the unit square depends on signal to noise ratio. If the noise level (amplitude) grows proportionally to frequency, the signal to noise ratio is improved linearly with frequency. Therefore the sensitivity limit is reverse to the frequency.

The time reaction is an important parameter for gas sensor. It is determined by the whole spectra of processes. One of them is equilibrium at adsorption, another is diffusion of gaseous molecules from surface where they are adsorbed in volume of thin film. In the case when time is determined by diffusion, and diffusion obeys to the Fick law, time of sensor switch on is reverse to forth power of frequency. However the experimental results are not coincident with it, as the process of diffusion is rather surface than volume.

Selectivity, time reaction and reversibility of processes on surface are strongly dependent on adsorption energy (from 1.0 to 100 Kkal/mole) of gas covering film.

As follows from analysis [8-11] typical sensitivity for $C_6H_5CH_3$, CCl_4, H_2, is equal to 1-2 Hz/ppm and for H_2S, NH_3, NO_2, SO_2 is about 100÷300 Hz/ppm.

Fig. 3 contains the scheme of SAW sensor for measurement of concentration of SO_2. Construction of the sensor consists of box 3, sensitive SAW element 2 and electronics circuit 1 (in its turn it has two amplifiers, references resonator, mixer and buffer). Sensitive element is a rectangular substrate of piezoelectric (it is resonator with thin film thermoresistor). The surface of substrate (that is ST,X-cut of the quarts) is covered by selectivity adsorbed layer.

We have investigated different kinds of gas sensors: with covering film Langmuir-Blodgett (3 layers of diacetilene , four layers of diacetilenamine), with film of SiO_2 (500 A^0) that was modified by diethylenamin, at last we used thin film of WO_3. The first two layers used for detection of SO_2, the last one used for H_2S. Measurements of SAW frequency change in oscillator were made at frequency about 360 MHz.

Measurements of sensitivity took place on the set-up shown in Fig. 4. This set-up allows to prepare mixture of SO_2 of required concentration in the region 5÷10000 ppm. As an intermediate gas a dry nitrogen was used . Procedure preparation of gaseous mixture is as follows. It is needed to evacuate the bypass volume V. Then from the valve II the gas removes in the bypass volume. After that dry nitrogen introduces up to 150 KPa.The relation between bypass volume and volume of work chamber is equal to 1:300. Box for measurement allows to adjust the flow of mixture.

Fig. 5 shows a dependence of differential frequency versus time evaluation for L-B thin film gas sensor. Region I is vacuum. Region II is insufflation of SO_2 with concentration of about 60 ppm, region III is pump out the mixture. Fig. 6 shows the similar dependence for the case of SO_2 concentration of about 100 ppm in the dry nitrogen. One can determine the sensitivity of gas sensor to SO_2 as 15 Hz/ppm.

Fig. 7 illustrates the deviation of differential frequency for the gas sensor sensitive element with modified SiO_2 thin film. Different regions of this curve behavior have respect to change of SO_2 concentration. At last, Fig. 8 demonstrates dependence of differential frequency for sensitive element with WO_3 thin film.

The following results are obtained:

Table 2

Gas	SO_2			H_2S
Sensitive layer	L-B layers	SiO_2 mod.	WO_3	WO_3
Sensitivity, Hz/ppm	15	1000	80	100

3. The pressure sensors.

Here are the results of design and performance of some types of SAW pressure sensors [12]. It is shown that such devices have a high sensitivity from 0.026 Hz/Pa for high pressure sensors (12000 KPa) to 27 Hz/Pa for low pressure sensor (20 KPa). Accuracy of these devices is about 0.15%.

In many cases the long term and temperature stability of semiconductor pressure sensors are not sufficiently high. That is why developing other types of pressure sensors is necessary. Here we present experimental data on two types of SAW pressure sensors [13,14].

The sensitive element of the first type sensor is a crystal membrane is fabricated on a Y-cut quartz substrate (Fig. 9). Two SAW resonators and one thermoresistor are placed on its surface. They all made of aluminum film with a thickness of 0.15 of a micrometer. One resonator is placed in the center of the membrane. The distance between the centers of the membrane and the other resonator is 0.75 of the membrane radius. Both resonators are connected by the feedback loop of two oscillators and serve as frequency control elements. Under the influence of measured pressure the membrane surface becomes curved. The surface deformation causes changes in the central frequencies of the resonators. As the resonator centers are placed in points of opposite signs of deformations their frequency shifts have different signs too (Fig. 10). The difference frequency is a sum of the resonator frequency shifts. The measured pressure is a function of this difference frequency and of temperature controlled by the thermoresistor. The frequencies of the signal from the mixer output and the thermoresistor value are transformed in a digital form and used for pressure calculation in a microcomputer.

The sensitive element of the second type sensor is a plate of

ST-cut quartz. The plate leaned on the edges of a rectangular hole in the steel unit (Fig. 11). The SAW resonators are fabricated on the plate surface. The sensor also includes a vacuum steel membrane with steel rod that welded to the center of the membrane. Under the influence of pressure the membrane center with the steel rod moves and squeezes the plate surface. The plate deforms and the resonator frequency changes. Its frequency shift and the thermoresistor value are processed in the same manner as in the previous sensor. The second oscillator uses a mechanically unloaded SAW resonator as a frequency control element. The material is chosen due to its small temperature coefficient and high sensitivity of SAW velocity to the static strains.

The short description of the technology of SAW pressure sensor sensitive element is as follows:

-the basic material is Y-cut (or ST-cut), X direction quartz, optically polished,

-10 nm thick vanadium layer, and

-150 nm thick aluminum layer, deposited by vacuum evaporation,

-etching of the aluminum pattern.

The main features of two different samples concerning the first type SAW pressure sensors are given in Table 3.

Table 3

Parameters	Sensor 5	Sensor 10
Diaphragm diameter, mm	30	9
Diaphragm thickness, μkm	200	2000
Pressure range, KPa	0÷16	0÷12000
Sensitivity, Hz/Pa	4.00	0.0263
Nonlinearity, %	0.56	0.7
Mean error, %	0.28	0.1
Threshold pressure (at the 10 Hz frequency accuracy), Pa	2.5	3800
Central frequency, MHz	197.1	310

For the SAW pressure sensor of second type we had opportunity to increase sensitivity by using the large diameter intermediate membrane. The main properties of these sensors are in Table 4.

Table 4

Parameters	Sensor 2	Sensor 3
Membrane material	vacuum steel	fused quartz
Dimension of ST-cut plate of resonator, mm	15*8*0.3	the same
Sensitivity, Hz/Pa	9.7	27.0
Pressure range, KPa	100÷116	0÷20
Hysteresis, %	0.15	0.15
Pressure threshold (at 10 Hz frequency accuracy), Pa	1.0	0.4

SAW pressure sensors under study have such attractive properties as high sensitivity, low threshold, design and technology simplicity, non complicated electronic equipment without any thermocompensation, that doesn't influence on sensor stability, small dimensions and frequency output.

They have a higher long-time and temperature zero drift as compared with bulk wave resonator pressure sensor. The strain transformation sensors have additional disadvantage. It is a high level hysteresis.

Calculated [15] and experimental difference of these relative frequency shifts and some other membrane parameters are in Table 5. It shows a good agreement between numerically calculated and experimental data. It has been found that ST-cut membranes with direction of propagation of $\varphi = 50^{0}$ are to have high pressure sensitivity and temperature stability simultaneously.

Table 5.

Membrane, N	1	2	3
Thickness, mm	2	2	1.1
Diameter, mm	9	9	12
Distance between resonator centers, mm	4	4	4.8
Resonator central frequencies, MHz	310	310	360
Substrate, Quartz	Y-cut	ST,X-cut	ST,X-cut
Experimental relative sensitivity, $(1/Pa)*10^{12}$	60	20.2	90
Theoretical relative sensitivity, $(1/Pa)*10^{12}$	75.4	24	100

The experimental data allow us to conclude that the sensitivities of both types of SAW pressure sensors are sufficiently high and they could be used in pressure measurement equipment.

Conclusion

The design and perfomance of different kinds of SAW sensors together with electronic circuits provide comparatively simple devices that could be used in a measurement equipment.

Acknowledgements

The authors would like to thank our colleagues who participate on different stage of work out of sample above mentioned sensors: Dr. N.S.Pashchin, Dr. R.M.Taziev, G.S.Ostanina, A.D.Mayer, A.A.Kreimer.

References

[1] Transducers'87, *Int. Conf. on Solid-State Sensors and Actuators,* Digest of technical papers, Tokyo, June 1987.

[2] Biosensors International Workshop 1987, GBF Monographs, vol.10 VCH Verlagsgesellschaft, Weinheim, 1987.

[3] 1988-1991 IEEE Ultrasonics Symposium Proceedings, New York.

[4] H.L.Bertoni, T.Tamir, "Unified theory of Rayleigh-angle phenomena for acoustic beams at liquid-solid interfaces", *Appl.Phys.*, 2:157, (1973).

[5] A.V.Arapov, V.S.Goncharov, I.B.Yakovkin, "Leaky waves in layered system", *Sov. Phys.Acoust.*, 31:439, (1985).

[6] A.V.Arapov, V.S.Goncharov, S.V.Ruchko and I.B.Yakovkin, "LSAW in Solid-Liquid-Solid Systems. Acoustoelectronic Microanalysis of Viscous-Elastic Properties for Liquids of Biological Nature",*Proc. Intern. Symp. Physical Acoustics, Fundamentals and Application,* Kortijk, Belgium, 1990, pp.213-217, Edited by O.Leroy and M.A.Breazeale, Plenum

Press. New York and London, 1990.
[7] H.Wohltjen, A.Snow, D.Ballantine, "The selective detection of vapours using surface acoustic wave devices", *Proc. Int. Conf. on Solid-State Sensors and Actuators,* 1985, pp.66 - 70.
[8] C.Caliendo, A.D'Amigo, P.Verardi, E.Verona, "SAW H_2 sensor on silicon substrate", *IEEE Ultras. Symp. Proc.*, 1988, pp.569-574.
[9] R.Lec, J.F.Vetelino, R.S.Falconer, Z.Xu, "Macroscopic theory of surface acoustic wave gas microsensor", *IEEE Ultras. Symp. Proc.*, 1988, pp.585-589.
[10] A.D'Amigo, A.Petri, P.Verardi, E.Verona, "NH_3 Surface acoustic wave gas detector", *IEEE Ultras. Symp. Proc.*, 1987, pp.633-636.
[11] M.S. Nievwenhizen et all, "First experimental results with a silicon based surface acoustic wave (SAW) chemical sensor for NO", *Eurosensor II, Twente*,1988, pp.52.
[12] N.S.Pashchin, I.B.Yakovkin, "SAW dispersion in anisotropic layer structure with static strain", *Acust.Journal*, 1986, v.32, pp.352-357.
[13] Yu.N.Vlassov, A.S.Kozlov, N.S.Pashchin, I.B.Yakovkin, "SAW pressure sensor", *Proc. international Acousto-Electro, St.Petersburg*, 1993, pp.138-139.
[14] Yu.N.Vlassov, A.S.Kozlov, N.S.Pashchin, I.B.Yakovkin, "Precision SAW pressure sensors", *Proc. the 1993 IEEE Frequency Control Symp.*, pp.665-667.
[15] R.M.Taziev, E.A.Kolosovsky, A.S.Kozlov, "Deformation - sensitive cuts for SAW in α-quartz", *Proc. the 1993 IEEE Frequency Control Symp.*, pp.660-664.

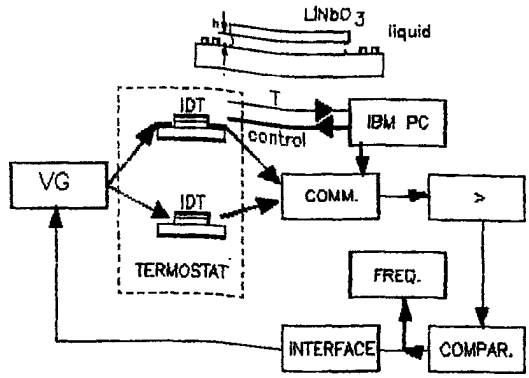


Fig. 1 Block-diagram of measurement system: Method of impulse oscillator, and the scheme of SLS-structure.

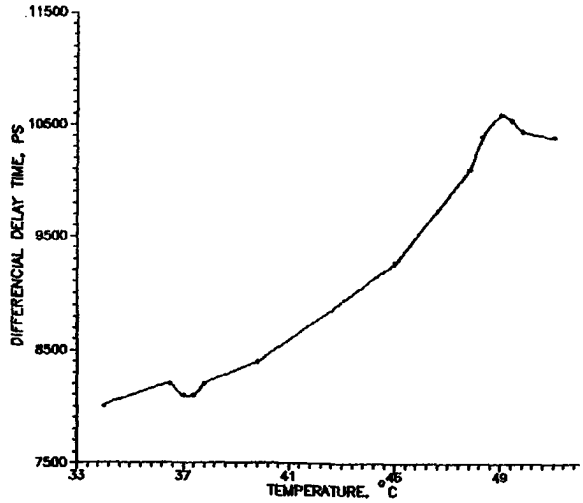
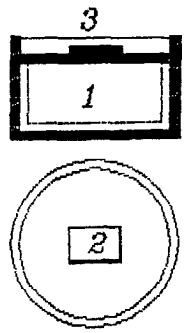
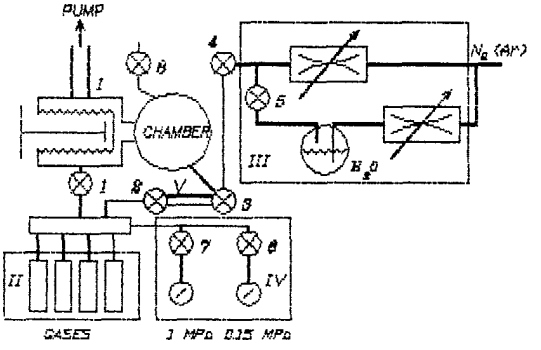


Fig. 2 Difference of delay times in two channels of sensors vs temperature, liquid-suspension of erytrocite shadows.



1 - electronics circuit; 2 - sensitive SAW element; 3 - box.
Fig. 3



1, 2, 3, 4, 5, 6, 7, 8 - valves; I - faucet; II - balloons with gases; III - insuffle system; IV- manometers; V - by pass volume. Relation volume of by pass to chamber is equal to 1:300.
Fig. 4 Block diagram of measurement gas sensitivity.

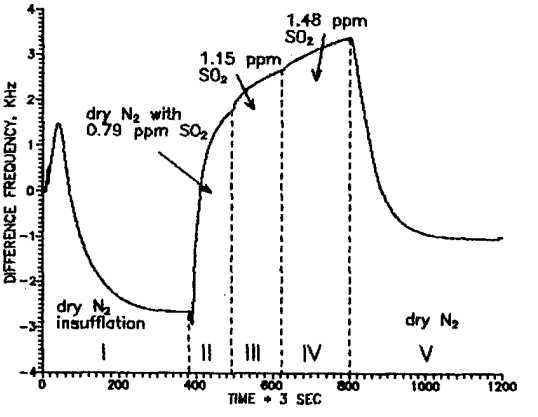


Fig. 7

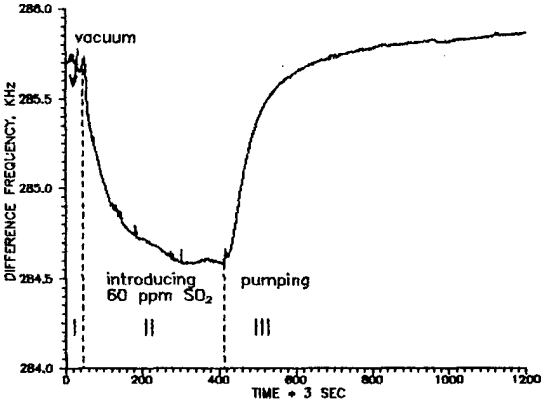


Fig. 5

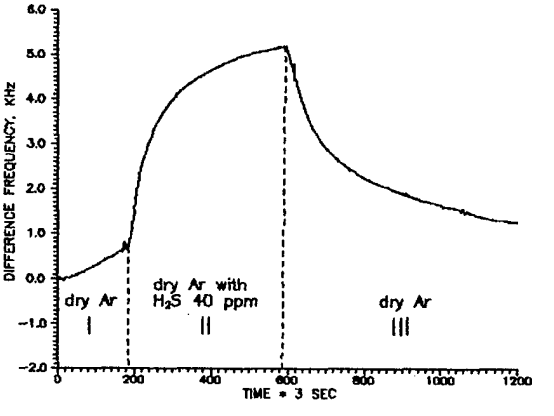


Fig. 8

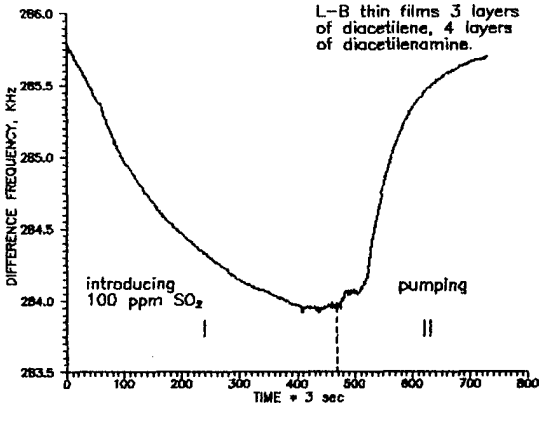


Fig. 6

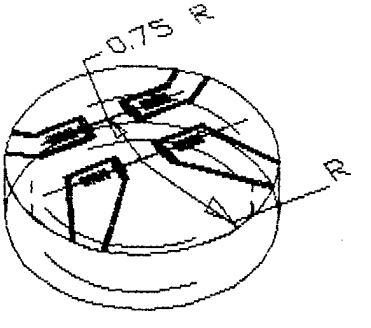


Fig. 9 Membrane two rezonator pressure sensor.

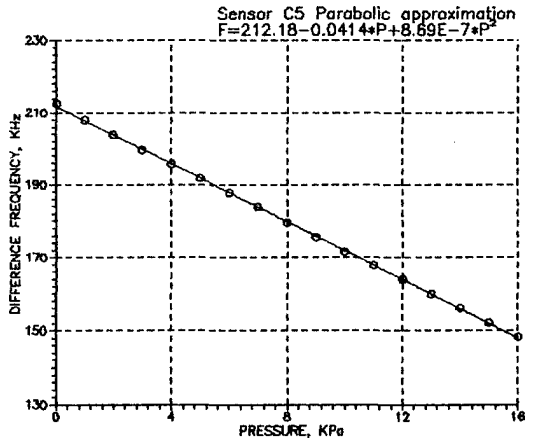


Fig. 10 a)

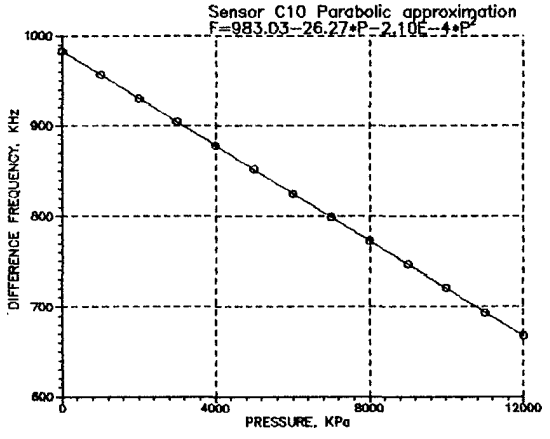


Fig. 10 b)

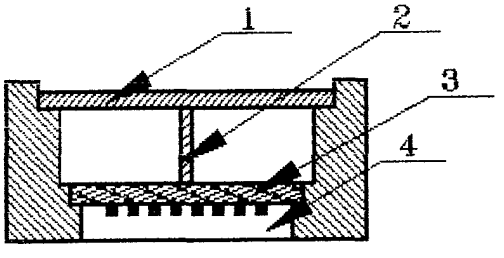


Fig. 11 Plate of ST-cut quartz pressure sensor.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

CHEMICAL/BIOLOGICAL CONTAMINANT DETECTOR FOR AQUEOUS ENVIRONMENTS: PRELIMINARY REPORT

R. McGowan, J. Foerster, D. Huynh, R. Lindenmuth &
T. Lukaszek
ARL, Electronics & Power Sources Directorate
Ft. Monmouth NJ 07703

Y. Lu
Rutgers University, Electrical Engineering Department
Piscataway, NJ 08855

Abstract

Paired surface transverse wave resonators operating near 500 MHz are being used as the sensing element in a chemical/biological detector capable of operating in water. Initially, the design of the surface transverse wave detector will be discussed. Then, the major problems uncovered with the prototype detector will be reviewed with an emphasis on solutions to these problems that will be implemented in a second-generation "smart" detector.

Introduction

One of the most challenging problems facing today's sensor industry, for both military and commercial applications, is the rapid detection of biological agents. This problem is even further complicated by the requirement that the detection technique be sensitive, selective, power-efficient and portable. Surface acoustic wave (SAW) sensors have demonstrated that these performance criteria can be met for the detection of chemical vapors in air [1]. However, when SAW devices are operated in a medium more viscous than air, such as water, mechanical damping of the acoustic wave produces excessive attenuation and makes the device unusable. The primary source of the mechanical damping in a SAW is the normal component of its elliptical particle motion. Fortunately, SAW is only one of several acoustic modes that will propagate in a piezoelectric substrate. There exist modes of vibration whose particle motion is parallel to the surface. One such mode is the acoustic plate mode whose transverse particle motion within the bulk of the crystal makes it possible to produce a sensitive gravimetric sensor capable of operating in a fluid. Other advantages of an APM device are 60% higher acoustic velocity, better aging behavior, increased power handling capability, higher intrinsic material Q and lower device phase noise.

The acoustic plate mode (APM) has two modes of practical importance which are the surface skimming bulk wave (SSBW) and the surface transverse wave (STW) [2]. Although both have parallel particle displacements, the STW travels close to the surface, with most of its energy several wavelengths below it. In contrast, the SSBW and its energy travel into the bulk of the substrate at an angle θ.

Detection sensitivity in most acoustic sensors is controlled to a great degree by the amount of interaction that occurs between the acoustic wave and a sorbing material. The superb sensitivity exhibited by SAW results from the normal component of its elliptical particle motion combined with the concentration of its wave energy at the surface. In an effort to optimize the mass sensitivity of the APM detector, the sensor element was designed as a STW device to take advantage of the confined wave energy near the surface. This will increase the wave's interaction with the contaminant and reduce the insertion loss of the sensor.

This paper presents some preliminary results obtained from a biosensor prototype utilizing a dual-channel surface transverse wave (STW) resonator as a quartz microbalance. Initially, the design of the chemical/biological detector will be reviewed concentrating primarily on the rf and digital electronics. Then, resonator and oscillator problems will be presented with an emphasis on solutions to these problems that will enhance the sensitivity of the biosensor. Finally, a second prototype biosensor will be proposed which solves the preliminary problems of the first prototype, while also meeting the evolutionary requirement of being compatible with the RS-232 and RS-485 communication protocols for possible sensor networking. This last feature will be critical to its proposed usage as a field deployable point-detector.

STW Sensor

Gravimetric effects produce frequency changes in acoustic devices because mass deposited on the active surface of the device slows and damps the particle motion in the piezoelectric substrate. Damping of the particle motion causes a reduction in the wave velocity of the acoustic signal which ultimately results in a decrease in the output frequency of the device. This phenomenon can be used to make a chemical or biological sensor, if two criteria are met. First, the desired analyte must be

0-7803-1945-1/94 $3.00 © 1994 IEEE

immobilized on the surface, usually, by some type of sorption via a selective coating. Second, this analyte/coating interaction must create enough mass loading to alter the velocity of the acoustic wave at useful analyte concentration levels [1]. Clearly, the selectivity and specificity of the STW sensor are a function of the physicochemical nature of the coating material.

A two-port STW resonator structure has five components, which are an input and output set of interdigital transducers (IDT), two sets of reflectors and an intertransducer grating. It has been shown that the response of a STW resonator near-resonance is very similar to that of SAW resonators [3]. Therefore, the general design of a STW resonator can be obtained from the SAW resonator design review in Gerber & Ballato, assuming that the wave velocity for the APM mode is used [4]. The major design differences between STW and SAW are an additional intertransducer grating used to guide the APM along the surface of the device and the wave propagation direction of the device relative to its crystallographic axes. The layout of the STW resonator can be seen in Figure 1. The STW resonators were fabricated on a 7.62 cm diameter ST-cut quartz wafer, a popular SAW cut. However, instead of propagating the acoustic wave in the x direction, which is the standard SAW orientation, the resonator structure was rotated 90° with respect to the crystallographic axis so that the acoustic wave propagates in the z direction producing the APM. Follow-on devices will be fabricated on an AT cut of quartz which is temperature-compensated for the STW mode.

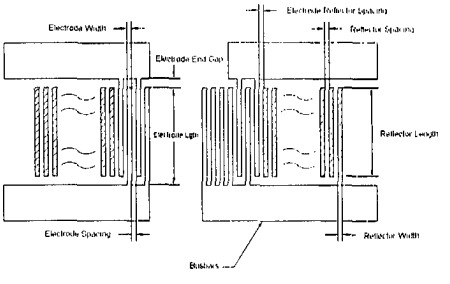


Figure 1 Schematic Diagram of 500 MHz STW Resonator

Biosensor Electronics

The biological agent detector can be partitioned into two primary sub-systems: sensing and displays. The sensing sub-system is formed primarily by the rf electronics which include two feedback-loop oscillators and a mixer. The basic feedback-loop oscillator, as seen in Figure 2, has six essential components: a phase shifter, loop amplifiers, a feedback-loop sampler, gain limiters, a frequency control device and a buffer amplifier. These components are required to insure the two conditions of oscillation are met within the loop. The first condition is that there must be 3 ± 1 dB of excess small signal gain in the feedback-loop. The second condition is a net $2\pi N$ phase shift around the entire loop [5].

In operation, one of the two resonators is covered with a selective coating and serves as the test channel, while the other resonator serves as the reference channel. As the desired analyte is flowed over the active region of the coated resonator, the analyte preferentially binds to the selective coating which mass loads the surface and results in a decrease in the output frequency of the test resonator. The detector is operated in a differential mode by mixing the output frequency of the test and reference oscillators and analyzing the difference frequency. A differential mode of operation is preferred because it minimizes systematic errors caused by changes in temperature, pressure and flow rate.

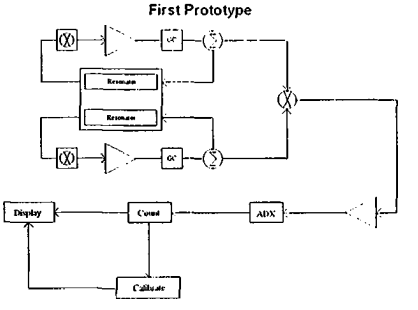


Figure 2 Block Diagram of Initial Biosensor

The first stage of the display sub-system is a Common Emitter configured Bipolar Junction Transistor that amplifies the differential frequency signal. The amplifier is followed by a wave shaping and level shifting stage that converts the differential frequency sinusoidal signal to a square signal from 0 to 12 Volts DC. The square wave drives the clock input of the frequency counters.

The display electronics sub-system includes a system clock that consists of a crystal oscillator and a frequency divider. The system clock creates one second sampling windows for counting the differential frequency and displaying the results on a liquid crystal display. Functionally, the differential frequency is accurately counted when the counters are allowed to operate for exactly one second.

Calibration data is initially stored in memory and then subtracted from subsequent readings. This difference is the true frequency shift due to mass loading on the surface of the detector and is displayed on the liquid crystal display for one second between counting intervals.

Experimental Findings

Testing an array of eight device designs which consisted of two different numerical apertures (NA) each having four different intertransducer spacings, confirmed that optimal wave coupling occurs for an intertransducer spacing of $n\lambda/2$ for STW devices [3]. The device selected for the sensor was the 80λ NA device because its insertion loss was 6dB lower than its 40λ counterpart. On average the devices had a resonance frequency of 487.5 MHz and an unloaded Q of 6700. The residual phase noise of the STW resonator used in the oscillator was -125 dBc/Hz at the 1 Hz intercept point.

One of the disadvantages of using the STW mode is that the analyte must come in direct contact with the active, metallized side of the device. Initially, the protective layer was a thin coat of SiO_2, approximately 1000Å thick, which would simply protect the electrodes from oxidizing or shorting due to any ionic content of the water. However, initial testing revealed that there was still some phenomenon that was severely degrading the device's performance when water was placed on the top surface. The problem was determined to be the strong electric field concentration at the edge of each electrode that was polarizing the water molecules above it and shorting out the device. The following equation was used to model the normalized electric field strength at the surface of the device [6],

$$E_1(\psi,a) = \frac{2\theta}{K(\sin^2\theta)} * \sum_{n=0}^{\infty} P_n(\cos 2\theta) * \cos[(2n+1)\pi\psi].$$

In this equation, ψ is the distance along the direction of wave propagation, P_n are the Legendre polynomials, K is the complete elliptic integral of the first kind, and $\theta = (l-a)\pi/2$ where l is the grating period distance and a is the electrode width. Figure 3 shows a Mathematica plot of the normalized electric field strength for a metallization ratio of 1/2, therefore the electrode and space are both equal to λ/4. The plot shows the electric field for two electrodes centered at 0.5 and 1.5, where the charge on each electrode is equal and opposite.

Given the presence of this strong electric field at the edges of the electrodes a shielding layer would have to be devised to reduce the strength of the electric field incident on the water. Modeling the electrodes as an array of constant line charges spaced λ/4 apart and assuming a low dielectric constant shielding layer ($\varepsilon_r \approx 3$), a shielding layer thickness of 2.0 um would reduce the field incident on the water by greater than an order of magnitude. This agreed well with preliminary experimental results that indicated acceptable device performance with water-loading for a 1.7 um thick rigid photoresist shielding layer.

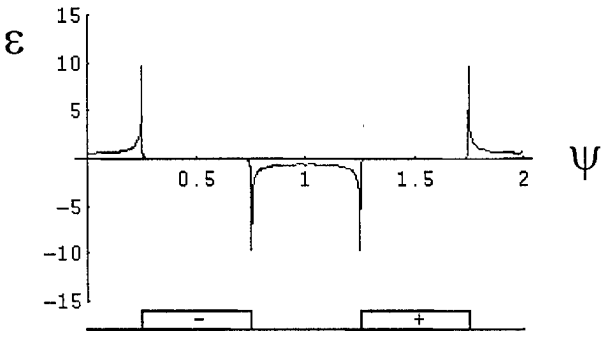


Figure 3 Normalized Electric Field Strength at Device Surface

The proposed sensor configuration can be seen in Figure 4a. The thin $Si0_2$ layer provides protection against oxidation and ionic shorting of the electrodes in solution. The thick, low dielectric polymer shielding layer prevents device shorting caused by polarization of the water molecules due to the strong electric field present at the electrode edges. The channel removed between the transducers allows the analyte to bind as close to the active surface as possible, improving the sensitivity of the device. The top layer is the chemically or biologically selective coating that will mass load the surface when exposed to the analyte.

The other problem that occurred was discovered when the short term stability of the differential oscillator was tested. There was a 1/600 Hz signal that was modulating the nominal 200 kHz output signal by ± 250 Hz. It was realized that the most probable source of the modulation signal was the common voltage regulator which supplied 15 V to each oscillator. Figure 4b shows the solution to this problem which was to place each oscillator on its own voltage regulator, giving 90 dB of isolation between each oscillator and the battery pack. The short term stability of the output signal improved by an order of magnitude as a result of introducing a second voltage regulator.

Currently, the modified sensor structure is being fabricated and appropriate test fixtures are being designed. Once a suitable detection layer is found and a flow cell is designed to deliver the analyte to the sensor surface, testing of the biosensor can begin.

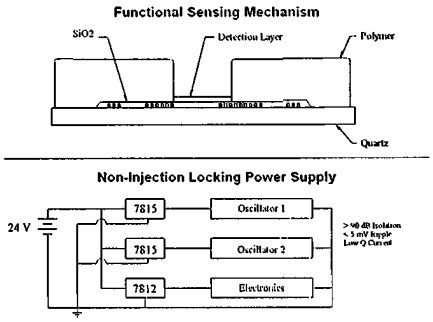


Figure 4 Solutions to Biosensor Problems:
a) Multi-layer Protective Coating
b) Non-Injection Locking Power Supply

The second generation biosensor, besides incorporating the enhancements listed above, will concentrate on reducing the power requirements and size of the rf electronics, providing more isolation for the oscillator loop, and evolving the digital electronics to incorporate a micro-controller. A micro-controller will enable calibration and compensation data to be stored and used to correct for environmental effects when networked with appropriate sensors. Sophisticated user interfaces such as high resolution displays and remote communications protocols, such as RS-485 and RS-232 ports, will be driven with the micro-controller. This enhanced capability resulting from the use of a micro-controller will provide the sophisticated data processing required for any sensor system to make an impact on future monitoring requirements.

References

[1] H. Wohltjen, "Mechanism of operation and design considerations for vapour sensors," Sensors & Actuators,(5), 1984, pp. 307-25.

[2] K.H. Yen, K.L. Wang & R.S. Kagiwada, "Efficient bulk wave excitation of ST quartz," Elect. Let., vol. 13, pp. 37-38.

[3] Avramov, I.D., "Microwave oscillators stabilized with surface transverse wave resonant devices," Proceedings of the 46th Annual Symposium on Frequency Control 1992, pp. 391-408.

[4] Gerber, E.A. & Ballato, A., Precision Frequency Control: Acoustic Resonators and Filters, Vol. 1, Academic Press, Inc., Orlando, Florida, 1985.

[5] Parker, T.E. & Montress, G.K., "Precision Surface-Acoustic-Wave (SAW) Oscillators," IEEE Trans. on UFFC, vol. 35, no. 3, pp. 342-64.

[6] Ballato, A., "Transduction of Acoustic Surface Waves by Interdigital Arrays, and Equivalent Circuit Representations," Research and Development Technical Report, ECOM-4359, October 1975.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

HIGH FREQUENCY MONOLITHIC CRYSTAL FILTERS (MCF) USING PIEZOELECTRIC ELEMENTS OPERATING AT THE FUNDAMENTAL FREQUENCY 30-150 MHz AND THEIR ADJUSTMENT METHOD

S.V.VOLKOV, V.A.MOSTIAEV, L.V.GRIGORJEV, V.S.FROLOV, Y.M. ERMAKOV,

Scientific Research Institute "Phonon", 105023, Moscow, Krasnobogatyrskaya str., 44, Russia

Abstract

The application of ion-plasma technology enabled us to develop and put into production MCFs using piezoelectric elements with reverse mesa structure operating at the fundamental frequency over the range from 30 to 50 MHz. The filters are manufactured in flat metal enclosures with the dimensions 12.0x12.0x4.2 mm. The advantage of this adjustment method consists in a practically maskless technology because no necessity exists in the mask alignment with the electrode configuration, which would be difficult with a gap of 0.05-0.18 mm. The parameters of 2nd order MCF FP2P-486 operating at the frequencies of 34338, 68736 and 149950 kHz and their frequency attenuation characteristics are given, including frequency attenuation characteristic of 4th order-MCF at 149950 kHz, synthetisized from two 2nd order MCFs.

The application of chemical or ion-plasma etching technology for manufacturing crystal elements having a reverse mesa-structure and used in piezoelectric crystal units and monolitic filters [1] allowed us to design and put into production high frequency monolitic filters operating at the fundamental frequency over the range 30-150 Mc.

The designing of a unique construction type and technology for the whole frequency range leads to the simplification of the production process satisfying simultaneously the needs of many consumers relating to frequency parameters. Filters are produced in metal-glass plane enclosure by the use of one continuous electrode (Fig.1,-1) at the side of mesa structure and two separate (Fig.1,-2,3) electrodes at the plane side of the crystal element (Fig.1,-2). Thin film control land areas of the crystal element are carried out by vacuum electroplating a three layers-film consisting of chrome, copper and nickel.

Thin film electrodes of the piezoelement intended for the frequency range 30-90 Mc are made of silver and for the range 90-150 Mc - of aluminum. The piezoelement is bonded to the sittale plate (Fig1,-5) in which a seating place for it is etched by photolithographic method; the bonding of the piezoelement is carried out by using an organic silicium containing compound characterized by low gas-secretion. The joining of contract land areas of the crystal element and terminal edges of the enclosure (Fig.1,-6) is realized by using a gold wire of 30 μm in diameter bonded to the land areas and terminal edges by microwelding technique. The photograph of the filter is given in Fig.2. Hermetic encapsulation of the filter is made by joint laser welding, the enclosure being filled with nitrogen of high purity after preliminary evacuation.

One of the distinctions of the technology process used for manufacturing the filters in question, as compared with the traditional production process, is the adjustment of filter parameters by ion etching means.

The adjustment method using material vacuum plating through special masks is unacceptable in this case because of the extremely sizes of thin film electrodes and the distance between them varying from 180 to 50 μm [2].

The most suitable adjustment method in our case is obviously ion etching technique widely used at the present time in crystal unit production process and described in literature [3,4]. The possibility of using electrode etching technique for the adjustment of the 2nd order monolitic filter parameters are analyzed in the filter mentioned above was carried on at one side for the piezoelement was mounted in parallel with the multi-terminal base of the enclosure at the minimum clearance.The surface of the piezoelement was situated normally to the ion flow in such a manner that the mutual arrangement of the piezoelement and the ion source insured the uniform simultaneous etching of the both electrodes. The piezoelement parameter adjustment (the aligning of the dot resonator frequencies, of the upper normal frequency and the normal frequency scatter) was realized by means of selective removing away part of the electrode plating material using a special moving mask.

0-7803-1945-1/94 $3.00 © 1994 IEEE

For piezoelements with electrode sized of some millimeters, the above described method of adjustment using a plane mask is expedient, yet in the case under discussion, when electrodes have the above mentioned small sized, this adjustment method is not practically realized because of difficulties associated with matching masks and electrodes, mount distortions, the necessity of approaching the mask closer to the element surface with the aim in view of avoiding possible etching of electrode parts, and uncontrollable shifts of the filter parameters as well.

For the adjustment of parameters of the filter under discussion it was proposed to use two ion sources (Fig.8,9) applying a vertical separating mask-shield (Fig.1,-7). The adjustment process arrangement is given in Fig.1.

The mask is removable, is placed on the filter enclosure using figured slots and its arrangement relative to the electrodes is controlled by applying a microscope.

For carrying out measurement and adjustment processed a short circuit is used by shorting one of the separate electrodes with the entire one.

While aligning frequencies of dot resonators, ion etching is applied to remove part of electrode material of that dot resonator which in the open circuit state has the least frequency according to the largest peak on the oscillograph screen. For this process to be realized an ion source is turned on operating at the side of the lower resonance frequency electrode and the process of material removing goes on up to the alignment of resonance peak amplitudes.

For adjusting the mid-band frequency of the filter both of the ion sources are turned on and the electrode etching process is carried on up to the necessary frequency value, after which both sources are turned off.

As a rule the most of filters do not need the operation of adjusting the frequency scatter, i.e. of adjusting the acoustic coupling factor. But in some cases of necessity masks may be used which are analogous to the above mentioned ones and have special slits for removing part of electrode material closely to the inter-electrode distance for narrowing the pass band and closely to the outer edges of electrodes for broadening the pass-bond.

The characteristics of the filter ФП2П4-486 at mean frequencies 34338, 68736 and 149950 kHz are presented in Table.1, and their attenuation frequency characteristics (AFC) are given in Fig.3-5.

To satisfy the increased requirements relative to the shape factor, attenuation frequency characteristic (AFC) and the guaranteed attenuation filters of higher orders may be synthesized on the basis of 9nd-order sections.

Fig.6 represents the characteristic of MCF at 149950 kHz of the 4th synthesized on the basis of two 2nd-order filters.

Table

Tech. data	FP2P4-486			
		01	03	09
Mode	Fundamental			
Order of filter	2nd			
Center freq. f, kHz	34338	41500	68736	149950
Pass-band width *) 3 dB, kHz 20 dB, kHz	±(10-50) ±(50-250)	±(15-60) ±(75-300)	±(20-100) ±(100-500)	±(30-100) ±(150-500)
Pass-band ripple, dB	< 1.0			
Insertion loss, dB	< 5.0			
Attenuation, dB	> 30		> 25	> 20
Temperature range, °C	-60÷+85			

*) Any value within this limit is avaible.

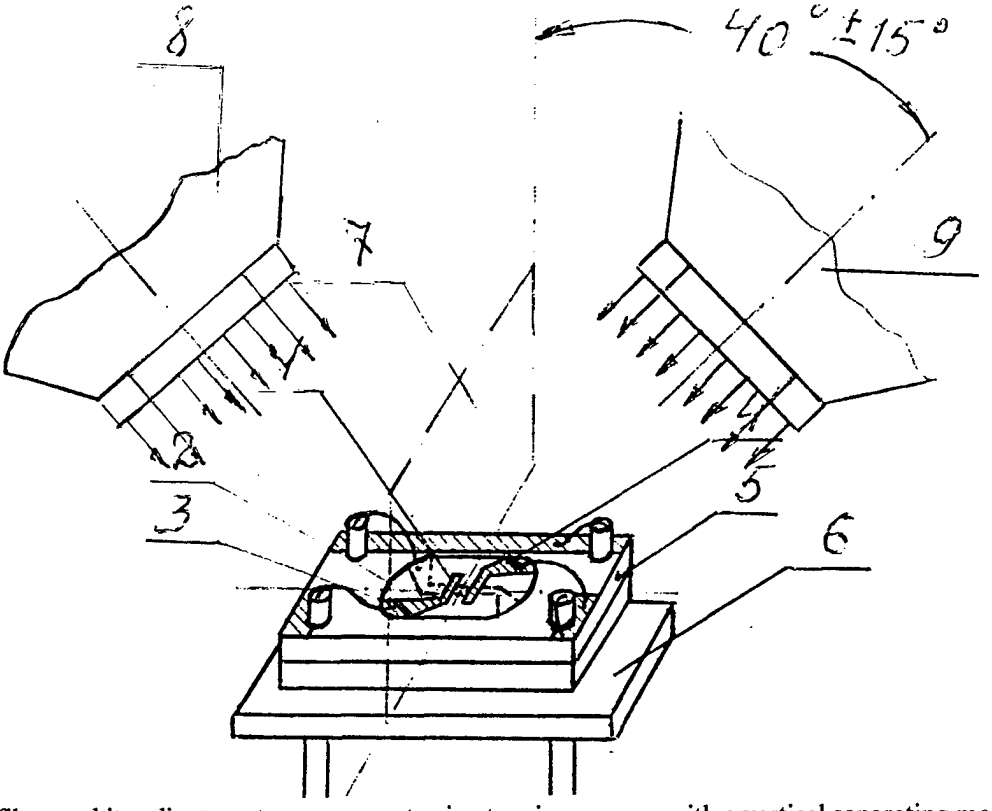


Fig.1. The filter and its adjustment arrangement using two ion sources with a vertical separating mask-shield.

1 — continuous electrode,
2 — crystal element,
3,4 — separated electrodes,
5 — sittale plate,
6 — enclosure,
7 — mask-shield,
8,9 — ion sources

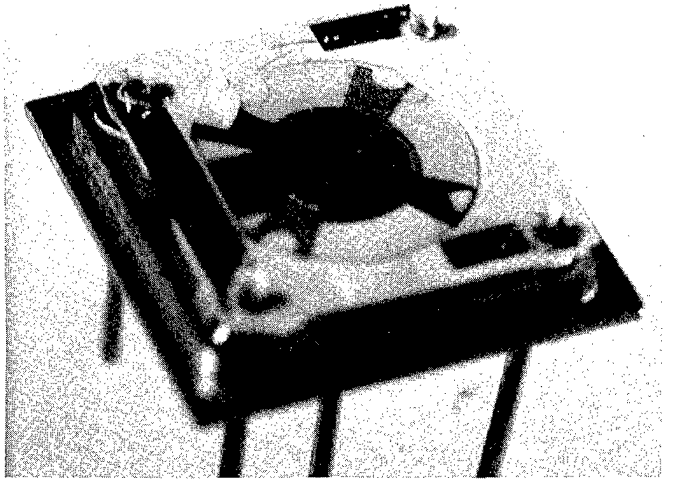
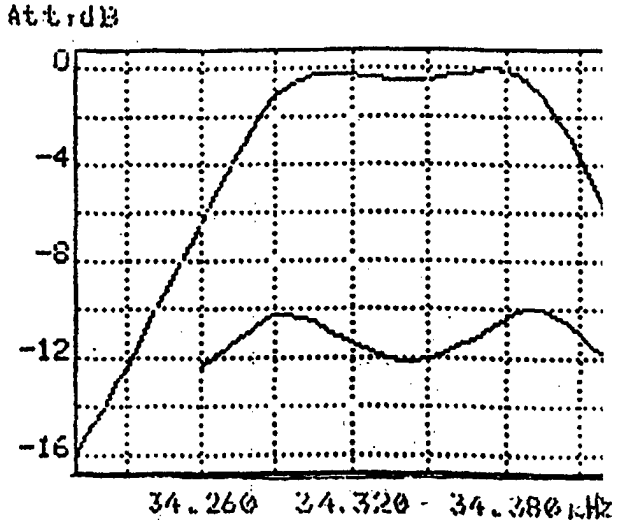


Fig.2. The filter ФП2П4-486 (outer sight).

FP2P4-486
1 crystal



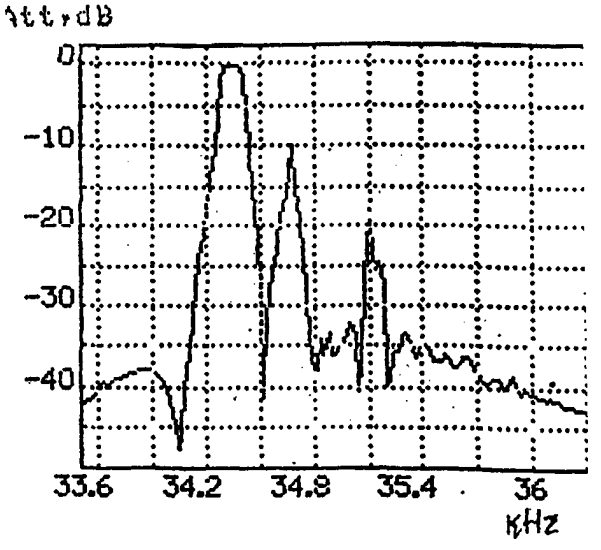
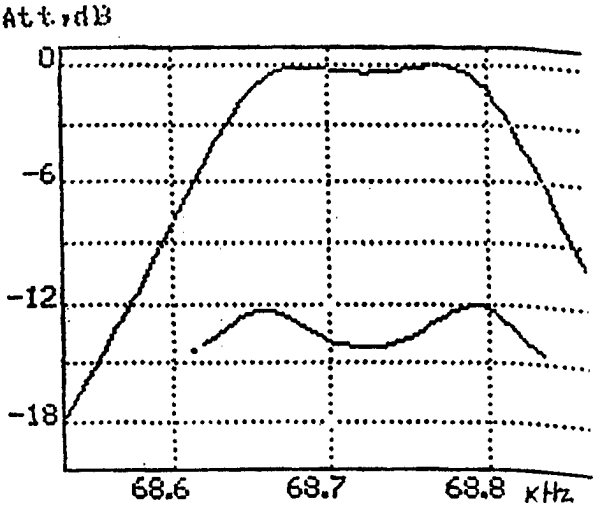


Fig.3 Attenuation frequency characteristic (AFC) of MCF - 486 at mid-frequencies 34338 кHz

FP2P4-486-03
1 crystal



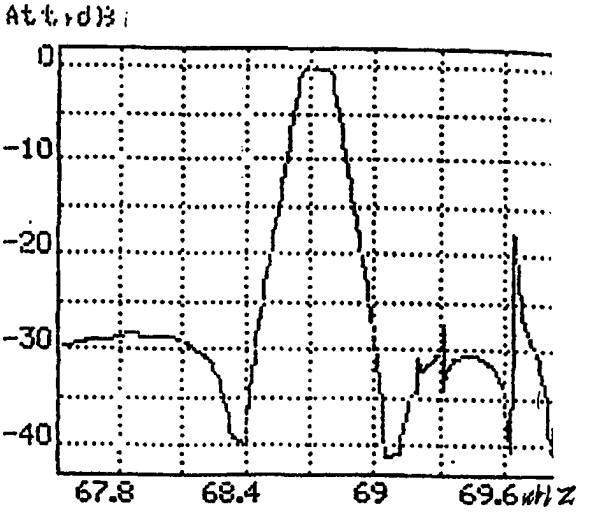
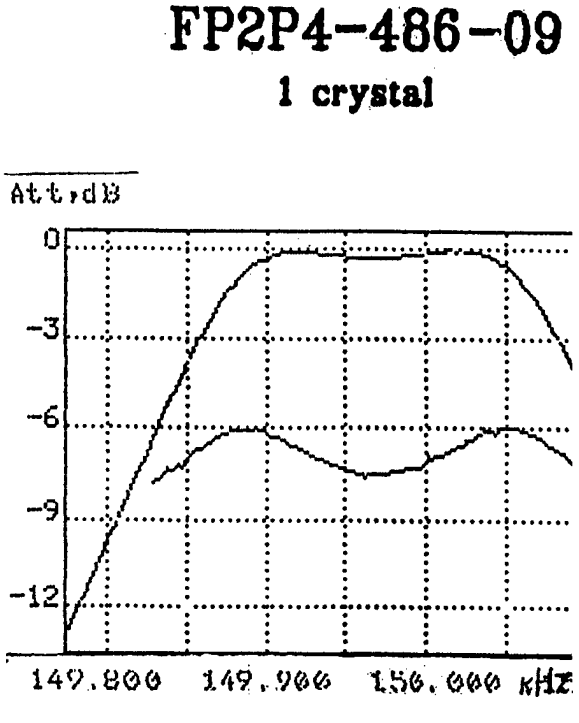
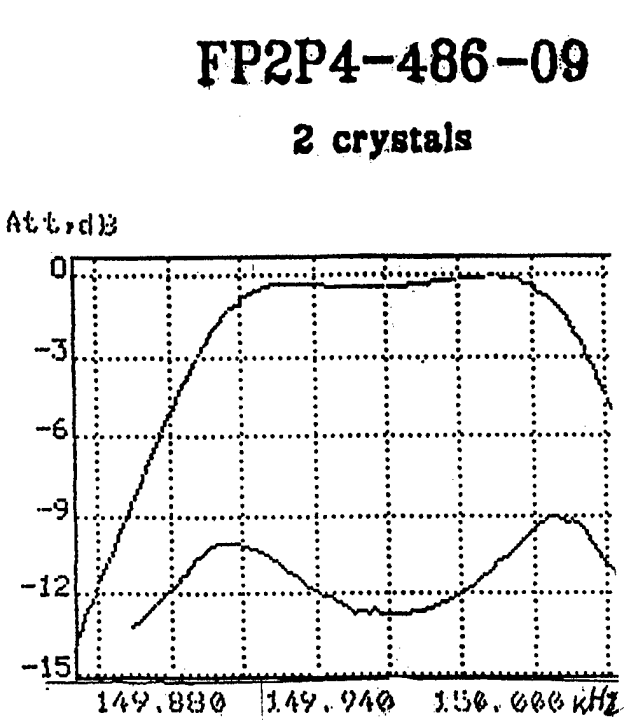
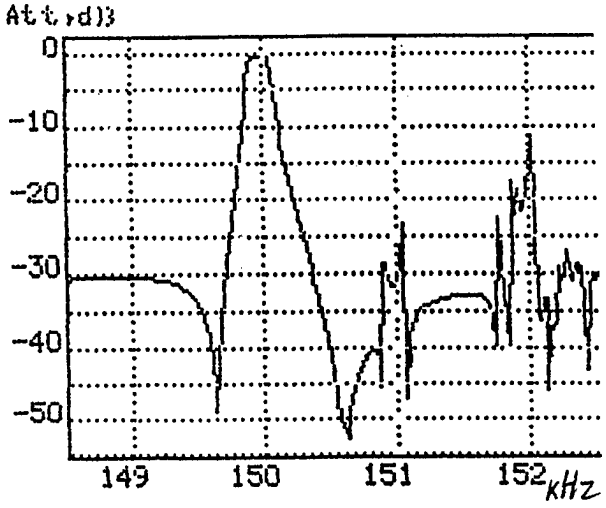


Fig.4 Attenuation frequency characteristic (AFC) of MCF - 486 at mid-frequencies 68736 кHz







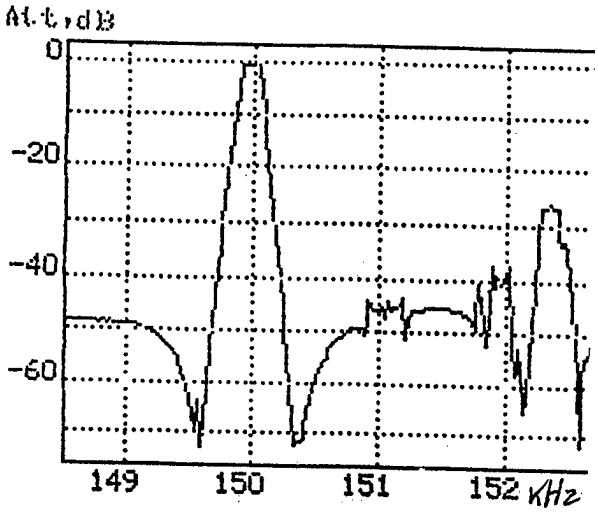


Fig.5 Attenuation frequency characteristic (AFC) of MCF - 486 at mid-frequencies 149950 kHz.

Fig.6. Attenuation frequency characteristic of MCF ФП2П4-486 of the 4th order at the mid-frequency 149950 kHz.

References (in Russian)

1. Ivanovsky G.F., Petrov V.S. Ion-plasma processing of materials. Radio and svyaz, 1986, s. 232.
2. USA Pat. N 4343827
3. Pashkov S.S., Pershina E.F., Ivanov V.V. et al. Frequency adjustment of crystal units by the method of ion etching the electrode material.- Elektronnaya technika, Ser.5. Radiodetaly i radiokomponenty. 1982, vyp.2(47), s.s.51-53.
4. Sevastyanov A.M., Pershina E.F., Lavrentzov V.D. et. al. Investigation of the operating regime of the frequency adjustment process for crystal units by the method of ion etching the electrode material. - Elektronnaya techika, Ser.5. Padiodetaly i radiokomponenty. 1982, vyp.2(47), s.s.53-55.
5. Pershina E.F., Ephanov A.S., Vatzuzo P.A., Mostyaev V.A. Adjustment of parameters of monolithic filters by ion etching the electrodes. - Electronnaya technika, Ser.5. Radiodetaly i radiokomponenty. 1988, vyp. 4(73), s.s.75-78.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

USE OF ACTIVE NETWORKS TO WIDEN THE SPECTRUM OF APPLICATION OF PIEZOELECTRIC FILTERS

A.T. Orlov and V.F. Skripov

Dept. of Microelectronics, Kiev Polytechnic Institute
37 Pobedy Ave., Kiev, Ukraine, 252056

This paper describes the design of active bulk acoustic wave (BAW) piezoelectric filters with properties that are unattainable with inductorless passive piezoelectric filters. Circuit realizations for wideband (fractional bandwidth of between 5% and 50%) bandpass filters (BPF), low pass filters (LPF), highpass filters (HPF), band reject filters (BRF), and tunable filter sections are proposed. This class of filter eliminates the characteristic restrictions related to the properties of the piezoelectric material. The proposed approach would be useful for the design of integrated circuit (IC) compatible thin film BAW piezoelectric filters over a wide frequency range of up to 1-2 GHz.

1. INTRODUCTION

Filters occupy a large share of the communication equipment volume. In comparison with other circuits (amplifiers, mixers, etc.) they are still unyielding in terms of miniaturization and IC compatibility. One of the ways to miniaturize the filters is by using piezoelectric SAW and BAW monolithic filters.

BAW piezoelectric filters operating in the fundamental mode are usually used at frequencies below 30 MHz. Their merits are low loss and simple electrode topology. However, the bandwidth of inductorless BAW filters is limited to 1% for quartz and to about 5% for $LiNbO_3$ resonators.

SAW filters have played a key role in frequency selection and determination in the frequency range from 10 MHz up to 2 GHz due to their small size, outstanding selectivity, high bandwidth capability (50% or more) and high reproducibility. The one disadvantage that they have had has been the relatively high insertion loss, often making them inappropriate for RF input signal selection. Novel types of SAW filters with low insertion loss have been recently developed [1,2]. Some of these devices have employed ladder or coupled SAW resonators. In these cases the inductorless low loss SAW filter's fractional bandwidth have been restricted ($\leq$ 5% for $LiNbO_3$) and the selectivity has been determined by the number of resonators.

Ladder and coupled networks are also basic to BAW filters [3]. Therefore the low loss SAW filters would get preference if the SAW resonators will be advantageous in the comparison with BAW resonators at the same frequency region.

Recent progress in thin piezoelectric film deposition promises small size (as low as 100 μm^2 independently of frequency) BAW resonators integrated with semiconductor ICs [4]. High quality thin film BAW resonators with resonant frequencies in the region of 1-2 GHz have just been fabricated [5, 6]. They use sputtered ZnO, AlN or PZT sol-gel deposed films on both GaAs and Si substrates. Their electrodes have a small size and a topology easily implemented by standard 3-5 μm IC processes.

At the same time, the progress in IC compatible active filters has been rapid. Active frequency-selective and active inductance simulating circuits based on only a few transistors have been reported for frequencies up to 2 GHz [7]. Unfortunately, these circuits have both a low Q factor and poor temperature stability. They could, however, be successfully used as elements in active piezoelectric filters.

In this paper active networks based mostly on multiport piezoelectric resonators (MPR) are proposed to extend the range of applications for BAW piezoelectric filters. These filters combine the advantages of piezoelectric and active filters.

Most of the examples are based on piezoceramic resonators. Piezoceramic filters have limited application because of the fractional bandwidth restrictions ranging from 1% to 8%, due to poor temperature stability and low electromechanical coupling coefficient k, respectively. We show that with our approach, the designer can overcome these severe restrictions imposed by materials with the use of active networks. The authors feel that the proposed approach will be useful for integrated thin film piezoelectric filters design.

0-7803-1945-1/94 $3.00 © 1994 IEEE

2. CASCADE REALIZATION

One way of designing active piezoelectric filters is by using a cascade realization of the transfer function. This is very similar to the design of active RC filters where it is necessary to combine stages with different poles and zeros and with different Q factors. MPR materials have fixed electromechanical properties and therefore require additional Q factor and frequency tuning. It is possible to incorporate special sections in the MPR for tuning purposes (fig.1). Passive tuning methods allow us to reduce the frequency of the MPR by less than 2% using a loading capacitance. A load resistor reduces the Q factor and the frequency simultaneously, the latter by a small amount because of the fixed electromechanical coupling coefficient.

The use of active networks as a loading impedance allows us to influence the Q factor and the frequency independently over a wide range of values. Among these active networks are negative capacitance and simulated inductance. Negative capacitance circuits and simulated lossy inductors, based on a single operational amplifier, allow us to change the resonant frequency by up to 20% and to reduce the Q factor by up to 100 times, respectively.

Independent frequency and Q factor transfer function pole control is achieved in an active network by applying a feedback signal to the additional MPR section through a phase-shift network (fig.2). A MPR centre frequency tuning range of ±3% is achieved while maintaining both a constant Q factor and a constant gain. This range depends on the gain of the feedback loop. Experiments have shown that the Q factor, originally at 500, can be varied over a range of 70 to 2600 [8].

In the case of high order BPFs, a basic 6th order circuit is suggested [9]. It uses a differential amplifier, which allows you to balance the response of each of the MPRs and to compensate the through capacitance of the MPR. MPRs with a Q factor are used to form the high steepness in the BPF transition regions. An active RC circuit with a low Q factor (of few units) smoothens the passband shape. Feedback loop further minimizes passband ripple.

The active BPF, designed using this circuit configuration, requires fewer active elements, and has an available fractional bandwidth in the range of 3 to 50%. Such filters have high temperature stability, low sensitivity and wide dynamic range compared to those of their active RC counterparts. The reason for this is that active RC circuits have a low Q factor. Besides, the amplifier is not used to improve the filter selectivity, as is done in high Q factor active RC filters. An example of an 8th order active piezoceramic BPF response is shown in fig.3.

3.INDUCTANCE SIMULATION AND MPR IMPEDANCE TRANSFORMATION

Lattice networks are useful for the realization of minimum phase filter networks. They have the advantage of permitting the independent design of the phase frequency characteristic [10]. MPRs are very convenient components for such network realisations without the need of real transformers. Wideband BPFs have been implemented with grounded inductors simulated by one or two operational amplifiers (fig.4). Losses of the widening inductors in that case can be accounted for by the loading resistors. By using simulated inductor networks, new classes of the active piezoelectric LPF, HPF and BPF are created. Their prototypes are LC lattice networks with piezoelectric resonators in the arms. MPRs used there define the transition region in the filter frequency response, with the sharpness of the response is conditioned by the quality of the piezoelectric resonator material. Since in the case of BPFs the demands for the simulated inductor are not high, the stability of such filters is mainly determined by the stability of MPR. Examples of active piezoelectric BRF, LPF and HPF based on MPRs and simulated inductors are shown in figs. 5 and 6.

Another method for designing active piezoelectric filters is to exploit the piezoelectric resonator impedance transformation. Examples of new circuits that use a gyrator and a frequency dependent negative resistor (FDNR) for the design of filters based on one-port piezoelectric resonator are shown in figs. 7 and 8, respectively. It is important, that resonant frequency and its sensitivity be mostly determined by the piezoelectric resonator. An example of a FDNR circuit transformation using a passive ladder prototype is shown in fig. 9. The LPFs and HPFs that we have built with this method have steep and stable frequency responses.

The proposed approach to the design of active piezoelectric filters has been shown to be applicable for the frequency range from 1 kHz to 10 MHz using different types of the piezoelectric resonator vibrational modes. MPRs for a higher frequency region of up to 1 GHz could be fabricated with multilayer piezoelectric thin film deposition. High value small size capacitors could be made using the same process, exploiting the high dielectric constant of the film.

4. CONCLUSION

The new approach to the design of active piezoelectric filters allows us to create wideband BPF, HPF, LPF and BRF with stable and steep frequency responses. Using this approach (with the latest achievements in thin piezoelectric films deposition), one can obtain IC compatible filters with good characteristics. They could compete with LC, active RC and switched capacitor filters at low frequencies as well as with SAW, dielectric resonator, stripline and other filters at the microwave frequencies.

ACKNOWLEDGEMENT

The authors wish to thank Dr. Peter Smith from McMaster University (Canada) for fruitful discussions.

REFERENCES

1. M. Hikita, T. Tabuchi, and N. Shibagaki, "Investigation of New Low-Loss and High-Power SAW Filters for Reverse-Frequency-Allocated Cellular Radios", *IEEE Trans. UFFC*, vol. 40, no. 3, pp. 224-231, May 1993.
2. Y. Xu and P.M. Smith, "A Theoretical and Experimental Study of Waveguide Coupled SAW Resonator Filters", Submitted to *IEEE Trans.UFFC*, March 1994.
3. K.M. Lakin, G.R. Kline,and K.T. McCarron, "Thin Film Bulk Acoustic Wave Filters for GPS", P*roc. 1992 Ultrasonics Symp.*, pp. 471-476, 1992.
4. G.F. Zhang and J.L. Gautier, "Broad-Band, Lossless Monolithic Microwave Active Floating Inductor", *IEEE Microwave and Guided Wave Letters*, vol. 3, no. 4, pp. 98-100, April 1993.
5. M. Sayer,"Piezoelectric Thin Films Devices", *Proc. 1991 Ultrasonics Symp.*, pp. 595-603, 1991.
6. Z. Awang and R.E. Miles, "Gallium Arsenide Bulk Acoustic Wave Resonators Fabricated Using Sol-Gel Technology", *Electron.Lett.*, vol.. 29, no.7, pp. 626-627, 1993.
7. G.R. Kline and K.M. Lakin, "1.0-GHz Thin Film Bulk Acoustic Wave Resonators on GaAs", *Appl. Phys. Lett.*,vol. 43 (8), 15 October 1983.
8. A.T. Orlov, D.V. Molodchik, V.N. Khorungij, "Methods of Tuning of the Piezoceramic Filter Stages", *Piezoelektricheskie Materiali i Preobrazovately*, Rostov-na-Donu, no. 7, pp. 97-101 ,1988 (in Russian).
9. A.T. Orlov, Yu.I. Yakimenko, S.F. Klochko, "Band-Pass Filter", Patent (USSR) no. 1096756.
10. A.T. Orlov, Yu.I. Yakimenko, "Details on the Design of Lattice Band-Pass Filters Based on Piezoelectric Transformers", *Vestnik KPI*, Serija "Radioelektronika", no. 23, pp. 92-94 , 1986 (in Russian).

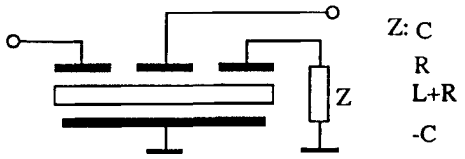


Fig. 1. Using special section of MPR for tuning.

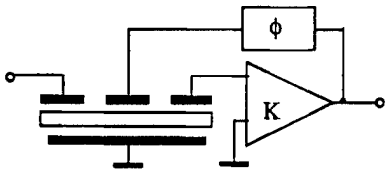


Fig. 2. Simple stage for BPF with active tuning.

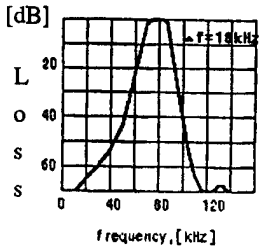


Fig. 3 8th order active piezoelectric BPF frequency responce

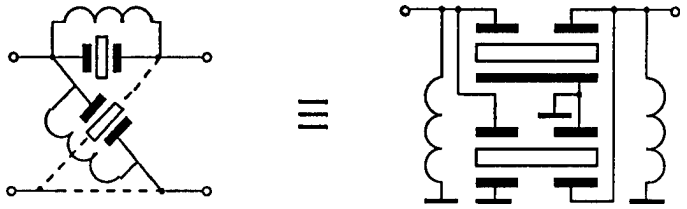


Fig 4 Lattice wideband BPF transformed for MPR active realization

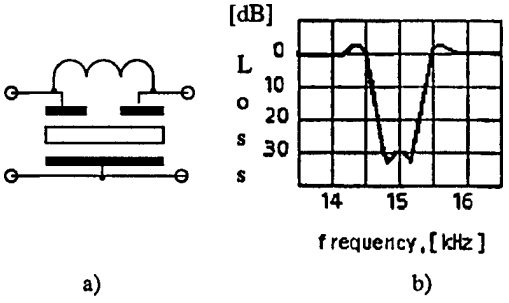


Fig. 5 Active BRF structure and frequency responce

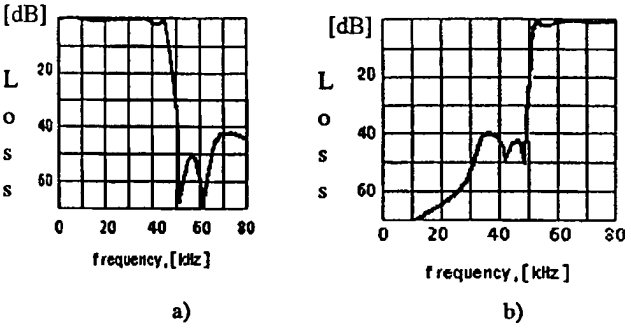


Fig. 6 Active piezoelectric LPF (a) and HPF (b) responces

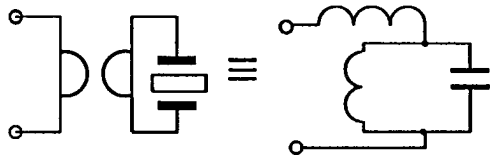


Fig. 7 Gyrator transformation of the one-port piezoelectric resonator impedance

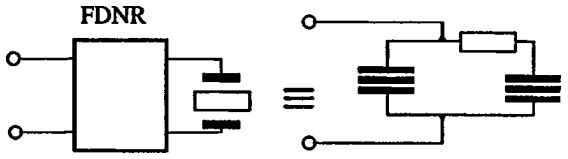


Fig. 8 FDNR transformation of one-port piezoelectric resonator impedance

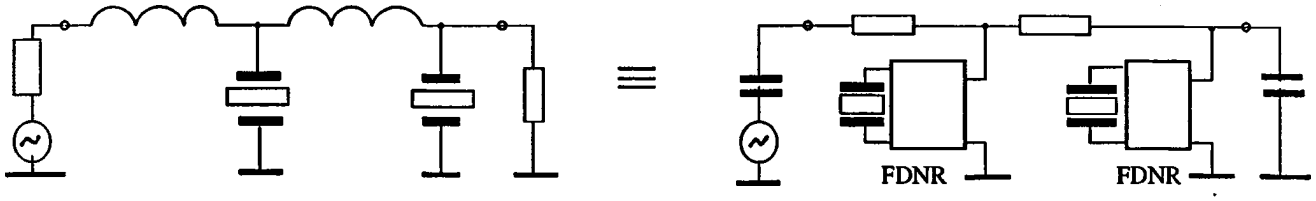


Fig. 9 Example of the filter circuit FDNR equivalent transformation

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

MONOLITHIC QUARTZ STRUCTURE VIBRATING BEAM ACCELEROMETER (VBA)

William C. Albert
Crystal Gage, Inc.
Boonton, NJ 07005

ABSTRACT

This paper describes an accelerometer for which the vibrating beam sensor, the vibration isolation system, the proof mass, the proof mass suspension system, and the mounting means are all machined from a single rectangular quartz crystal blank. As such, this sensor represents a significant advancement in combining the sensor art with the quartz crystal resonator art. The design uses the well-known force-frequency modulation effect of a vibrating beam in axial tension or compression.

INTRODUCTION

The use of frequency-modulated vibratory sensors has always been attractive in accelerometer as well as other sensor applications due to the inherent digital nature of the output signal without need for A/D conversion. When the sensor material used is quartz crystal, high stability of bias frequency and scale factor is realized as well as low temperature sensitivity. In addition, the piezoelectric property of quartz crystal provides a simple means of sustaining vibrations using an oscillator circuit, electrically connected to electrodes plated on the resonator.

A unique feature of the patented design is its two thickness structure. This allows the mass, mass suspension and mount requirements of the relatively thick outer structure to be independent of the sensitivity requirements of the relatively thin vibrating beam inner structure. This simple yet powerful design approach allows the concept to be applied to both high g low sensitivity and low g high sensitivity applications. A plated metalization pattern
leads from the mount portion of the structure to electrodes on the vibrating beam. A simple piezoelectric shear drive is used to sustain the beam flexure vibrations.

The monolithic quartz structure is "mini-machined" using ultrasonic machining, as opposed to photo-etch micro-machined techniques used on DETF quartz or silicon sensor applications. Since the structure is a .1" x .6" x .7" rectangular "quartz sensor chip", it lends itself to IC packaging and sealing techniques. The oscillator circuit may be packaged along with the "chip" or remain external. Depending on the application, many packaging options are possible.

The use of quartz resonators in sensors (primarily pressure sensors and accelerometers) is not new. However, previous sensor designs used assemblies of several quartz parts or a combination of quartz and metal parts. The design being presented uses a monolithic quartz structure.

The following paragraphs will describe the construction and operation of the quartz structure beginning with the innermost vibrating flexure beam, through the Q-maintaining isolation mechanism and end with a description of the proof mass and mounting means of the outer portion of the structure. The advantages of the unique feature of having the thickness of the inner and outer sections markedly different will also be discussed. The piezoelectric drive will then be covered, followed by a description of packaging. All design aspects of this sensor stress simplicity and producibility for a high performance-to-cost ratio.

Since the paper concentrates on quartz resonator design, machining and packaging, details on performance will not be discussed. An exception is temperature sensitivity performance because it differs from what is normally experienced from quartz flexure resonators.

DESCRIPTION

Figure 1 illustrates the basic sensing means which is the well-known force-frequency (F-f) effect of a flexure beam in axial tension or compression. Figure 1 also describes the vibratory distinction between a beam and a string, specifically that the beam has a no-load resonant frequency and can also respond to compression, as well as the force effect on the crystal equivalent circuit. A more detailed discussion of the F-f effect appears in the Appendix.

0-7803-1945-1/94 $3.00 © 1994 IEEE

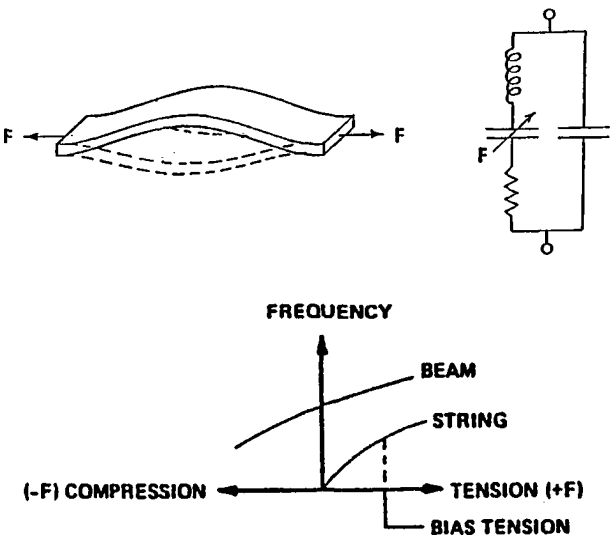


FIGURE 1
FORCE-FREQUENCY (F-f) EFFECT
OF A
VIBRATING FLEXURE BEAM

Without a means isolation to maintain a high Q, the beam of Figure 1 is not practical. A typical simple isolation means is shown in Figure 2 and operates as follows. The vibrating beam is actually a virtual fixed-fixed beam, since there is some linear and angular displacement at the beam roots. By design, this root motion is kept small and is on the order of 1% of the maximum linear and angular displacements. Since the isolator mass reaction to M and V is primarily inertial, the relative phasing is as shown in Figure 2. The overall result is that the M_{MT} and V_{MT} reactions at the outer structure will be much less than M and V respectively which prevents energy loss. Typical Q is on the order of 20,000 or more.

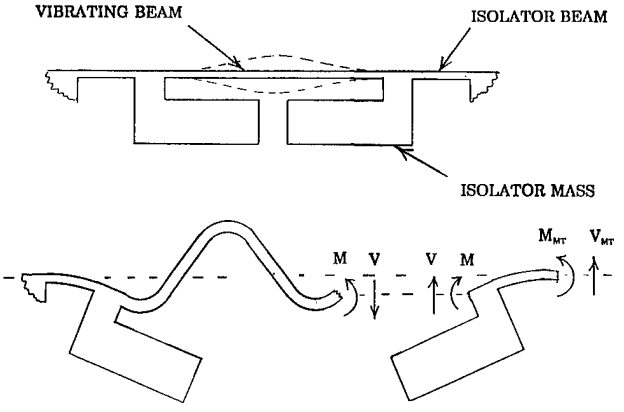


FIGURE 2
VIBRATION ISOLATOR TO
MAINTAIN HIGH Q

The vibrating beam and its isolator system comprise the inner structure of the overall structure shown in Figure 3. Note that without the inner structure in place, the flexure beams would permit x-axis motion of the proof mass with respect to the mount portion of the structure. Therefore, the sensing portion inner structure is placed in tension or compression in response to an acceleration along the input axis. The use of a mount appendage insures that the error producing mounting stresses are not transmitted to other portions of the structure. Successful mounting has been achieved using epoxy, solder and mechanical fasteners.

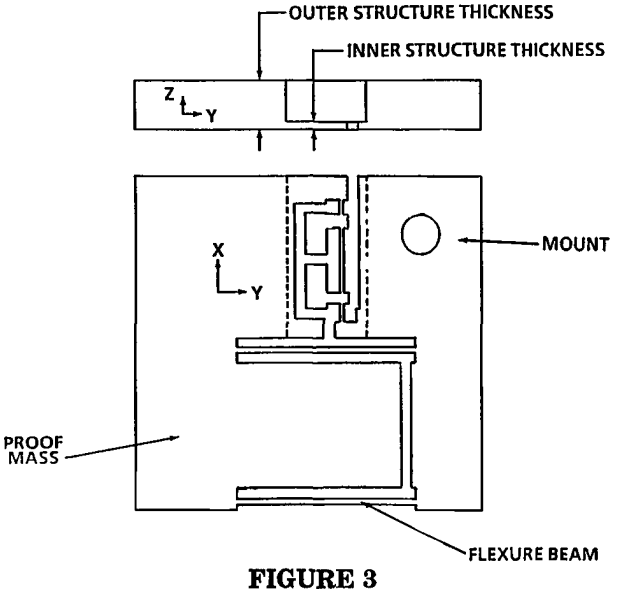


FIGURE 3
MONOLITHIC QUARTZ STRUCTURE

A key feature of the design is the marked difference between the thickness of the outer and inner portions. The following comments apply:

- For almost all applications, a single thickness structure is not practical because for a reasonable overall geometry, the sensitivity and isolation requirements of the inner structure cannot be made compatible with the mass, mounting and cross-axis stiffness requirements of the outer structure.
- Having two different thicknesses introduces a high degree of design flexibility that completely decouples the relative design requirements of the two portions so that high g, low g and high cross-axis g requirements can all be accommodated.
- The outer to inner thickness ratio is typically 8:1.

The various fabrication steps leading up to a completely machined and metalized structure are illustrated in

Figure 4. Ultrasonic machining is the primary fabrication means. Overall dimensions are .1 x .6 x .7 inches.

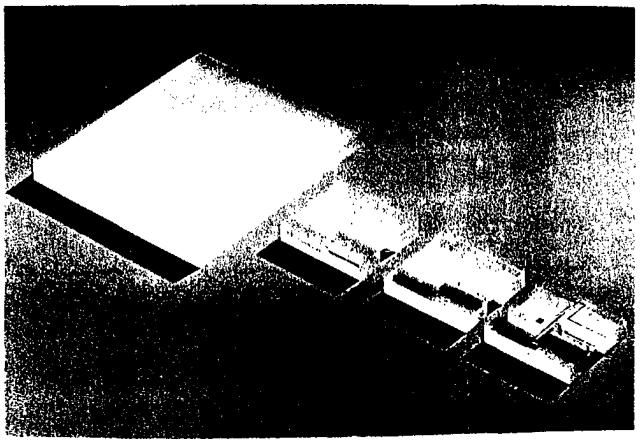


FIGURE 4
FABRICATION STEPS
QUARTZ BLANK TO METALIZATION

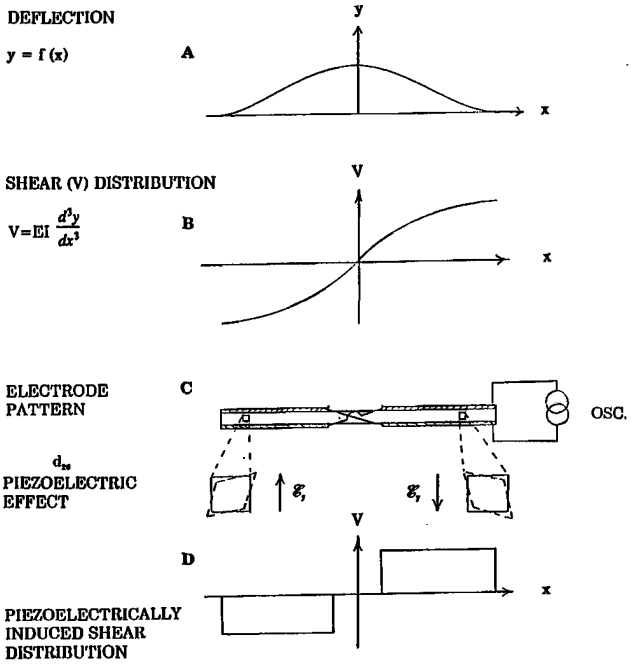


FIGURE 5
PIEZOELECTRIC FLEXURE DRIVE

FLEXURE DRIVE AND METALIZATION PATTERN

The piezoelectric flexure drive uses the well known Y-cut shear drive[1] described in Figure 5. It's operation is as follows:

- The displacement distribution of a fixed-fixed fundamental flexure vibration mode is shown in Figure 5-A.
- From the third derivative of y with respect to x, the shear stress distribution is shown in Figure 5-B.
- Figure 5-C shows the electrode pattern along with the piezoelectric shear effect. This drive uses the d_{26} piezoelectric coefficient whereby an electric field in the y direction causes a shear stress about the z axis.
- Figure 5-D shows the piezoelectrically induced shear stress distribution. Note that it is in phase with Figure 5-B and therefore will drive the beam at its fundamental frequency when connected to an oscillator.

This type of drive was selected because it is simple and the electrodes are easily placed on the Y surface of the thin (0.004 in.) beam. Because of the thin ribbon-like geometry of the beam, a direct X-cut flexure drive electrode pattern was considered too difficult to apply using conventional masking and plating techniques. Because the shear strain energy is small compared to the flexure strain energy, the electrical-to-mechanical coupling is low for the shear drive, resulting in a relatively high R_m on the order of 0.5 megohms.

Figure 6 illustrates how the deposited metalization on the Z surface of the outer structure leads to the

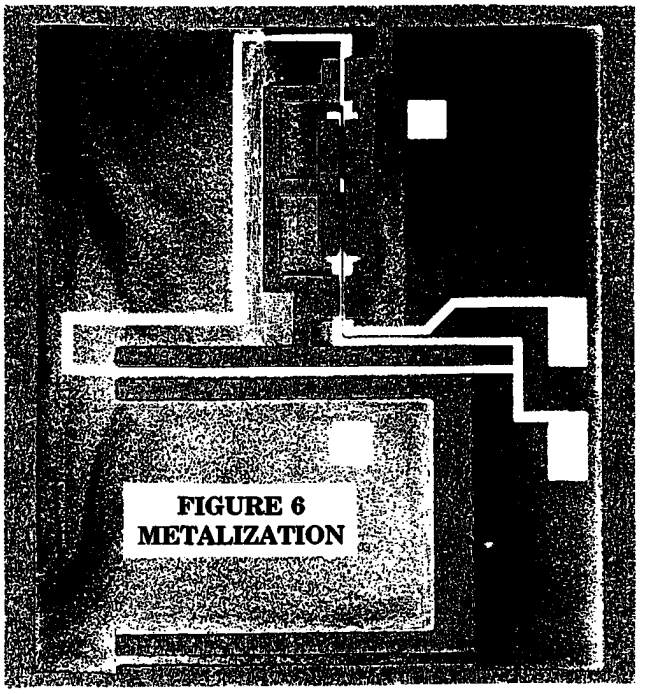


FIGURE 6
METALIZATION

electrodes, which are deposited on the Y surfaces of the vibrating beam. The electrode phasing is as shown in Figure 5-C.

PACKAGING

Since the quartz structure can be thought of as a rectangular "g-sensitive chip", it lends itself very nicely to IC packaging techniques as shown in Figure 7. IC packaging techniques are well developed and use low cost formed enclosures with very reliable glass-to-metal feedthrough seals. The seam weld or solder lid sealing techniques are also high volume low cost operations. The packages are usually evacuated but can also be back filled with a partial atmosphere of inert gas.

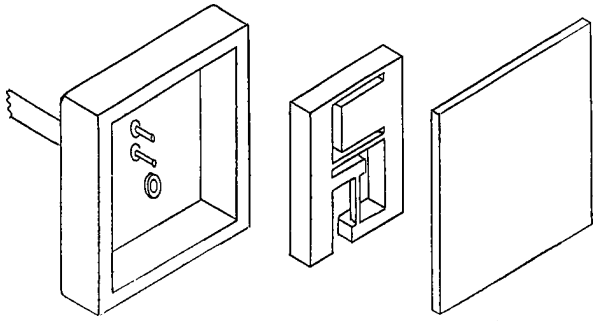


FIGURE 7
"g-SENSITIVE CHIP" PACKAGING

While for some low accuracy applications the package of Figure 7 can be used directly, VBA's usually employ two units in a push-pull arrangement so that an input acceleration places one beam in tension and the other in compression. The output is taken as the difference frequency. This dual beam mechanization and the reasons for its use are described in the Appendix.

A typical dual beam package arrangement is illustrated in Figure 8. Because of the versatility of the basic IC packaging technique many other arrangements are also possible.

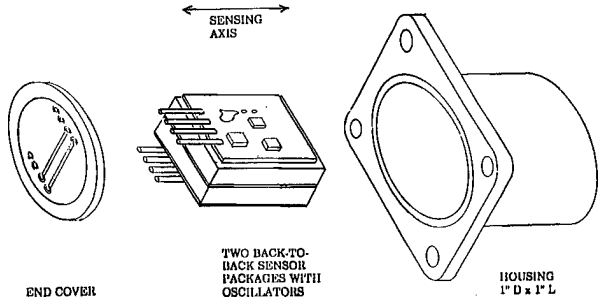


FIGURE 8
DUAL BEAM VBA PACKAGE

MONOLITHIC VBA CHARACTERISTICS

The instrument described differs from conventional accelerometers and also many other VBA designs in the following ways:

- The basic sensing element is not an assembly and therefore error producing sources such as joints and dissimilar materials are eliminated.
- The sensor design and packaging are very simple and the overall parts count is very low for low cost.
- The quartz sensor operates in a vacuum and is therefore undamped. This greatly simplifies the design and construction by eliminating critical damping gaps and the error sources associated with squeeze film damping. All structure resonances are designed to be well above 2 kHz. Vibration tests on engineering prototypes to 15 g and 2 kHz along all three axes revealed no resonances and no failures.

It is the small size of the quartz structure that results in the very high resonant frequencies and eliminates the need for a damping mechanism. It is well known that the resonant frequency and shock capability of a given structure design geometry is generally inversely proportional to its size. Future generation monolithic VBA designs will be even smaller.

Other characteristics of the individual resonator, as well as the dual beam VBA appear in Table 1.

TABLE 1
TYPICAL CHARACTERISTICS

INDIVIDUAL BEAM

FREQUENCY (f_o)	**52 kHz**
SCALE FACTOR (k_1)	**26 Hz/g**

DIFFERENCE FREQUENCY

BIAS ($\Delta f_o/2k_1$)	**10 g Per % f_o Mismatch**
SCALE FACTOR	**52 Hz/g**
K_2	**1.3 $\mu g/g^2$ per % k_2 Mismatch**
BIAS TC	**50 μg/°C per °C Turnover Mismatch**
SCALE FACTOR TC	**10 PPM/°C**

PERFORMANCE

The dual beam VBA is intended for applications requiring 1 mg bias and 100 ppm scale factor stability and input up to 50g. Single beam devices can be used in less stringent applications. With better quartz and quartz processing along with a more sophisticated oscillator and temperature control, there is no reason why the design cannot qualify for 10 μg and 1 ppm applications. More details on performance will appear in future papers.

One performance characteristic of special interest was temperature sensitivity. The bias frequency versus temperature characteristics exhibited the classic parabola as illustrated in Figure 9. Turnover temperature was -15° C. Of significance was the observed second order sensitivity of -25 $(10^{-9})/°C^2$. For other flexure resonators (X-cut and Y-cut), this sensitivity is typically -45 $(10^{-9})/°C^2$. It is suspected that the second order temperature sensitivity of the outer structure, as well as that of the vibrating beam, combined to result in the lower than expected value.

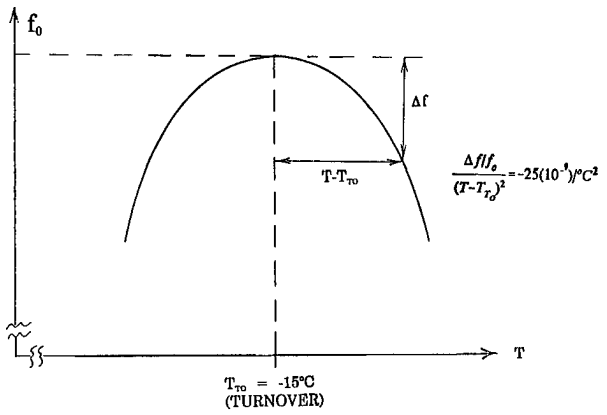


FIGURE 9
TEMPERATURE SENSITIVITY

APPENDIX

A previously published Rayleigh method solution of the frequency-force effect of a vibrating flexure beam resulted in the following equations [2] [3].

$$f=f_o\sqrt{1+SF} \tag{1A}$$

Where:

$$f_o=a_o\frac{t}{L^2}\sqrt{\frac{E}{\rho}} \tag{2A}$$

$$S=a_s\frac{L^2}{Ebt^3} \tag{3A}$$

Where:

The beam geometry, vibration mode, and boundary conditions are described by Figure A1.

E and ρ are respectively the elastic modulus and density of the beam material.

The dimensionless constant a_o and a_s are respectively 1.026 and 0.294 for f in Hz, the fundamental vibration mode, and the fixed-fixed boundary conditions.

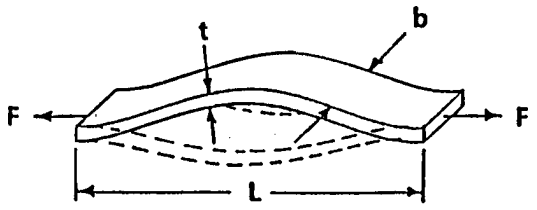


FIGURE A1
FIXED-FIXED VIBRATING BEAM
UNDER AXIAL FORCE

Another useful relationship is obtained by applying the binomial expansion to Equation 1A.

$$f=f_o(1+\frac{S}{2}F-\frac{S^2}{8}F\cdots) \tag{4A}$$

For an accelerometer application, F = mg where m is mass and g is acceleration. When the indicated substitutions are made Equation 4A becomes:

$$f=f_o+k_1g-k_2g^2\cdots \tag{5A}$$

Dual Beam Accelerometer Mechanization

Figure A2 is a schematic representation of dual beam operation where an input acceleration places one beam in tension and the other in compression.

Equations 6A thru 8A describe analytically and Figure A3 graphically the behavior of the individual beam and difference frequencies. (Note the additional subscript to denote beam number as well as coefficient order. Also note that higher than second order terms are neglected.)

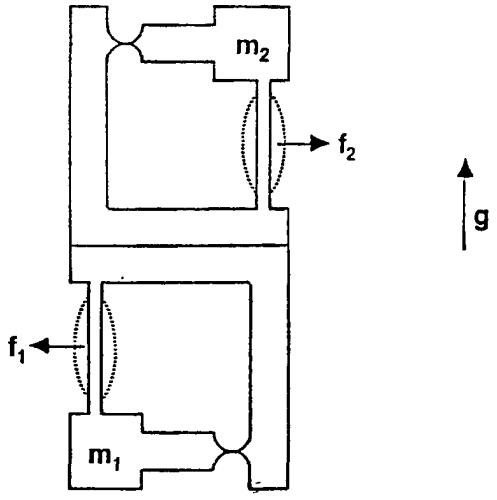


FIGURE A2
DUAL BEAM PUSH-PULL MECHANIZATION

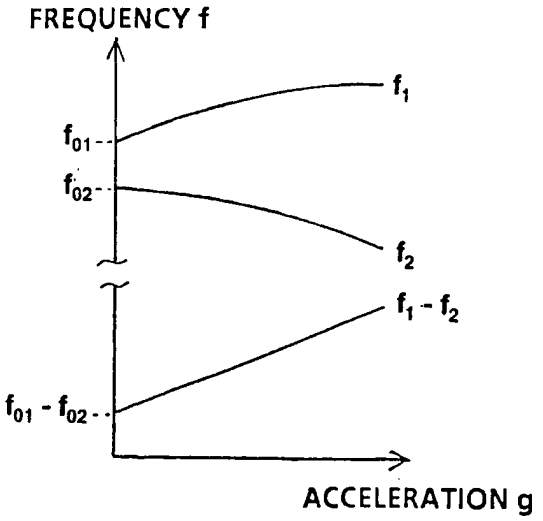


FIGURE A3
DUAL BEAM
FREQUENCY BEHAVIOR

For beam one in tension:

$$f_1 = f_{01} + k_{11} g + k_{21} g^2 \qquad (6A)$$

For beam two in compression:

$$f_2 = f_{02} - k_{12} g + k_{22} g^2 \qquad (7A)$$

The difference frequency now becomes:

$$\Delta f = f_1 - f_2 = f_{01} - f_{02} + (k_{11} + k_{12}) g + (k_{21} - k_{22}) g^2 \qquad (8A)$$

Frequency differencing achieves the following:

- The bias frequency (f_{01} - f_{02}) becomes small. More importantly, the benefit of common mode rejection (CMR) of bias frequency thermal drift is achieved since both essentially identical crystals will have similar temperature sensitivities. Other error sources are also reduced by CMR.

- The dominant first order scale factor (k_{11} + k_{12}) becomes double that of an individual resonator.

- The most dominant nonlinearity term (k_{21} - k_{22}) is greatly reduced by cancellation for improved linearity.

REFERENCES

(1) Cady, Walter G., Piezoelectricity. McGraw-Hill, New York, 1946, pp 448-449.

(2) Albert, W.C., "Force Sensing Using Quartz Crystal Flexure Resonators". Proceedings of the 38th Frequency Control Symposium, 1984, pp 233-239.

(3) Albert, W.C., "Vibrating Quartz Crystal Beam Accelerometer", ISA 28th International Symposium, Vol. 28, No. 1, 1982, pp 33-44.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

TEMPERATURE COMPENSATED SAPPHIRE RESONATOR FOR ULTRA-STABLE OSCILLATOR CAPABILITY AT TEMPERATURES ABOVE 77 KELVIN*

G. J. Dick, D. G. Santiago and R. T. Wang
California Institute of Technology, Jet Propulsion Laboratory
4800 Oak Grove Drive, Bldg 298
Pasadena, California 91109

Abstract

We report on the design and test of a whispering gallery sapphire resonator for which the dominant (WGH_{n11}) microwave mode family shows frequency-stable, compensated operation for temperatures above 77 kelvin. The resonator makes possible a new ultra–stable oscillator (USO) capability that promises performance improvements over the best available crystal quartz oscillators in a compact cryogenic package. A mechanical compensation mechanism, enabled by the difference between copper and sapphire expansion coefficients, tunes the resonator to cancel the temperature variation of sapphire's dielectric constant. In experimental tests, the WGH_{811} mode showed a frequency turn-over temperature of 87 K in agreement with finite element calculations. Preliminary tests of oscillator operation show an Allan Deviation of frequency variation of $1.4 - 6 \times 10^{-12}$ for measuring times 1 second $\leq \tau \leq$ 100 seconds with unstabilized resonator housing temperature and a mode Q of 2×10^6. We project a frequency stability 10^{-14} for this resonator with stabilized housing temperature and with a mode Q of 10^7.

1 Introduction

A sapphire whispering gallery resonator consists of a wheel or disk of sapphire inside a cylindrical metallic shielding can. By confining resonating rf fields to the sapphire element, these resonators effectively eliminate metallic conduction losses – and so make possible resonators which are only limited by performance of the sapphire itself. The sapphire is typically oriented with its crystal c–axis along the central axis of the container in order to achieve cylindrical symmetry for the excited resonance modes.

Whispering gallery electromagnetic modes can be divided into families depending on their field configuration, and further characterized by the number of full waves n around the perimeter of the sapphire element. The modes are doubly degenerate, with azimuthal phase of the two sub–modes differing by 90 degrees. Modes typically used are the WGH_{n11} family for wheel resonators and the WGE_{n11} family for flat disks, where $n \geq 5$.

With very high microwave quality factors (Q's) at easily reached cryogenic temperatures, the sapphire resonators already make possible excellent phase noise performance. In principle, the high Q values also make possible high frequency stability – if the resonator itself were stable[1–6]. However, temperature fluctuations in the sapphire cause unwanted frequency fluctuations. If these frequency variations could be cancelled or *compensated*, high stability could be achieved.

The resonator Q's increase rapidly as the temperature is cooled, from approximately Q = 300,000 at room temperature to 30 million at 77 Kelvin (for X-band frequencies $\approx$ 8 GHz). This compares to Q values of 1 to 2 million for the best available crystal quartz oscillators, and 10,000 to 20,000 for metallic microwave cavities. When coupled with low noise microwave circuitry, the high sapphire Q could make

*This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

0-7803-1945-1/94 $3.00 © 1994 IEEE

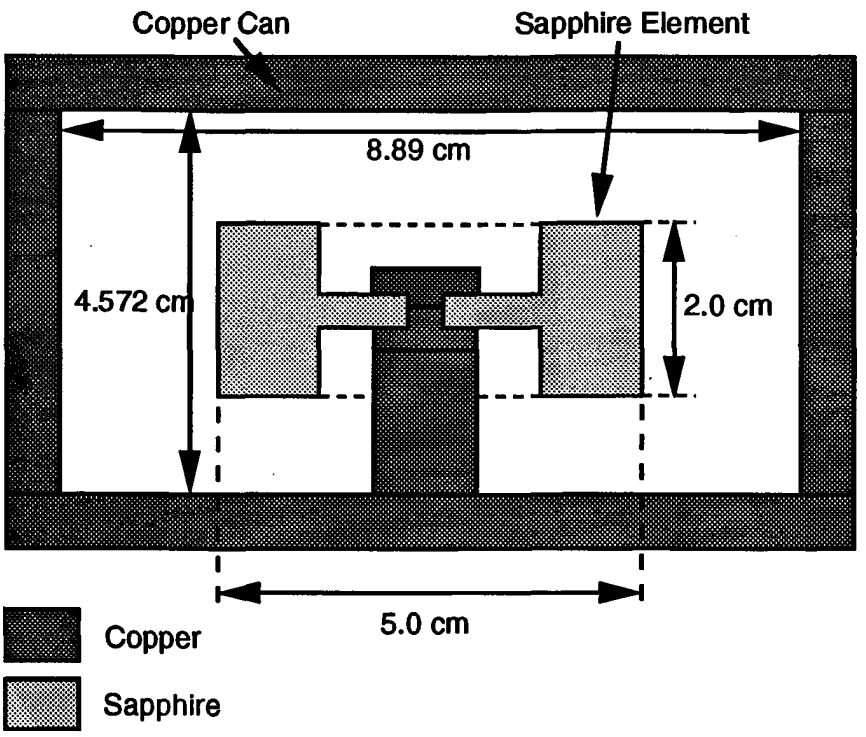


Figure 1: Whispering gallery sapphire resonator (cylindrical cross section)

possible a frequency stability as low as 10^{-14} [1]. Such a stability would be 20 times better than that achievable by quartz oscillators of the highest quality, which presently give a stability of 2×10^{-13}.

Finite element calculations of frequencies and frequency tuning rates with temperature were made using the CYRES 2-D program [see companion paper, this conference]. A comparison of calculated and measured tuning rates at 77K shows excellent agreement. In particular, both calculation and experiment show a weakening of the compensation effect for higher mode numbers that was not predicted by a simple circuit model that was initially used to design the resonator.

2 Background

Various approaches have been developed to reduce the thermal variation in electromagnetic or acoustic resonators and so achieve high frequency stability. Compensated operation for bulk acoustic wave (BAW) quartz oscillators is achieved by means of an appropriate choice of orientation for the quartz crystal. This is possible due to a very strong variability of acoustic parameters with crystal direction. The *electromagnetic* sapphire resonators have a much smaller anisotropy ($\approx 35\%$) and no sign reversal for any of its thermal dependencies. In fact, up to the present time useful compensation of sapphire resonators has only been possible at liquid helium temperatures, where incidental or added paramagnetic impurities give an effective compensating effect[2,7]. But helium temperature operation is expensive, and impractical for most applications. A compensation mechanism at 77 kelvin or above would allow liquid nitrogen could be used as the coolant in a very much smaller and less expensive system.

Temperature sensitivity of the operating frequency is characteristic of all electromagnetic and acoustic (piezoelectric) resonators due to thermal variation of the size, dielectric constants, speed of sound, etc. for solid state materials. Variation of these parameters is typically parts in 10^4 to 10^5 per Kelvin. Thus, achieving resonator stabilities of 10^{-13} to 10^{-14} would require *nano-degree* temperature stability – an impossible task.

Available techniques for higher stability and re-

duced thermal variation in resonator frequencies are:

- Very low cryogenic temperatures ($T < 10$ Kelvin) can be used to "freeze out" the thermal-induced variation, which varies as T^3 as the components are cooled. This technique has been successfully applied to superconducting[9], superconductor–on–sapphire[7], and sapphire whispering gallery resonators[2]. However, the very low temperature required makes such systems large and expensive, and therefore impractical for most applications.

- An inherently weak tuning mechanism may be used at the lowest temperatures to provide complete cancellation. In this way paramagnetic impurities can compensate the thermal variation in sapphire resonators for $T \leq 6$ Kelvin[1,2], but again, operation at such temperatures is impractical for most applications.

- The differing thermal coefficients for various properties of the resonator material can be played against each other in such a way that, for some operating temperature, thermal frequency variations are compensated or cancelled. Piezoelectric quartz resonators are compensated in this way by an appropriate orientation of this strongly anisotropic crystal (e.g. "SC" or "AT" cut quartz resonators) [10]. Unfortunately, an orientation–dependent cancellation does not occur for electromagnetic resonators where the anisotropy is much smaller (the temperature dependencies vary by only $\approx$ 30% as the orientation is changed).

- A resonator may be constructed using several similar materials with compensating thermal characteristics. For example, dielectric resonators for DRO's are typically stabilized by use of several materials with thermal dielectric variations of opposite sign [11].

- A mechanical tuning mechanism may be driven by thermal expansion coefficients of the construction materials. This mechanism has been previously applied to a sapphire resonator at room temperature using a highly reentrant geometry to achieve very low phase noise and a stability of 4×10^{-8} at $\tau = 10$ seconds [3]. Ultra–high stability was probably precluded by susceptibility of the design to thermal gradients. This is also the methodology of the present work.

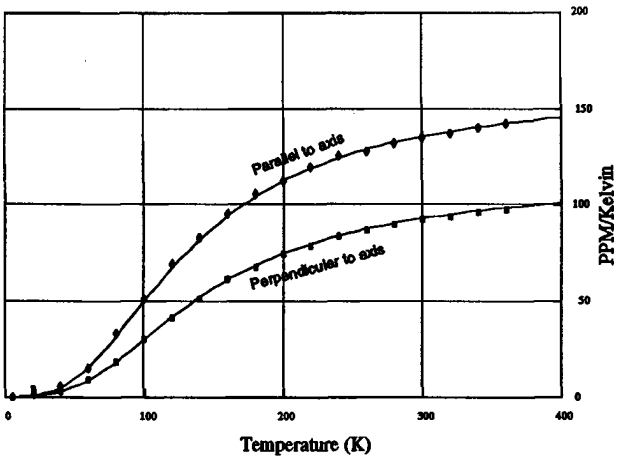


Figure 2: Temperature coefficient of the dielectric constant of sapphire for the components parallel and perpendicular to the c–axis.

3 Material Properties

Sensitivity of the sapphire resonator's frequency to temperature is due to several factors.

- Variation of the dielectric constants with temperature is the largest factor. As shown in Fig. 2, they vary by 80–140 parts per million (PPM) per Kelvin at room temperature (300 Kelvins)[12]. The resulting frequency change is just half this value, or 40–70 PPM/Kelvin (since $f \propto 1/\sqrt{\epsilon}$).

- The expansion coefficients of sapphire impact the frequency directly. As shown in Fig. 3 they give rise to a frequency change of 5–6 PPM/Kelvin.

- Thermal expansion of the copper containing can is a small but significant factor. Because microwave energy density at the walls is greatly reduced, (typically 100 to 10,000 times, to enable a high sapphire Q) the frequency sensitivity to can size is reduced by this same factor. Thus the 15 PPM/Kelvin copper expansion (Fig. 3) is reduced to 0.15 PPM/Kelvin or smaller [13].

Short term thermal stability of approximately 1 μKelvin can be attained at room temperature or at 77K [14]. However, even this very low variability, when coupled with a sapphire frequency sensitivity of $\approx 6 \times 10^{-5}$/K (room temperature) or $\approx 1.25 \times 10^{-5}$/K (77K)[13], gives frequency variations of 1 –

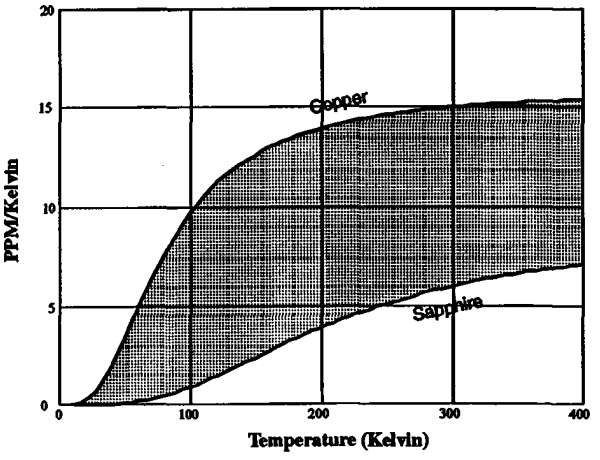


Figure 3: Thermal expansion coefficients for copper and sapphire. Differential motion (shaded area) can be used to tune a variable frequency resonator.

6×10^{-11}. These values are substantially worse than attainable with excellent quality quartz oscillators.

On the other hand, if this sensitivity could be compensated to first order, the remaining response is a second-order effect of $\approx 10^{-7}/\mathrm{K}^2$. If the temperature could be kept within only .01K of the compensation point, the remaining linear sensitivity would be $\leq 1 \times 10^{-9}/\mathrm{K}$. Coupling this sensitivity value with the achievable temperature stability of 1 μKelvin would result in a frequency stability of $\leq 10^{-15}$, a very attractive prospect.

Because there are no available internal compensation mechanisms that would give an effectively unchanging dielectric constant in the sapphire itself, we must look for solutions where one physical part of the resonator structure compensates for the effect of a different part.

In particular, if the thermal expansion profile of a second construction material were sufficiently different from sapphire, and if a mechanism could be found to use this difference to give a frequency variation of opposite sign to that of the sapphire itself, a practical compensated resonator could be constructed.

> A consequence of this type of design is that the various parts must be in excellent thermal contact with each other. For example, a temperature *differential* of 1μK, between the parts would give rise to the same (large) $1 - 6 \times 10^{-11}$ variation alluded to above.

> Fortunately, sapphire has one of the highest thermal conductivities for any solid material in the 77K – 300K temperature range, together with relatively low thermal mass. Thus, sapphire could be mated with some other high–conductivity material in a composite resonator with a very short thermal time constant and overall high conductivity to provide a structure with high immunity to internal temperature gradients.

Since sapphire's expansion coefficient is relatively small compared to most materials, a natural choice is for a second material with a *larger* expansion coefficient. Figure 3 shows a comparison between sapphire and copper – a likely candidate by virtue of its high thermal conductivity. The **difference** between sapphire and copper values, as shown in the figure is the driving force for our compensation mechanism. It is useful to compare this difference with the temperature coefficients of the dielectric constant as shown in Figure 2. Such a comparison shows that the compensation task is much easier at 77K than 300K, since, dielectric coefficient variations are strongly reduced as the temperature decreases from 300K to 77K, while the copper–sapphire expansion difference holds constant.

A comparison of the *magnitudes* of the two effects shows that a very effective tuning mechanism is required to achieve compensation. However, at 77K the task does not seem out of reach. Here the difference between sapphire and copper expansion coefficients is

$$\frac{1}{x}\frac{\partial x}{\partial T} = \sigma_c(77K) - \sigma_s(77K) = 7 \text{ PPM/Kelvin},$$

while the dielectric tuning effect (one half of the dielectric constant variation, averaging perpendicular and parallel components) is

$$\frac{1}{\omega}\frac{\partial \omega}{\partial T} = 13.5 \text{ PPM/Kelvin}.$$

Combining these two equations we find that the required tuning sensitivity is given by

$$\left|\frac{\delta\omega}{\omega}\right| = \frac{13.5}{7} \times \left|\frac{\delta x}{x}\right|. \qquad (1)$$

That is, differential thermal expansion between copper and sapphire could be used to compensate the dielectric constant variation in sapphire at 77 Kelvin if a mechanism could be found that is able to tune about twice (actually 13.5/7) as much as it moves on a fractional basis, comparing Hertz per Hertz with centimeters per centimeter.

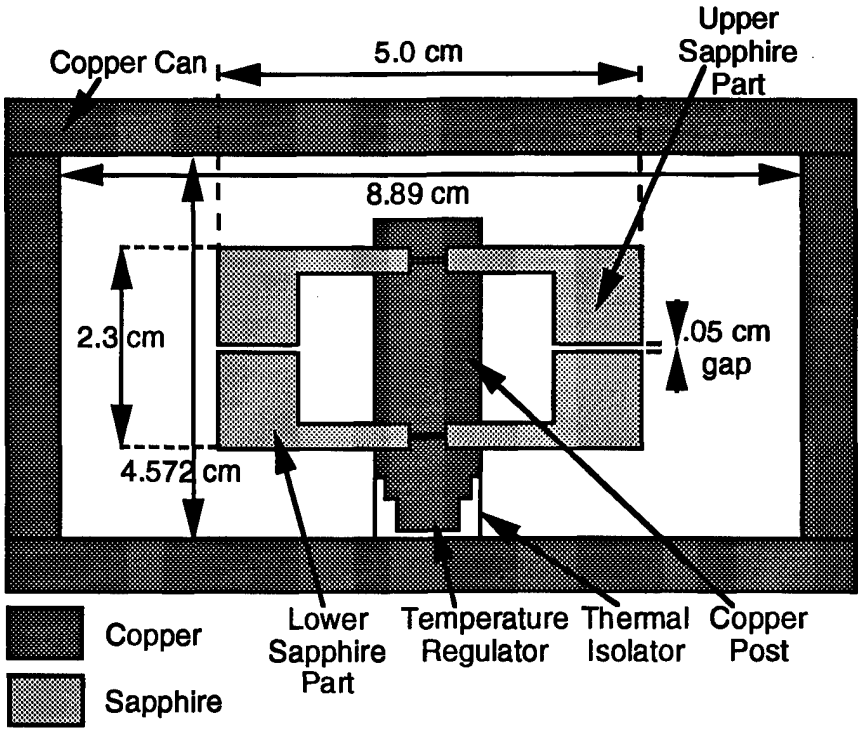


Figure 4: Compensated sapphire resonator.

It is worth noting that compensation at room temperature is much more difficult. It could be accomplished by the use of a material with a greater coefficient of expansion, such as zinc, or by the use of relatively extremely geometries. A comparison of Figs. 2 and 3 shows that increasing the temperature from 77 to 300 Kelvins increases the required tuning sensitivity by more than 4 times.

4 A Compensated Resonator

Figure 4 shows a composite microwave resonator that uses thermal expansion in an added non-sapphire tuning element to compensate for sapphire's thermal frequency variation. The resonator consists of two sapphire parts (each approximately half the thickness of a conventional whispering gallery resonator) separated by a copper post which has an expansion coefficient larger than that of the sapphire. If the gap between the two parts is small the resonant frequencies of some of the whispering gallery modes are strongly tuned as the gap spacing changes. However, for available materials, a weak tuning effect results if the post is only as tall as the gap spacing. Thus the sapphire parts must be made re-entrant, so that the post is approximately as tall as the entire resonator itself. In this case a strong thermal tuning effect, due to the *difference* between post-material and sapphire expansion coefficients can completely cancel sapphire's inherent frequency variation. The post can be made of copper for compensation at temperatures up to about 100 Kelvin, while materials with higher expansion coefficients (e.g. zinc) could be used up to room temperature.

In order to achieve high stability in an oscillator, the high Q of the sapphire resonator must not be degraded by the presence of the post. Furthermore, because the compensating tuning effects are due to physically distinct parts, thermal gradients must be minimized, requiring high thermal conductivity through the post. We find that a copper post of approximately 20–30% of the sapphire diameter provides the required thermal conductivity. An axial position for the post minimizes any Q degradation, since electromagnetic energy in the modes is concentrated near the outer perimeter of the sapphire disks.

The compensation mechanism can be understood

as follows. As the temperature is e.g. raised, the mode frequencies are *lowered* due to the increasing dielectric constant and thermal expansion of the sapphire (Figure 2). However, the gap is widened due to the large thermal expansion of the copper post (Figure 3). The resulting increase in vacuum gap volume (dielectric constant $\epsilon = 1$ compared to ≈ 10 for sapphire) tends to *raise* the frequencies. These cancelling effects can give rise to complete compensation at some temperature.

As previously discussed, the sapphire element itself is the primary temperature–dependent element in this resonator. Thus, the compensated central subassembly (consisting of sapphire parts and the copper post) is thermally isolated from the copper can, and held at a stabilized operating temperature above 77 K by action of a small heater and thermometer (not shown). The exact temperature depends on the experimentally determined turnover temperature for the sapphire resonator subassembly. The can's temperature must also be controlled for the highest frequency stability, even though its thermal sensitivity is 100 to 10,000 times reduced from that of the sapphire. This is accomplished by means of a second heater/thermometer feedback system (not shown) that stabilizes the can temperature to a value just slightly above 77 K.

4.1 Variation of mode Frequency with Gap Spacing

As indicated in Eq. 1, the copper/sapphire composite resonator requires a high tuning sensitivity in order to achieve compensation at 77 Kelvins. We present two different approaches to evaluate the sensitivity of the frequency of the fundamental $\mathrm{WGH_{n11}}$ mode to changes in a small gap at the resonator center plane.

- With a knowledge of the mode configuration, we can estimate this sensitivity using simple *circuit models* that incorporate the resonator dimensions in a natural way. This approach has the advantage of illuminating the qualitative features of the design problem.

- A *finite element* computational technique can be used to estimate mode frequencies, both with and without a gap. Accuracy of this methodology depends on the number of *nodes* used to characterize the geometry, with fields being evaluated only at the node points, and with the space between fitted to a simple power law behavior.

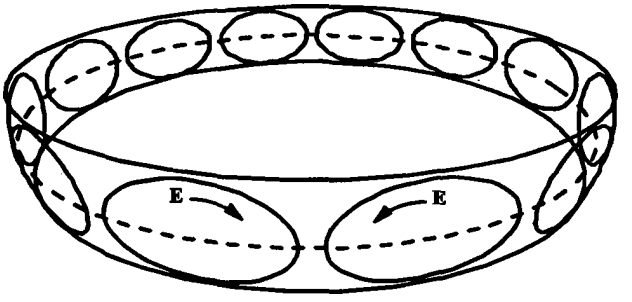


Figure 5: Electric field configuration for the WGH_{611} mode in a sapphire ring resonator. A horizontal gap at the dashed center line will cut the loops of electric field and strongly tune the mode frequency. Other modes, e.g. the WGE_{611} mode with horizontal loops, would not be strongly affected by such a gap.

We know that the magnitude of the tuning can be relatively large, as required by Eq. 1, because electromagnetic boundary conditions can give rise to larger energy in the gap region than in the (high ϵ) sapphire. In particular, this is true for modes with large electric fields perpendicular to the gap such as the $\mathrm{WGH_{n11}}$ mode family chosen for our application.

4.1.1 Gap Sensitivity Estimation by Circuit Analysis

In order to demonstrate that our approach is sound, the circuit analysis approach is used to estimate a lower bound for the tuning sensitivity. The $\mathrm{WGH_{n11}}$ mode is characterized by a chain of approximately elliptical loops of electric field in the $z - \phi$ plane (as shown in Fig. 5) linked by loops of magnetic field in the $r - \phi$ plane. Because of the continuity of the electric field lines and of displacement current, we can estimate the effect of the gap using a simple series–capacitance model. Conventional circuit analysis gives $\omega = \sqrt{1/LC}$, and assuming that any change in effective inductance is small we find:

$$\frac{\Delta\omega}{\omega} \approx -\frac{1}{2}\frac{\Delta C}{C}.$$

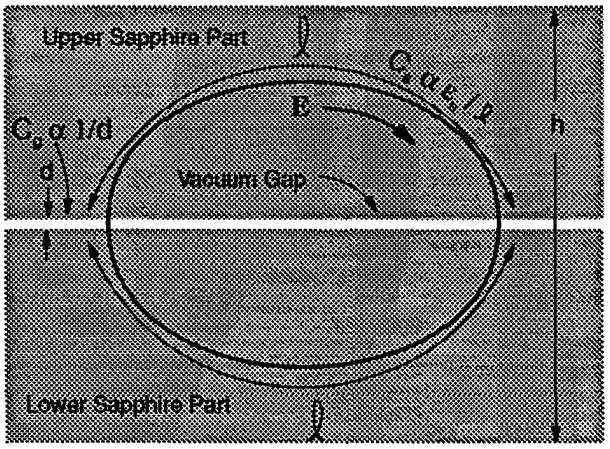


Figure 6: Detail view of one E–field loop showing elements of series–capacitance model.

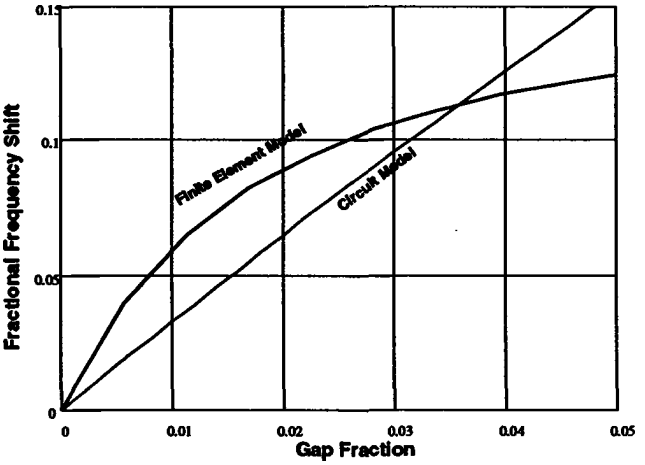


Figure 7: Gap Sensitivity Estimated by Circuit Analysis and Finite Element Calculation.

Each electric field loop traverses a path ℓ through the sapphire with dielectric constant ϵ_s and then a gap distance d with $\epsilon = 1$ as shown in Fig 6. Because of the continuity of displacement current[15] we can approximate the sapphire capacitance as $C_s \propto \epsilon_s/\ell$ and the gap capacitance as $C_g \propto 1/d$. By combining these capacitances in series we estimate the dependence of the capacitance C on the gap spacing d as:

$$\frac{\Delta C}{C} \approx -\frac{d\epsilon_s}{\ell + d\epsilon_s}.$$

We estimate a lower bound for the tuning effect by assuming that the loops are as long as possible, touching both the resonator top and bottom. Approximating the elliptical loops as circular the loop length becomes

$$\ell \approx \frac{\pi}{2} \times h$$

where h is the sapphire height. Combining these two equations in the limit of small g, shows the frequency sensitivity in terms of distance sensitivity to be

$$\frac{\Delta\omega}{\omega} \approx -\frac{1}{2}\frac{\Delta C}{C} \approx \frac{1}{2}\frac{g\epsilon_s}{\ell} \approx \frac{\epsilon_s}{\pi} \times \frac{g}{h}$$

or

$$\left|\frac{\delta\omega}{\omega}\right| \geq \frac{\epsilon_s}{\pi} \times \left|\frac{\delta x}{x}\right| \qquad (2)$$

where $\delta x/x \equiv \delta g/h$ is the fractional variation of the resonator height.

Since $\epsilon_s \approx 10.5$, a comparison of Eqs. 1 and 2 shows that the circuit model predicts a sensitivity more than sufficient to achieve complete compensation at 77 Kelvins. That is, the tuning sensitivity of $\epsilon_s/\pi \approx 3$ is larger than the required value of 13.5/7 from Eq. 1.

4.1.2 Gap Sensitivity Estimation by Finite Element Calculation

Figure 7 shows a comparison of the circuit model prediction with calculations using a recently developed finite–element methodology. This CYRES 2-D program[8] method takes advantage of sapphire's cylindrically symmetric dielectric properties to allow a simplified and more accurate calculation of whispering gallery mode frequencies and fields than was previously possible. The finite–element approach allows relatively complicated geometries such as ours to be easily treated.

As expected, this more accurate calculation gives a larger tuning effect than the (lower bound) circuit model prediction, but, somewhat surprisingly, shows an additional effect. As shown in Fig. 7, the finite element method predicts that tuning effectiveness (the *slope* of the curve) will be degraded for gaps as small as 0.02 of the resonator height.

> This reducing sensitivity is shown to be due to a feature not included in our simple circuit model. For larger gaps (and also for large n, where the wide elliptical loops for E as shown in Fig. 6 tend to be narrow and tall instead) finite element solutions exhibit a substantial horizontal (azimuthal) electric field component near the gap, showing that the gap capacitance C_g is bypassed by azimuthal displacement currents in the

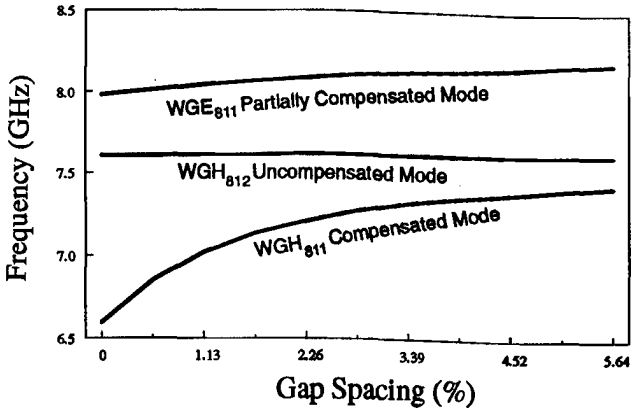


Figure 8: Frequency dependence on gap spacing for modes from various families (finite element calculation).

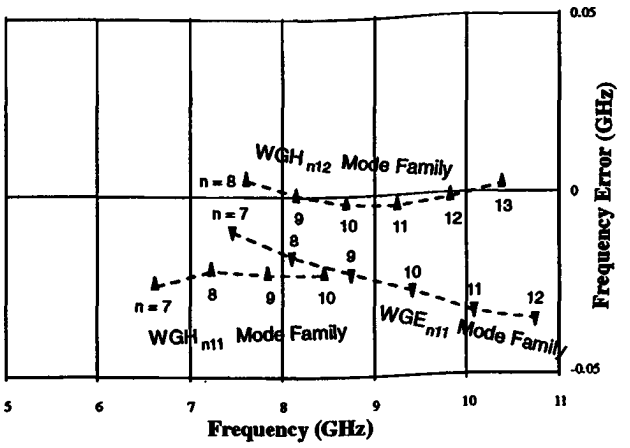


Figure 9: Frequency difference between experimental results and finite element calculations for several mode families.

sapphire. That is, a more accurate circuit model would contain an additional capacitance C_{sg} acting in parallel with C_g.

The variation in tuning sensitivity with gap spacing provides a means to adjust its strength in order to match other requirements. Thus the gap may be varied to provide compensated operation in a particular mode at a particular temperature.

Not all mode families are found to be strongly affected by changes in the gap spacing. This is to be expected, for example, for modes with very small E fields in the gap region. Figure 8 shows the frequency dependence on gap spacing for three examples of modes with different characters for their electric fields. In order of decreasing sensitivity to gap spacing they are:

- The WGH_{811} mode with a maximum of vertical E field in the gap shows a rapid increase in frequency with increasing gap spacing:
- The WGE_{811} mode has a maximum of radial (horizontal) E field at the gap and shows a slight frequency increase:
- The WGH_{812} mode with a sign reversal of vertical E field at the center, and very small values in the gap shows almost no change in frequency.

These results quantitatively confirm that a strong vertical E field in the gap region, as exhibited by the WGH_{n11} mode family, and displayed in Figure 5, is essential to achieve high sensitivity to gap spacing. For this geometry it is also the "fundamental" mode family, showing the lowest microwave frequency and highest mode confinement for any given azimuthal wave number n [1]. This family is thus an ideal candidate for use in a composite compensated resonator. Similar effects have been previously used to provide frequency variability for sapphire resonators [16].

4.2 Experimental Tests

4.2.1 Mode Frequencies

A resonator was constructed with configuration and dimensions as shown in Fig. 4, and with the parts mechanically and thermally bonded by means of pure indium solder. A clean (scraped) molten indium pool on each end of the copper post was mated in turn to an evaporated gold layer on the sapphire parts. After cooling to 77 kelvin, the frequency, Q, and coupling coefficient was measured for each of 69 resonant modes from 6.6 GHz to 10.75 GHz. This list was then preliminarily matched by frequency with the finite element data. Analysis of the electromagnetic visualization of the resonator cross section using the CYRES 2–D software conclusively identified the experimental modes for each family.

Figure 9 shows the excellent agreement between theory and experiment for the three mode families previously discussed. The data indicates a frequency difference of less than 0.4%.

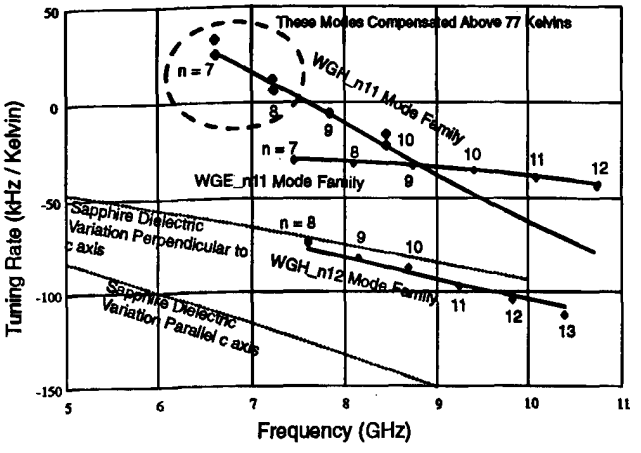


Figure 10: Calculated (lines) and experimental (points) temperature tuning rates at 77 Kelvin for the 3 dominant (lowest frequency) mode families in a sapphire/copper compensated resonator with a gap of .050 cm.

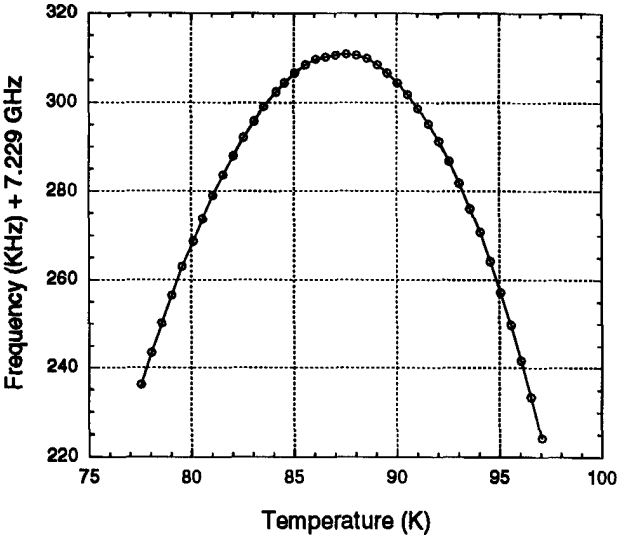


Figure 11: Temperature dependence of frequency for WGH_{811} compensated mode at 7.23 GHz.

4.2.2 Temperature Tuning Rates

A direct demonstration of the effectiveness of our compensation mechanism can be shown by experimental measurement of the rate of frequency change with temperature at 77 Kelvins. As shown in Figure 10, the experimental points with positive values indicate modes that are actually overcompensated at 77 K. They will have turnover temperatures (complete compensation) above 77 K, as desired. Negative values indicates under–compensation or even no compensation (if values are approximately the same as expected from the sapphire dielectric variation alone), and a zero value would indicate a turnover temperature at exactly 77 K.

A comparison of calculated and measured tuning rates in Figure 10 shows excellent agreement. Sensitivity to small changes in the gap spacing was calculated with the finite element software. The results were combined with values for the expansion coefficients of copper and sapphire (Fig. 3) and a *fitted* value for the sapphire dielectric temperature dependence. As shown in Fig. 2, the dielectric variation values can be expected to vary between 9.4 PPM/Kelvin (perpendicular) and 16.75 PPM/Kelvin (parallel) at 77 Kelvins. However this represents data measured at kilohertz frequencies[12], and may be modified at microwave frequencies.

The fitted values were 11PPM/Kelvin and 10.5PPM/Kelvin for the WGH_{n11} and WGH_{n12} mode families, respectively, and 7 PPM/Kelvin for the WGE_{n11} family. It is to be expected that the WGE families would have a lower value, because the electric fields in this case are almost entirely in the $r - \phi$ plane, and so correspond to the lower "perpendicular" results, while the values for the WGH modes have a substantial fraction of their electric fields "parallel" to the z axis.

Both calculated and experimental values in Fig. 10 show a weakening of the compensation effect for higher mode numbers (higher frequencies) that was not predicted by the simple circuit model as discussed in relation to Fig. 7. Thus, modes with $n \geq 9$ in the WGH_{n11} mode family are under–compensated at 77 Kelvins, and so would require operation at a lower temperature. However the $n = 8$ modes at 7.23 GHz are just slightly over–compensated, and so can be operated at a temperature a little above 77 K, as desired. More detailed calculations show that higher frequency compensated operation is possible using a smaller gap spacing – where the compensation effectiveness is not so strongly dependent on the mode number.

4.2.3 Compensated Resonator Operation

One of the two WGH_{811} modes was chosen for further study. This mode showed the highest quality factor of any of the compensated modes with $Q = 1.8 \times 10^6$. Figure 11 shows a plot of the res-

onance frequency for this mode showing a turn–over temperature of 87.09 Kelvins. A quadratic approximation in the vicinity of the peak gives:

$$\frac{\delta f}{f} \approx 1.17 \times 10^{-7}(T - 87.09)^2.$$

A residual linear thermal coefficient due to imperfect temperature adjustment $\delta T = T - 87.09$ can be derived from the slope of the curve as

$$\frac{1}{f}\frac{\partial f}{\partial T} \approx 2.34 \times 10^{-7}\delta T. \quad (3)$$

Equation 3 allows us to estimate the thermal requirements that would allow such a resonator to achieve its ultimate stability of $\delta f/f \approx 10^{-14}$. If the temperature is held at the turnover temperature with an *accuracy* of $\delta T = 1$ millidegree, the slope given by Eq. 3 will be less than 2.34×10^{-10} per Kelvin, requiring a *stability* of 43 micro–Kelvins to achieve 10^{-14} stability. Accuracy and stability are distinguished in this discussion, because the δT accuracy needs to be held over a relatively long operational time period (possibly days or months), while the strength of the sapphire oscillators is in short–term stability. Thus, in order to achieve a stability of 10^{-14} for a time period of e.g. 100 seconds, the temperature would need to be stable to 43 micro–Kelvins for 100 seconds, but could vary up to 1 milliKelvin over the time period of operation. These requirements are easily met using conventional thermal regulation technology as developed for use by other types of frequency standards[17].

4.2.4 Oscillator Stability

An oscillator was constructed, stabilized by the WGH_{811} mode of the compensated resonator. Preliminary oscillator tests were accomplished with open loop control of the resonator temperature, and with the can temperature not regulated, but determined by direct contact with a liquid nitrogen bath. The stability of the oscillator was characterized using a hydrogen maser frequency standard as reference.

As shown in Figure 12, the Allan Deviation of frequency variation was measured to be 1.4 - 6 $\times 10^{-12}$ for measuring times 1 second $\leq \tau \leq$ 100 seconds. There was a large but constant frequency drift during the course of the measurements which we attribute to the uncontrolled can temperature and the changing level of liquid nitrogen. It seems likely that the frequency variation observed was also caused by can temperature variations. We project a frequency stability 10^{-14} for this resonator with stabilized can temperature and with a mode quality factor of $Q = 10^7$.

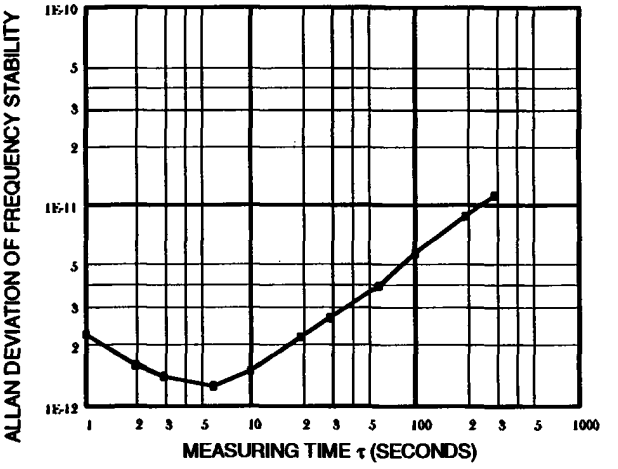


Figure 12: Frequency stability of sapphire oscillator with uncontrolled can temperature

5 Conclusion

We have demonstrated a new ultra–stable oscillator (USO) capability which promises performance improvements over the best available crystal quartz oscillators in a compact cryogenic package. First tests of this mechanically compensated sapphire oscillator show stability in the low 10^{-12} region for measuring times $\tau \leq 100$ seconds. We project a stability of 10^{-14} for this technology at liquid nitrogen temperatures.

The authors would like to acknowledge valuable contributions made by many of the members of the JPL Frequency Standards Laboratory, and in particular by T. Tucker, P. Reder, L. Maleki, W. Diener, and A. Kirk; and by A. Revilla of UTEP.

REFERENCES

[1] 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); also see 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) and G.J. Dick and J. Saunders, "Method and Apparatus for Reducing Microwave Oscillator Output Noise," U.S. Patent #5,036,299, July 30, 1991.

[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); also see 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); available commercially through Sapphire Technologies, technical marketing for the University of Western Australia.

[3] S.L. Abramov, E.N. Ivanov, and D.P. Tsarapkin, "Low noise microwave oscillator with a temperature stabilized disk dielectric resonator," *Radiotechnika,* No. 11, 81–83 (1988) (in Russian); also see D.P. Tsarapkin, "An Uncooled Microwave Oscillator with 1–Million Effective Q–Factor," *Proc. 1993 IEEE International Frequency Control Symposium,* pp. 779–783 (1993).

[4] M.M. Driscoll, J.T. Haynes, S.S. Horwitz, R.A. Jelen, R.W. Weinert, J.R. Gavaler, J. Talvacchio, G.R. Wagner, K.A. Zaki and X–P Liang, "Cooled, Ultra–high Q Sapphire Dielectric Resonators for Low Noise Microwave Signal Generation," *Proc. 45th Symposium on Freq. Control,* 700–706 (1991).

[5] I. Panov and P.R. Stankov, "Frequency Stabilization of Oscillators with high–Q Leucosapphire Dielectric Resonators," *Radiotekhnika i Electronika* **31**, 213, (1986), (In Russian); also "VEGA–M" commercial frequency source with Sapphire Resonator at 6 GHz, NPO, 48 Ivan Franko Str., 121351 Moscow, Russia.

[6] C.A. Flory and R.C. Taber, "Microwave Oscillators Incorporating Cryogenic Sapphire Dielectric Resonators," *Proc. 1993 IEEE International Frequency Control Symposium,* 763–773 (1993).

[7] 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).

[8] 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.

[8] R.A. Osegueda, J.H. Pierluissi, L.M. Gill, A. Revilla, G.J. Villalva, G.J. Dick, D.G. Santiago and R.T. Wang, "Two Dimensional Finite Element Determination of Modes and Frequencies for an Anisotropic Cylindrical Dielectric Resonator" to be published in *Proc. 10th Ann. Rev. of Progress in Appl. Computational Electromagnetics (ACES)* 1994.

[9] J.N. Hollenhorst, R.C. Taber, L.S. Cutler, T.L. Bagwell, and N. Newman, "High–Temperature Superconducting Resonators," *Proc. 45th Symposium on Freq. Control,* 452–459 (1991).

[10] R. Besson and U.R. Peier, "Further Advances on BVA Quartz Resonators," *Proc. 34th Ann. Symposium on Frequency Control,* 175 (1980).

[11] "Dielectric Resonators," D. Kajfez and P. Guillon, Editors, p. 362 ARTECH HOUSE, Norwood, MA, 1986

[12] R. Shelby and J. Fontanella, "The low temperature electrical properties of some anisotropic crystals," *J. Phys. Chem. Solids,* **41** pp. 69-74, 1980.

[13] "Thermal Expansion of Technical Solids at Low Temperatures, A Compilation From the Literature," R.J. Corruccini and J.J. Gniewek, *National Bureau of Standards Monograph 29,* 1961.

[14] LISA Flight Mission Study, European Space Agency, L. Maleki (JPL), Private Communication.

[15] "Principles of Electricity," L. Page and N. I. Adams, Jr., D. Van Nostrand Inc., Princeton, 1958.

[16] M.E. Tobar and D.G. Blair, "Analysis of a Low Noise Tunable Oscillator based on a Tunable Sapphire Loaded Superconducting Cavity," *Proc. 45th Symposium on Freq. Control,* 495–499 (1991); also see "Mode control of Cryogenic

Whispering–Gallery Mode Sapphire Dielectric–Ring Resonators," D.G. Santiago, G.J. Dick, and A. Prata, Jr., *IEEE Trans. MTT*, **42**, 52–55 (January, 1994).

[17] For example, see J. Vanier, "The Active Hydrogen Maser: State of the Art and Forecast," *Metrologia* **18**, pp 173–186 (1982); also see H.E. Peters, H.B. Owings, and P.A. Koppang, "Hydrogen Masers with Cavity Frequency Switching Servos," *Proc. 22nd Ann. Precise Time and Time Interval (PTTI) Applications and Planning Meeting (NASA Conference Publication 3116)*, pp 283-292 (1991).

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

LOW NOISE MICROWAVE OSCILLATORS BASED ON HIGH-Q TEMPERATURE STABILIZED SAPPHIRE RESONATORS

*M.E. Tobar, *E.N. Ivanov, *R.A. Woode, †J.H. Searls

*Department of Physics, University of Western Australia, Nedlands, WA, 6009, Australia.
†Poseidon Scientific Instruments Pty. Ltd., 14 Mouat St. Fremantle, WA, 6160, Australia.

Abstract: This paper reports on development of X-band microwave oscillators incorporating recently developed sapphire resonators. Two resonators have been constructed for oscillator development. One resonator is a room temperature thermoelectric (TE) stabilized sapphire resonator operating at 9.00000 GHz and configured as a two port device with an unloaded Q factor of 187,000 and port couplings of 0.74 and 0.17. The second resonator is a liquid nitrogen cooled sapphire resonator with two stage temperature control operating at 8.95 GHz with an unloaded Q factor of about 40 million. The single sideband phase noise of a free running loop oscillator incorporating the room temperature resonator has been measured to be -104 dBc/Hz at 1 kHz with a f^{-3} dependence and is limited by the resonator Q and the flicker noise of the active components. By incorporating specially designed feedback electronics we have measured a reduced single sideband phase noise of -125 dBc/Hz at 1 kHz with a f^{-3} dependence. For this oscillator configuration the phase noise was limited by the flicker noise of the phase detector in the feed back electronics. Based on the room temperature results we predict that we will obtain better than -155 dBc/Hz at 1 kHz with the liquid nitrogen cooled resonator-oscillator.

Introduction

Sapphire resonator technology now provides oscillators at microwave frequencies with phase noise performance superior to existing technologies. Low noise oscillators incorporating sapphire technology have been constructed and are under further development by various research groups. Such devices have been made at room temperature down to cryogenic temperatures [1-10].

The dielectric loss tangent of sapphire is greatly enhanced at cryogenic temperatures with resonator Q-factors of $Q\sim2\times10^5$ at 300K, $Q\sim5\times10^7$ at 77K and $Q\sim8\times10^9$ at 4K previously achieved. With the gain in Q-factor the complexity and size of the resonator also increases. Ideally to gain maximum performance a liquid helium or sub liquid helium cooling system could be used to achieve maximum Q. However, the expense of construction and maintenance of such systems means that they are probably only ever going to be used for special purposes where the very best performance is essential. One such purpose is for fly wheel oscillators for atomic standards such as the trapped ion frequency standard at CSIRO National Measurement Laboratory, Sydney, Australia [11]. Liquid helium systems have been in the research stage over the last ten years at the University of Western Australia (UWA) [1,7] and the Jet Propulsion Laboratories (JPL) [3], with outstanding short term frequency stabilities. Recently stabilities between 10^{-15} to 10^{-14} have been achieved over the range of .1 to 10,000 seconds [12].

For general commercial applications at room temperature sapphire technology still holds an advantage over other technologies such as quartz and DROs. In this paper we report the measured phase noise of a 9.000 GHz room temperature oscillator with superior performance between 100 Hz to 10 kHz. This oscillator is based on the recently developed thermoelectric stabilized resonator which is now commercially available [10]. These resonators are compact and robust and unlike the cryogenic version they can easily be used for mobile applications.

With the advent of high Tc superconducting technology there has been a flurry of research to make compact superconducting devices cooled in liquid nitrogen cryostats and by closed cycle systems. This technology is also important for sapphire resonators as the Q-factor is improved by two orders of magnitude from 300K to 77 K. Because of the reduced complexity of a nitrogen system relative to a helium system, and the large potential improvement in performance relative to a room temperature system, research on nitrogen cooled sapphire resonators and oscillators have also been under investigation [4,5,13]. At UWA we are now developing a liquid nitrogen oscillator with two-stage temperature control. In this paper we report on our design strategy and our expected results based on our presented room temperature oscillator results. This oscillator will soon be operational.

Mode Temperature Dependencies

The types of modes most commonly used in cylindrical sapphire dielectric resonators are of the whispering gallery

0-7803-1945-1/94 $3.00 © 1994 IEEE

of sapphire (~10) requires modes with high azimuthal mode numbers for high confinement. High confinement is necessary to exploit the low loss tangent of sapphire and limit other loss mechanisms such as radiative and support structure losses. Sapphire is uniaxially anisotropic, and to maximise confinement cylindrical dielectric resonators must be made with the crystal c-axis parallel to the cylindrical z-axis. Whispering gallery modes in dielectric resonators may be classified in two main groups; quasi TE (Transverse Electric) or H modes, and quasi TM (Transverse Magnetic) or E modes. Other higher order modes exist in the sapphire which have significant components of both TE and TM, these modes are called hybrid modes and are usually denoted as HE or EH.

The sensitivity of mode frequencies (f_o) to perturbations in permittivity and dimension are related by [14];

$$\frac{\partial f_o}{\partial \varepsilon_r}\frac{\varepsilon_r}{f_o} + \frac{\partial f_o}{\partial \varepsilon_z}\frac{\varepsilon_z}{f_o} = -\frac{1}{2} \qquad (1)$$

$$\frac{\partial f_o}{\partial d}\frac{d}{f_o} + \frac{\partial f_o}{\partial h}\frac{h}{f_o} = -1 \qquad (2)$$

where ε_r and ε_z are the permittivities perpendicular and parallel to the c-axis respectively and h and d are the height and diameter of the cylindrical dielectric respectively. Equation (1) is true if;

$$f_o \propto (u\,\varepsilon_r^2 + v\,\varepsilon_r^2)^{-1/4} \qquad (3)$$

as expected for a uniaxially anisotropic crystal. Here u and v are constants between 0 and 1 which depend on the mode structure.

The fractional change in permittivities with temperature are a about a factor of 20 greater than the fractional dimensional change [15,16] when cooling from room temperature to 4 K. Thus by comparing (1) and (2) the effect of temperature on resonant frequency is about 10 times greater due to the permittivity change over the dimensional change. Therefore the resonant frequency-temperature dependence over this temperature range is mainly determined by the dielectric constant of sapphire. For whispering gallery E-modes the TE component is small as v->1 and u->0, which means $(\partial f_o/\partial \varepsilon_z)(\varepsilon_z/f_o)$ -> -1/2 and $(\partial f_o/\partial \varepsilon_r)(\varepsilon_r/f_o)$ -> 0. Similarly for whispering gallery H-modes the TM component is small as v->0 and u->1, which means $(\partial f_o/\partial \varepsilon_r)(\varepsilon_r/f_o)$ -> -1/2 and $(\partial f_o/\partial \varepsilon_z)(\varepsilon_z/f_o)$ -> 0. Thus the temperature dependence of the resonant frequency for whispering gallery E and H modes is given by;

$$\frac{\partial f_o}{\partial T}\frac{T}{f_o} = -\frac{\partial \varepsilon_z}{\partial T}\frac{T}{2\varepsilon_z} \quad \text{(E-modes)} \qquad (4)$$

$$\frac{\partial f_o}{\partial T}\frac{T}{f_o} = -\frac{\partial \varepsilon_r}{\partial T}\frac{T}{2\varepsilon_r} \quad \text{(H-modes)} \qquad (5)$$

The temperature dependence of the permittivity of sapphire has been measured previously [16,17]. Comparing these references suggests that the absolute values of permittivity are accurate to about 0.3 %, however relative values are accurate to better than 0.05 %. The change in resonant frequency with respect to temperature only depends on the relative change. From the experimental data in [16] we have calculated the expected dimensionless frequency-temperature coefficient for E and H modes, and plotted the magnitude in figure 1.

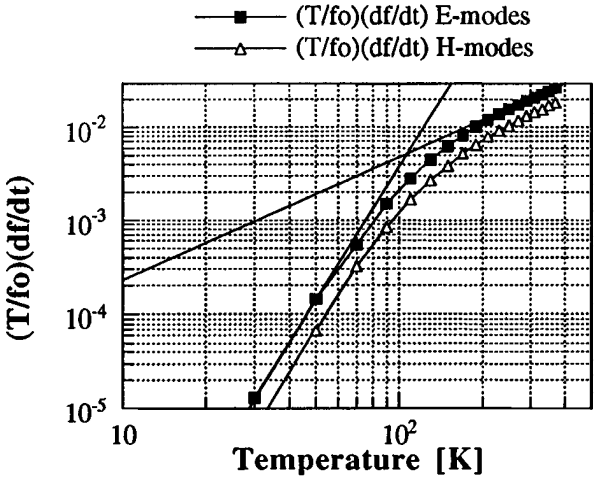


figure 1. Dimensionless frequency-temperature coefficient for E and H modes

The frequency-temperature coefficient is proportional to $T^{1.3}$ above 110 K and proportional to $T^{4.7}$ below 110 K. The magnitude of this coefficient for H-modes is approximately half that of the E-modes. Hybrid modes that have both H and E components are expected to have a temperature coefficient between the two values. At 77K and 300 K respectively E-modes have (df/dT)/fo = -11.3 ppm/K and -67 ppm/K, while H-mode have (df/dT)/fo = -6.5 ppm/K and -46 ppm/K. The temperature coefficient for the room temperature TE resonator was measured to be -70 ppm/K, which is consistent with -67 ppm calculated for E-modes.

At liquid helium temperatures (~ 4K) the temperature-frequency coefficient can be reduced to zero due to paramagnetic impurities having the opposite effect [18]. This is desirable as temperature fluctuations is a major source of frequency instability in a sapphire resonator.

Operating on these turning points has produced oscillators with very small frequency instabilities of the order 10^{-15} [1,7,12].

Thermal Time Constants

To obtain a good frequency stability in a sapphire resonator temperature control is necessary. The rate of control is determined by the thermal resistance and the heat capacity of the sapphire dielectric. Here we plot data taken from Touloukian et al [19] in figure 2. We show the thermal conductivity (k), heat capacity (Cp) and ratio of the two (Cp/k). The thermal time constant is proportional to Cp/k with the proportionality constant depending on the mass of sapphire and the path of heat flow from the sapphire to the surrounding environment.

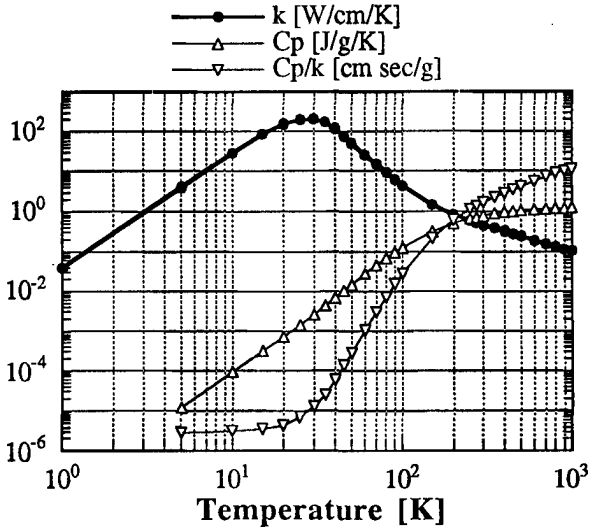


Figure 2. heat capacity, thermal conductivity and ratio as a function of temperature for mono-crystalline sapphire.

From figure 2 it is evident that the thermal time constant governed by sapphire is reduced by more than two orders of magnitude when cooling from room temperature to 77K, and a further three orders of magnitude from 77 K to 4 K.

Typically at room temperature the thermal time constant is dominated by the sapphire, and can be of the order of 10 minutes. Depending on the power dissipation in the cavity, such a long time constant can cause problems in temperature control locking. At cryogenic temperatures below 77 K the time constant is sufficiently small so that the sapphire no longer continues to dominate.

Thermoelectric Stabilized Resonator

The resonator that we are reporting on was designed to operate at 9.000 GHz with a loaded Q factor of 98,000 to match the Q factor of an Interferometric Sapphire Transducer [20]. The resonator is a two-port device with port couplings of 0.74 and 0.17. Previously we have reported on two other Thermoelectric (TE) stabilized resonators with centre frequencies of 10.000,000 GHz and unloaded Q factors of 200,000 [9,21]. These resonators are configured as shown in figure 3. They consists of a cylindrical piece of high-purity single crystal sapphire which is mounted in a metallic cavity. The resonator centre frequency and Q factor are determined almost completely by the sapphire [22].

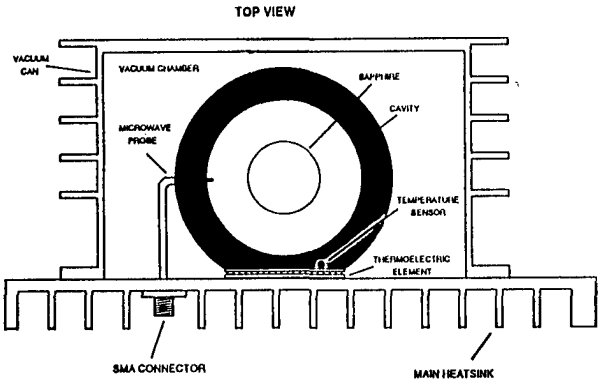


figure 3. Schematic of the TE stabilized high-Q microwave resonator.

This configuration forms a complex set of boundary conditions resulting in a system which has a very large number of resonances of varying frequency and Q factor. The desired mode of operation is quasi-Whispering Gallery E-mode, of moderate azimuthal number. Its frequency was determined by solving Maxwell's equations and matching the field solutions at the dielectric and cavity boundaries [22]. Unwanted resonances are suppressed by modifying the cavity walls to cause such modes to be radiative [9].

The resonant cavity and sapphire dielectric are cooled and temperature controlled via a servo which includes a Lock-In Amplifier, PID controller and a TE module. The thermal time constant of the cavity has been specially designed to enable time constants of only 40 seconds at room temperature.

Temperature Controlled Resonator at 77K

This resonator consists of a 5cm diameter sapphire spindle that has been characterised previously [14]. The mode chosen for operation is the 8.95 GHz $E_{12\,1\,\delta}$ mode. We have chosen this mode because of its proximity in frequency to our room temperature resonator and its high Q-factor at 77K of 40 million. A density plot of the axial electric field intensity in the azimuthal plane is shown in figure 4 below.

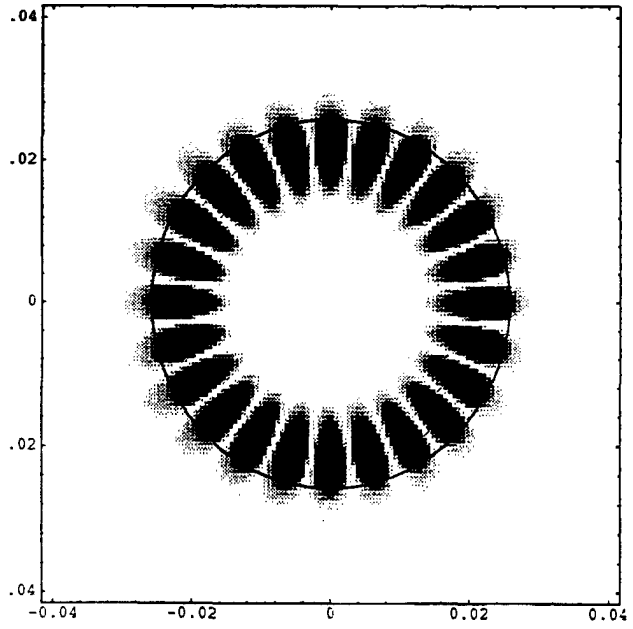


figure 4. Density plot of the axial electric field intensity of the $E_{12\,1\,\delta}$ mode in the azimuthal plane of the resonator, dimensions are in m.

From one fixed point of the permittivity and equation (4) we can calculate the approximate temperature dependence of the resonant frequency for this mode as shown below in figure 5. Alternatively a more accurate calculation can be done using the methods in [14,22]. However fig. 5 shows our approximation to be quite accurate as the experimental points lie close to the predicted frequency curve.

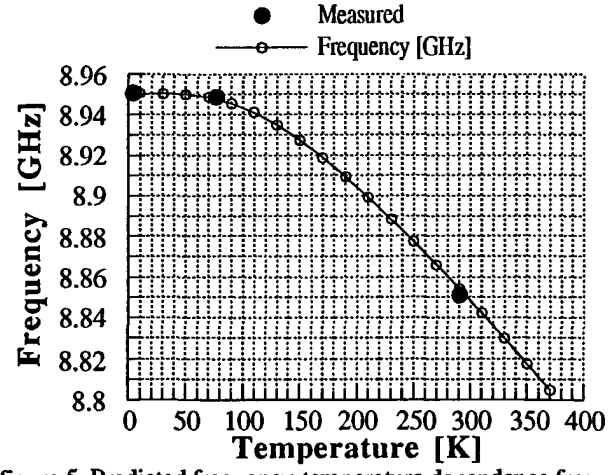


figure 5. Predicted frequency-temperature dependence from 4K to 300 K for the $E_{12\,1\,\delta}$ mode. Measured values compare favourably.

The measured frequency shift from 4K to 300 K is actually slightly larger than predicted. This is due to the component of shift due to dimensional change. From equation (4) we can also calculate the temperature coefficient df/dT as a function of temperature, shown below in figure 6.

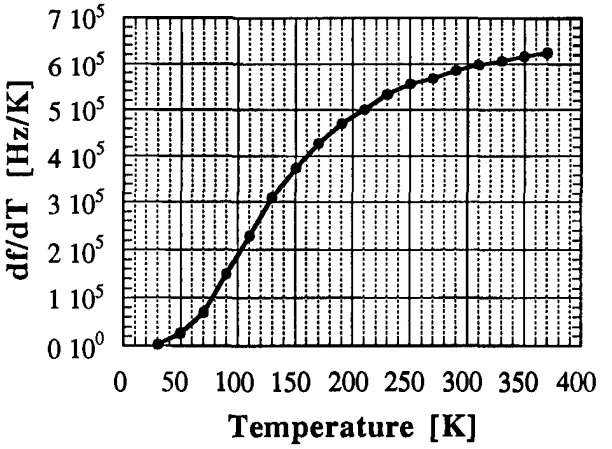


Figure 6. Frequency temperature coefficient for the $E_{12\,1\,\delta}$ mode as predicted from equation (4).

The temperature coefficient at the operation temperature of 77 K is calculated to be 10^5 Hz/K. With a two-stage temperature control system we conservatively estimate that we can obtain a temperature stability of better than 1 mK. This means we anticipate a fractional frequency stability of better than 10^{-8}. To overcome this bad temperature coefficient a mechanical temperature compensation technique could be used as described in [23]. This type of scheme will however enhance sensitivity to mechanical vibrations in the same way as shown previously in a mechanically tunable sapphire resonator [6].

To estimate the thermal time constant (τ) of the sapphire dielectric we use the relation $\tau = R\ m\ Cp$, where m is the mass of the sapphire, Cp is the heat capacity, and R is the thermal resistance given by $R = l/(A\ k)$ where l and A are the length and cross sectional area respectively of a sapphire segment. Heat will be generated by the microwave power inside the sapphire dielectric when it is configured as an oscillator. The axial electric field density in the z-plane of the sapphire crystal is shown in figure 7, and the direction of heat flow will occur as indicated. If we assume that all the heat is generated in the centre azimuthal plane then the thermal resistance from the centre plane of the sapphire to the top spindle can be approximated by considering the two segments consisting of the top half of the sapphire cylinder and the top spindle. Assuming that the temperature sensor is on the

top spindle and by ignoring the effect of the bottom half, the thermal time constant for our resonator has been calculated. Its dependence as a function of temperature is shown in figure 8.

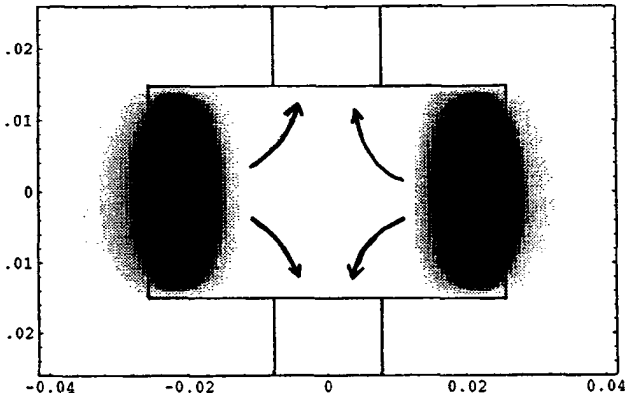


figure 7. Density plot of the axial electric field intensity of the $E_{12\,1\,\delta}$ mode in the axial plane of the resonator, dimensions are in m. Arrows show the direction of heat flow.

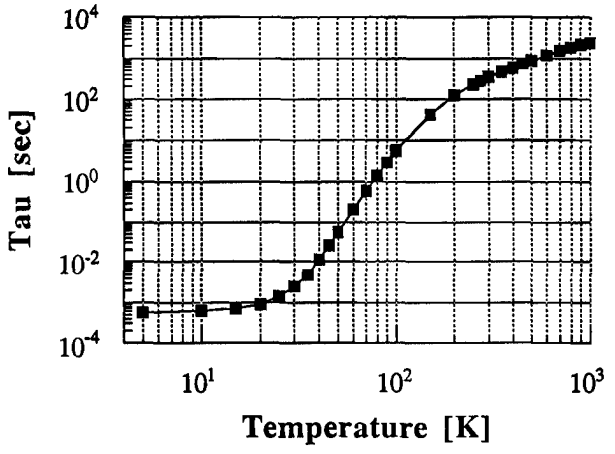


figure 8. Approximate temperature dependence of the thermal time constant due to the sapphire dielectric.

At 300 K the time constant is of the order of a few hundred seconds which can cause tight temperature control to be difficult. At temperatures below 77 K the values are quite reasonable. Successful stable temperature control to 10 μK with a one stage system has already been implemented with these type of resonators at around 4 K [12].

Oscillator Phase Noise Results

A low phase noise oscillator was constructed using the room temperature TE stabilized resonator at 9.000 GHz. A noise reduction technique similar to the methodology described by Galani et al [24] and Walls et al [25] was implemented. A schematic of the stabilized oscillator is shown below in figure 9.

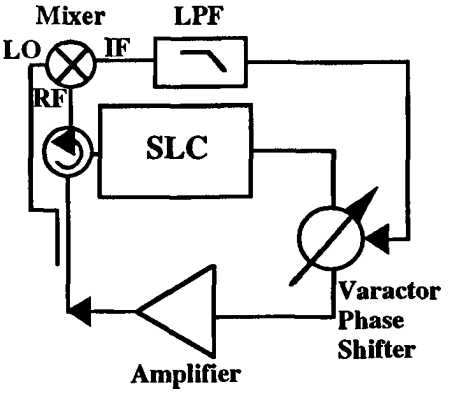


figure 9. Schematic of the TE stabilized resonator-oscillator with phase noise cancellation circuitry.

The loop gain characteristics of our noise cancellation was determined mainly by an active filter in the feed back path to the varactor phase shifter. This filter was designed to have maximum gain at 1 kHz, and was measured by sweeping a modulation signal to the varactor and measuring the difference in response to the open and closed loop configurations. The measured response is shown below in figure 10.

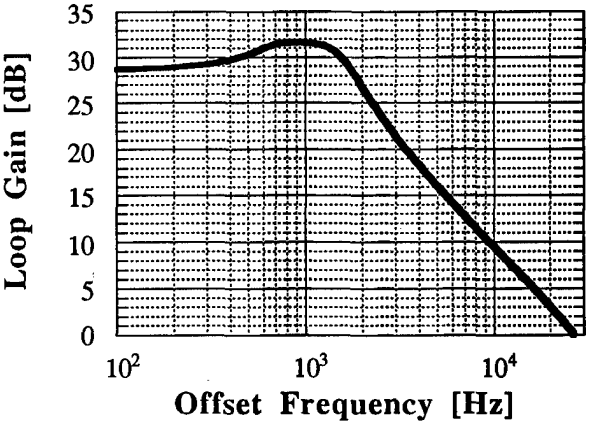


figure 10. Measured loop gain of the phase noise cancellation feed back loop.

The phase noise of this oscillator was measured using the following setup shown in figure 11. The result of the phase noise measurements reveal an excellent phase noise performance (shown in figure 12). To obtain this measurement we did not require any vibration isolation, which substantiates the fundamentally good performance possible from whispering gallery resonator-oscillators.

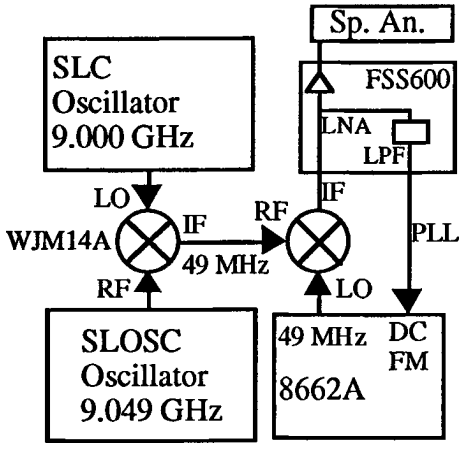


figure 11. Schematic of the oscillator phase noise measurement system.

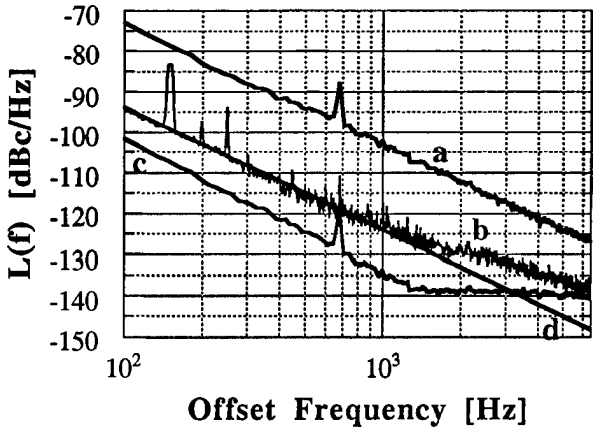


figure 12. Measured single sideband phase noise for the TE stabilized resonator-oscillator: **a.** open loop configuration; **b.** closed loop configuration; below 2 kHz we measured the noise limited due to the phase detector; above 2 kHz we measured noise due to the 8662A synthesizer; **c.** open loop noise divided by the loop gain; **d.** Oscillator noise limit due to the phase detector .

The noise in the 8662A was determined by beating two separate synthesizers together at 49 MHz and alternating them in the measurement system. Both are fairly old and regularly exhibited phase noise levels above specifications that intermittently changed. Figure 12 shows results obtained with the best noise floor from one of the 8662A synthesizers.

Above 6 kHz the measured phase noise spectrum was flat out to 100 kHz offset at a value close to -140 dBc/Hz. This noise level is due to the 8662A synthesizer. In the future we will use both synthesizers in a two channel measurement system to lower the noise floor of the measurement [26, 27].

Below 2 kHz offset frequency we have measured the flicker noise limitation due to the WJ M14A mixer in the phase detector. Thus to obtain a lower oscillator noise a more sensitive or lower noise mixer would have a direct impact.

If we had measured the phase noise beyond 10 kHz offset then the phase noise limit in our oscillator would be due to the microwave amplifier noise, because our current cancellation scheme does not have enough gain at 10 kHz offset. Theoretically simulating the oscillator reveals that if we increased the bandwidth of the loop gain response we could lower the noise at 10 kHz to better than -145 dBc/Hz. However if we increased the bandwidth much further a degradation of the noise performance beyond 100 kHz would become significant.

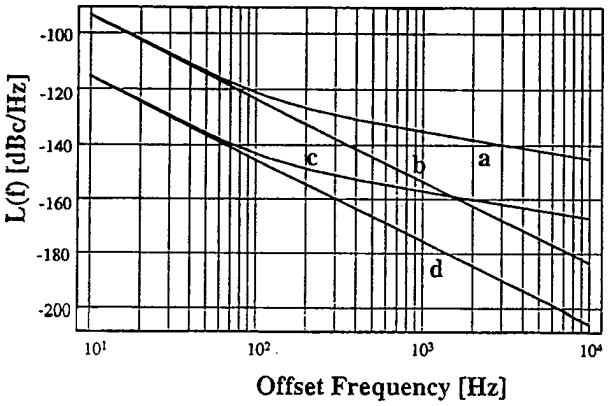


figure 13. Predicted single sideband phase noise performance of the liquid nitrogen cooled resonator-oscillator: **a.** open loop noise from non-filtered port; **b.** open loop noise from filtered port; **c.** closed loop noise from non-filtered port; **d.** closed loop noise from filtered port.

Assuming the same mixer noise level in the liquid nitrogen cooled oscillator under development, we predict the following noise performance shown in figure 13.

Conclusion

In a recently developed room temperature TE stabilized sapphire resonator-oscillator, we have measured a phase noise of -125 dBc/Hz at 1 kHz offset with a f^{-3} dependence. This noise was limited by the mixer noise floor in the phase detection circuitry. Using a similar type of phase noise cancellation in a liquid nitrogen cooled oscillator we expect to obtain a phase noise performance of -145 dBc/Hz at 100 Hz offset with a f^{-3} dependence. Any improvement in current mixer technology would

have a direct impact on the noise performance of this type of oscillator.

Acknowledgments

The authors thank Simon Edwards and Cameron McNeilage for making the servo electronics and the room temperature cavity. Also they thank Andre Luiten for his helpful advice during noise measurements, and the rest of the sapphire clock group at UWA for their general assistance. This project was supported by the Australian Research Council.

References

[1] 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," in Proc. IEEE 43rd Annu. Symp. Freq. Contr., 1989, pp. 89-93.

[2] 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.

[3] G.J. Dick and R.T. Wang, "Ultra-stable performance of the superconducting maser," IEEE Trans. Instrum. Meas., vol. 40, no. 12, pp. 174-177, Dec. 1991.

[4] 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.

[5] 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.

[6] 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.

[7] A.N. Luiten, A.G. Mann and D.G. Blair, "Ultra high Q-factor cryogenic sapphire resonator," Electron. Lett., vol. 29, no. 10, pp. 879-881, 1993.

[8] 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.

[9] 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.

[10] 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.

[11] P.T.H. Fisk et al, "Performance of a prototype microwave frequency standard based on trapped $^{171}Yb^{+}$ ions," in these proceedings.

[12] A.N. Luiten, A.G. Mann, M.E. Costa and D.G. Blair, "Cryogenic sapphire resonator-oscillator: an update," in these proceedings.

[13] Z.Y. Shen, C. Wilker, P. Pang and W. Holstein, "High Tc superconductor-sapphire microwave resonator with extremely high Q-values up to 90K," in 1992 IEEE MTT-S Int. Microwave Symp. Dig., F-3, pp. 193-196.

[14] M.E. Tobar and A.G. Mann, "Resonant frequencies of higher order modes in cylindrical anisotropic dielectric resonators," IEEE Trans. MTT., vol. 39, no. 12, pp. 2077- 2081 Feb. 1991.

[15] V.B. Braginsky, V.P. Mitrofanov and V.I. Panov, *Systems with small dissapation*, The University of Chicago Press Press, 1985.

[16] R. Shelby and J. Fontanella, "The low temperature electrical properties of some anisotropic crystals," J. Phys. Chem. Solids, vol. 41, pp. 69-74, 1980

[17] E. Lowenstein, D. Smith and R. Morgan, "Optical constants of far infrared materials," Appl. Opt., vol. 12, pp. 398, 1973.

[18] A.G. Mann, A.J. Giles, D.G. Blair and M.J. Buckingham, "Ultra-stable cryogenic sapphire dielectric microwave resonators: mode frequency-temperature compensation by residual paramagnetic impurities," J.Phys D: Appl. Phys., vol. 25, pp. 1105-1109, 1992.

[19] Y.S. Touloukian et al, *Thermophysical properties of matter vol 2 and 4*, IFI/Plenum, New-York Washington, 1970.

[20] H. Peng, D.G. Blair and E.N. Ivanov, "An ultra high sensitivity sapphire transducer for vibration measurements," to be published in J.Phys D: Applied Physics, 1994.

[21] J.H. Searls, S.J. Edwards, E.N. Ivanov, M.E. Tobar and R.A. Woode, "High-Q T.E. stabilised sapphire microwave resonators and low noise oscillators," in Proc. of 8th European Frequency and Time Forum, 1994.

[22] E.N. Ivanov, D.G. Blair and V.I. Kalinichev, "Approximate approach to the design of shielded dielectric disk resonators with whispering gallery modes," IEEE Trans. Microwave Theory Tech., vol. 41, Apr. 1993.

[23] V.I. Panov and P.R. Stankov, "Frequency stabilization of oscillators by high-Q dielectric resonators made from leucosapphire," Radiotekhnika i electronica, no. 1, pp. 213-216, 1986.

[24] Z. Galani et al, IEEE Trans. Microwave Theory Tech., vol. 32, pp. 1556-1565, 1984.

[25] F.L. Walls, C.M. Felton, T.D. Martin, "High spectral purity X-band source," Proc. 44th Symp. Freq. Contr., pp. 542-547, 1990.

[26] F.L. Walls and S.R Stein, "Acurate measurements of spectral density of phase noise in devices," in Proc. 30th Annu. Freq. Contr. Symp., pp. 335-343, 1977.

[27] D. Fest, J. Groslambert and J-J. Gagnepain, "Individual characterization of an oscillator by means of cross-correlation or cross variance method," IEEE Trans. Instrum. Meas., vol. IM-27, no. 4, pp. 447-450, 1983.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

CRYOGENIC SAPPHIRE RESONATOR-OSCILLATOR WITH EXCEPTIONAL STABILITY: AN UPDATE

A.N. Luiten, A.G. Mann, M.E. Costa and D.G. Blair,

Department of Physics, University of Western Australia, Nedlands, W. A. 6009, Australia.

Abstract

Microwave oscillators of exceptional short term stability have been realized from cryogenic sapphire resonators with loaded Q factors in excess of 10^9 at 11.9 GHz and 6 K. This has been achieved using a synthesis of a power stabilized loop oscillator and active Pound frequency stabilization. These oscillators have exhibited a fractional frequency stability of 3-4 x 10^{-15} for integration times from 0.3 to 100 seconds. The relative drift of these two oscillators over one day is a few times 10^{-13}. To reduce the long term drift, which is principally due to excessive room temperature sensitivity, we have added cryogenic sensors for the power and frequency stabilization servos on one of these oscillators. We have also implemented a servo to reduce the room temperature sensitivity of our phase modulators. Testing of this oscillator against a Shanghai Observatory H-maser has shown an Allan deviation of less than 4 x 10^{-15} from 600 to 2000 s.

Introduction

With the impending development of a new generation of trapped ion atomic frequency standards a requirement exists for new "flywheel" oscillators at least an order of magnitude more stable than state of the art quartz oscillators. Cryogenic oscillators with resonator quality factors in excess of 10^8 have shown excellent short term stability [1, 2, 3]. A microwave oscillator based on a 3 cm diameter cryogenic sapphire dielectric resonator (SDR) with loaded Q-factor ~10^8 achieved a fractional frequency stability of 10^{-14} for integration times of 1-300 seconds [2]. The stability was chiefly limited by the frequency pulling effects of the external circuit. These effects as well as the frequency stabilization servo noise floor scale inversely as the Q-factor. There is a significant advantage, therefore, in obtaining higher Q resonators for high stability oscillators. Higher Q-factors are obtainable in the more highly confined modes of 5 cm diameter resonators. We have measured ~4 x 10^9 at 4.2 K and 5-8 x 10^9 at 2 K [4] in resonators cut from Crystal Systems "HEMEX" material. In 1993 we constructed two loop oscillators based upon these resonators [3, 5]. In this paper we present measurements of the intrinsic resonator properties which are important in determining the stability of an oscillator. We then focus on the frequency stability results from a comparison of the two sapphire oscillators and the noise floors of their servo systems. Finally we report improvements to the long term stability of one of these sapphire oscillators which has been achieved by installing cryogenic sensors on its servo systems, and implementing a further servo to reduce temperature sensitivity of the microwave phase shifters. Verification of this improvement is provided by comparison with a H-maser.

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 resonator 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.

Resonator Frequency-Temperature Dependence

Extensive details of the two sapphire resonators have been presented elsewhere [6, 7]. Resonator 1 was

0-7803-1945-1/94 $3.00 © 1994 IEEE

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 to permit $\sim 10^{-15}$ fractional frequency stability. Unfortunately, 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 combination of the temperature dependence of the sapphire dielectric constant and the ac susceptibility of paramagnetic ions which have an electron spin resonance above 100 GHz [7]. 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.

Resonator Frequency-Power Dependence

For some modes with high coupling coefficients, principally $H_{m\ 1\ 1}$, we have been able to measure the dependence of mode frequency on the microwave power dissipated in the resonator. We determine the dissipated microwave power by monitoring changes in the output power of the temperature control system, on variation of the incident microwave power. We have considerably refined our technique since our first measurements [7], which relied on direct measurement of the change in output power when varying the incident signal power. These measurements have subsequently been shown to be corrupted, primarily by drift in the thermal resistance between the resonator and the helium bath, due to changes in the residual helium gas pressure. Our new measurements rely upon measuring dissipated power indirectly. We measure the incident power with a detector which is then calibrated in terms of the dissipated power. This calibration is made by making large changes in incident power. The microwave dissipated power can now be swept continuously under computer control using the power stabilisation servo.

The power dependence of the $H_{14\ 1\ 1}$ mode frequency for each resonator is shown in Figure 1. In resonator 1 and at high power (> 0.5 mW) in resonator 2 there is an approximate linear dependence of mode frequency on the microwave power dissipated in the resonator. The magnitude of $1/f\ df/dP$ is about $-10^{-10}\ mW^{-1}$ for both resonators which is consistent with the permittivity change due to electromagnetic radiation pressure for a mode with an unloaded Q of 2×10^9 [8]. This effect would limit the long term frequency stability to the 10^{-14} level if the power level were unregulated since power fluctuations of up to 0.1 % are possible in our microwave amplifiers. To limit the influence of power instability we have implemented a power level servo. The curvature visible in Figure 1 at the lowest power (< 0.03 mW) is due to offsets in the frequency stabilization servo that become apparent because of the reduced gain of this servo at low power.

At low power in resonator 1 the extremum reported earlier in the $H_{13\ 1\ 1}$ (11.25 GHz) mode and the high initial (negative) slope in the $H_{14\ 1\ 1}$ (11.9 GHz) mode [7] do not exist. They were artefacts of the residual gas pressure drift. No frequency-power extremum exists for any $H_{m\ 1\ 1}$ modes, where m runs from 11 to 15. In resonator 2 the $H_{13\ 1\ 1}$ and $H_{14\ 1\ 1}$ modes have a high initial (negative) slope, whose origin is unknown.

Small multiple hysteretic effects are present in both resonators which probably don't affect the frequency stability and were not visible before the advent of extremely high resolution swept power measurements. In resonator 1 below 0.25 mW, one hysteresis loop and several small frequency steps are due to superconducting contamination on the surface of the sapphire, which is also seen in high resolution measurements of the temperature dependence of the electrical Q-factor [9]. In resonator 1 at 0.4 mW and resonator 2 at 1.4 mW a narrow hysteresis loop (~0.05 Hz high) is observed in the $H_{14\ 1\ 1}$ mode. This phenomenon, which is not related to superconductivity, may be due to saturation of an electron spin resonance and warrants further investigation.

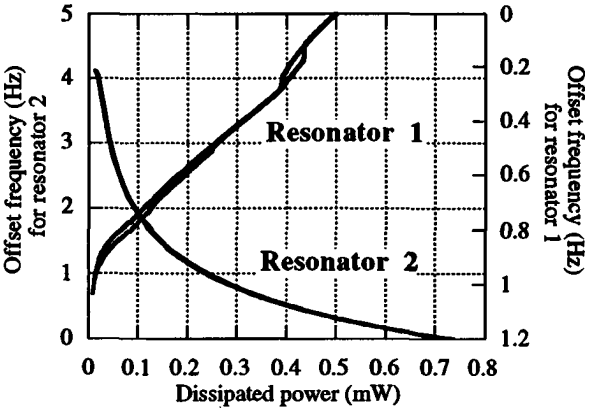


Figure 1 The dependence of mode frequency on power dissipated in resonators 1 and 2, for the $H_{14\ 1\ 1}$ (11.9 GHz) mode.

Loop Oscillator

The improved loop oscillator circuit [3] is essentially our original configuration [2] with the addition of a power level servo and is shown in Figure 2. 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 the active Pound-type frequency stabilisation scheme [10]. A bandpass filter (Q~10^3) is necessary to select the required SDR mode. The circuit losses are compensated by the gain of the GaAs FET amplifier 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 centre 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 or 1.4 MHz, 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) and thereby maximizes discriminator sensitivity. 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. We have operated one of the loop oscillators using a custom made low noise 1.4 MHz demodulation system and a commercial lock-in amplifier performing the same operation at 80 kHz. No difference in frequency stability was detected.

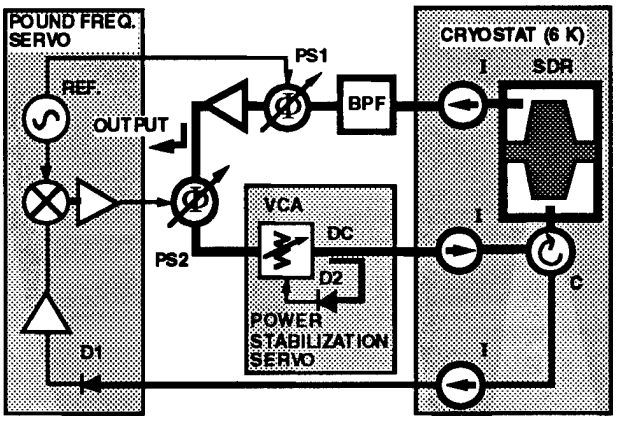


Figure 2 The loop oscillator circuit.

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 partially regulated by the power servo comprising the tunnel diode detector (D2) and voltage controlled attenuator (VCA). This servo fully compensates for amplifier power fluctuations and the variation of insertion loss of PS2 with the error voltage from the frequency stabilization servo. In oscillator 2 (based on resonator 2) D2 is at room temperature to allow for the power to be varied over a wide range in order to explore the mode frequency-power dependence described above. In oscillator 1 (based on resonator 1) D2 is instead placed near the resonator in the cryogenic environment. This eliminates both the long term variation of incident power on the resonator due to variation of transmission line attenuation with liquid helium bath level as well as the room temperature dependence of the detector sensitivity.

In oscillator 1 the frequency stabilization detector (D1) has also been placed near the resonator. This improves the noise floor of the frequency stabilization servo because of the elimination of the 5 dB transmission line loss before D1. The power servo detector (D2) is operated at a power level corresponding to its highest sensitivity. By varying 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.

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 synthesizer which has less absolute instability. For time domain measurements longer than a few seconds the synthesizer is locked to a HP5065A rubidium frequency standard and is set to generate a second difference frequency near 100 Hz which is then counted in a 300 Hz bandwidth on a HP5316A frequency counter. The noise floor of this method (C) is shown in Figure 3. At short times this floor is inadequate to characterize our oscillators.

To circumvent this counter noise for measurement times under a few seconds we use a "zero-beat" technique [5]. This technique provides the fractional frequency stability and phase noise spectral density from the same measurements. The method involves digitising and recording the beat waveform of the two oscillators. The waveform is generated by the usual phase detector technique in which a high-sensitivity mixer is used to compare two oscillator signals. If the signals have the

same frequency and are in phase quadrature the output is proportional to the phase difference between the signals. A low phase noise frequency synthesizer (HP8662A, phase-locked to a low phase noise CO246 Vectron crystal oscillator) is used to offset the difference frequency and trim the phase so that the mixer is least sensitive to amplitude modulation. A Sun SPARCstation 2 calculates the Allan deviation from the time evolution of phase, and by taking the Fourier transform of the time series it can calculate the phase noise. The noise floor of this method (Z) is shown on Figure 3 and is about 2.5 times better than the conventional counter method.

Short Term Stability

The oscillator stability (O), combining conventional frequency counter data and our "zero-beat" technique data is shown in Figure 3. A factor of $\sqrt{2}$ has been removed 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 also shown in Figure 3. At short times 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. Above 0.2 second it reaches an apparent floor of 3 to 4 x 10^{-15}, which extends to about 100 seconds. Amplifier flicker phase noise will be present in the oscillator output as it is coupled from the loop immediately after the amplifier. The measured stability below 0.1 s is consistent with amplifier flicker phase noise of S_ϕ = -100/f dBc Hz^{-1} which translates to an Allan deviation of approximately 5 x $10^{-16}/\tau$. The small degradation in the Allan deviation around 2 s is due either to the 0.2 Hz beat between the two 1.4 MHz frequency stabilization servo reference oscillators, or to the oceanic microseism peak at 0.2 Hz which has an amplitude of up to 2 x 10^{-6} g. We had no accelerometers sensitive enough at 0.2 Hz to eliminate the latter possibility.

We have 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. We were unable to make an equivalent measurement for oscillator 1 because its cryogenic detectors are inaccessible. The inferred Allan deviation due to the noise floor of the frequency control servo (F) improves as $1/\sqrt{\tau}$, as is expected for the white frequency noise of the discriminator. 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. In preliminary measurements the floor in oscillator 1 was much lower than that in oscillator 2, primarily due to the lower frequency vs power slope at the operating power in oscillator 1.

The requirement for temperature stability in resonator 2 is very strict as its minimum slope is 3 Hz K^{-1}. The resonator temperature is controlled by a Lake Shore DRC-91CA temperature controller. The oscillator's stability indicates that it must be controlling the temperature to better than about 10 μK although we have been unable to verify this independently. The inferred Allan deviation from a measurement of a germanium thermometer near the resonator by a Quantum Design Model 1802 digital R/G bridge is 10^{-14} which is the same as the quoted stability of this bridge. We need a higher resolution measurement system if we are to determine the actual temperature stability.

The above measurements show that the short term stability is not limited by any of the servo systems except between 0.07-0.2 s where it appears to be limited by the frequency control servo on oscillator 2. This is confirmed by the lack of correlation between the inputs or outputs of the servo systems and the zero-beat phase noise of the two oscillators, except for a few bright lines. Temperature instability of resonator 2 hasn't been eliminated as a candidate although it would seem unlikely that it would dominate from 0.1 to 1000 s.

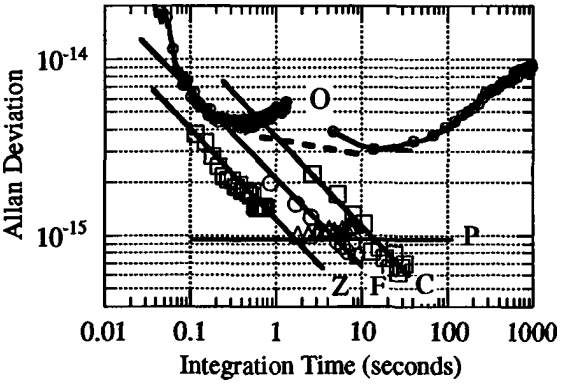


Figure 3 Frequency stability of sapphire oscillator (O), noise floors of the measurement system (Z, C) and oscillator 2 servo systems (P, F).

The origin of our apparent flicker frequency floor is uncertain although it may be due to intrinsic resonator flicker frequency noise. Ground vibrations of order 10^{-6} g might also be significant in the short term since the phase noise below 1 Hz is partially correlated with accelerometers mounted on the cryostat and the acceleration sensitivity in resonator 1 appears to be a few times 10^{-9} g^{-1}. Stability of the order of 10^{-15} is only

realized with one other oscillator - the superconducting cavity maser [1], which also uses a cryogenic SDR and exhibits a similar flicker frequency floor from 1 to 1000 s.

This extreme level of short term stability could easily be transferred to an atomic frequency standard, using the intrinsic resonator power sensitivity to enable a low bandwidth frequency lock loop. The relatively large frequency-power coefficient gives the system adequate range to cope with long term drifts of the sapphire oscillator.

Improvements in the Long Term Stability

The addition of a power servo significantly improves the Allan deviation in the medium term although the performance in the long term still begins to degrade as approximately $3 \times 10^{-16} \tau^{1/2}$. The relative drift of the two oscillators over 1 day is a few times 10^{-13}, which is an order of magnitude improvement over the oscillators based on 3 cm resonators [2]. This long term drift is dominated by the temperature coefficient of the power detector sensitivity on oscillator 2 which resulted in a sensitivity to room temperature of about 5×10^{-13} K^{-1}.

In order to better determine the intrinsic long term stability of the sapphire resonator we have reduced the room temperature sensitivity on one oscillator (oscillator 1) and compared it to a hydrogen maser. The room temperature sensitivity was reduced to ~ 10^{-13} K^{-1} by employing cryogenic tunnel diode detectors for the power and frequency control servo detectors. The residual temperature sensitivity is then dominated by the reference phase modulator (PS1). 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 in-phase component of the AM from the reference signal in PS1 masquerades as a frequency offset error at the discriminator. In a few 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. Unfortunately, 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 a 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). With this AM servo system the temperature sensitivity of our oscillator is less than 3×10^{-14} K^{-1}.

Comparison with H-maser: The sapphire oscillator was compared with a Shanghai Observatory H-maser model H-M12A. The stability of this maser has been previously determined against a Smithsonian VLG-10-1 maser [11], at 800 MHz. The sapphire oscillator is heterodyned with a HP8673H microwave synthesizer whose 5 MHz reference is phase locked to the 100 MHz output of the H-maser. Both the sapphire oscillator and H-M12A are housed in the same temperature controlled room, whose diurnal temperature fluctuations are less than 0.5 K. Short term temperature oscillations of 1 K amplitude with a typical period of 1600 s are due to the air-conditioner thermostat.

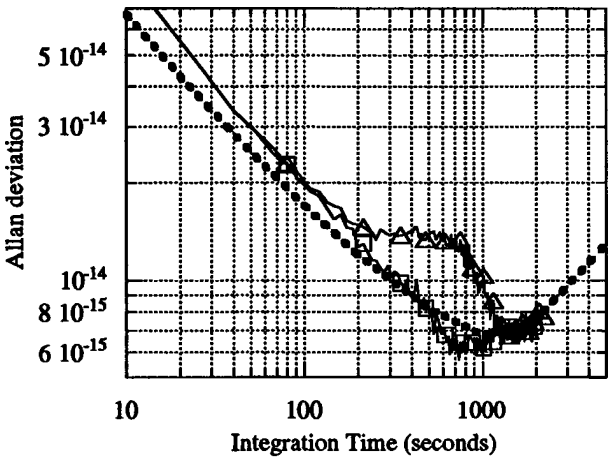


Figure 4 Comparison of Shanghai H-maser: with VLG-10-1 (thick dotted line), with the Sapphire Oscillator (triangles), and with the Sapphire Oscillator with the effects of room temperature oscillation removed (squares).

The stability measured over 13 hours (triangles) is shown in Figure 4 along with the results of the two maser intercomparison (thick dotted line). We attribute the difference in the curves at small integration times to the poor phase noise of the maser quartz oscillator when multiplied to 11.9 GHz. The peculiar Allan deviation curve (triangles) shown in Figure 4 probably arises from frequency modulation of the H-maser signal because of oscillation in the room temperature. The magnitude of the frequency modulation is consistent with the maser sensing the room temperature oscillation filtered by a 90 minute internal thermal time constant. The temperature coefficient of the H-maser is approximately 4×10^{-14} K^{-1}. This frequency modulation could also be explained by assigning a room temperature sensitivity of 1×10^{-14} K^{-1} to the sapphire oscillator. Further work needs to be done to isolate the source of the temperature sensitivity. By filtering out the effect of the temperature modulation from the data we can obtain the result

(squares) for a temperature stable measurement. The filtered stability from 600 to 2000 s is less than or equal to the results of the two maser intercomparison [11] Taking into account the errors in the Allan deviation at these integration times (50 data points) we arrive at an estimate of the instability associated with the sapphire oscillator of less than 4×10^{-15} from 600 to 2000 s.

Conclusions

We have demonstrated that a power, phase, AM and temperature stabilized loop oscillator based on a cryogenic sapphire resonator can achieve fractional frequency stability of 3 to 4×10^{-15} for integration times from 0.3 to 100 seconds and less than 4×10^{-15} from 600 to 2000 s. We have shown that the noise floors of the servo systems aren't responsible for imposing the present flicker frequency floor. The origin of this floor is uncertain although it may be due to intrinsic resonator flicker frequency noise. We have also shown that the long term drift can be significantly reduced using cryogenic power servo detectors and by implementing a new AM servo system which reduces the effect of the temperature sensitivity of the phase modulators.

Acknowledgments

This research was supported by the Australian Research Council. We wish to thank Dr. Warwick Wilson of the Australia Telescope National Facility for providing us with data on the comparison of the Smithsonian and Shanghai Observatory H-masers. We also thank the ATNF and Dr. C.F. Lin and J.W. He of Shanghai Observatory for the loan of the H-maser.

References

[1] G.J. Dick and R.T. Wang, "Ultra-Stable Performance of the Superconducting Cavity Maser", IEEE Trans.Instrum. Meas., IM-40, 1991, pp 174-177.

[2] 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," in Proc. IEEE 43rd Annual Frequency Control Symposium, 1989, pp 89-93.

[3] A.N. Luiten, A.G. Mann and D.G. Blair ,"Cryogenic Sapphire Microwave Resonator-Oscillator with Exceptional Stability", Electronics Letters, 30, 1994, pp 417-419.

[4] A N Luiten, A G Mann and D G Blair, "Ultra High Q Factor Cryogenic Sapphire Resonator", Electronics Letters, vol 29, 1993, 29, pp. 879-881.

[5] M.E. Costa, A.N. Luiten, M.E. Tobar and D.G. Blair "Oscillator Performance from the time evolution of relative phase", Electronics Letters, 30, 1994, pp 149-151.

[6] A.N. Luiten, A.G. Mann, A.J. Giles and D.G. Blair, "Ultra-Stable Sapphire Resonator-Oscillator", IEEE Trans. Instrumentation and Measurement, vol 42, 1993, pp 439-443.

[7] A.N. Luiten, A.G. Mann and D.G. Blair,"Improved Sapphire Dielectric Resonators for Ultrastable Oscillators", Proc. 47th Annu. Freq. Control. Symp., Salt Lake City UT, 1993, pp. 757-762.

[8] V.B. Braginskii, V.P. Mitrafanov and V.I. Panov Systems with Small Dissipation. Chicago: University of Chicago, 1985, p 78.

[9] A.N. Luiten, A.G. Mann and D.G. Blair , "Automated System for High Resolution Measurements of Q, Coupling and Resonant Frequency of a Microwave Resonator", to be submitted to IEEE Trans. Instrum. Meas.

[10] S. R. Stein and J. P. Turneaure, "The Development of the Superconducting Cavity Stabilized Oscillator", Proc. 27th Annual Frequency Control Symposium, 1973, p414 .

[11] W.E. Wilson, "Shanghai Maser Tests at Parkes", ATNF, CSIRO, Jan 1994.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SAPPHIRE DISK DIELECTRIC RESONATOR TEMPERATURE COEFFICIENT OF FREQUENCY DEPENDENCE ON TEMPERATURE, DISK CONFIGURATION AND RESONANT MODE

D.P.Tsarapkin

Radio Engineering faculty
Moscow Power Engineering Institute, Moscow 111250, RUSSIA

Abstract: Results are presented for leucosapphire disk dielectric resonator temperature coefficient of frequency (TCF) dependence on temperature, disk configuration and a resonant mode type in the temperature range 70...350 K. It is found that the most intensively the enumerated factors influence resonant oscillations with HE-type polarization. The revealed TCF(T) difference for different modes permitted to build a novel long term frequency stabilization method.

Introduction

A "whispering gallery" mode sapphire disk dielectric resonator (DDR) [1-3] is capable of extremely high Q-value due to the fact that DDR's high number circumferential modes isolate electromagnetic energy into the sapphire element itself and away from lossy environment. It results in room and cryogenic temperature leucosapphire DDRs are growing an important component for low phase noise microwave and millimeter wave oscillators.

Long term frequency stability DDR-based oscillators depends first of all on the DDR eigenmode frequency changing with temperature. A sapphire DDR is produced from a high quality monocrystal. Unlike quartz resonators a specimen crystal axes orientation is strictly fixed in a given case. Nevertheless one can find in literature [4-8] the data on sapphire DDR TCF ranging from -40 to -70 ppm/K at room temperature and from -2 to -8 ppm/K at 77K. These scatterings are much in excess of usual experimental errors but their origin (and often even their presence) is still unknown.

This paper provides computer analysis of this problem. The results of modelling is being compared partly with experimental data.

Main Results

The electromagnetic field within a DDR contains all six components. Thus there are no pure H- or E-type mode polarization in the case. Usually one distinguish HE-modes with E-vector on the cylindrical surface strongly polarized in parallel with the optical axe (z-axe) which is a disk axe at the same time and EH-ones with the ortogonal orientation of the E-vector in respect of the previous case. We analyze main family of resonances characterized with the indexes set $HE(EH)_{N,1,1}$.

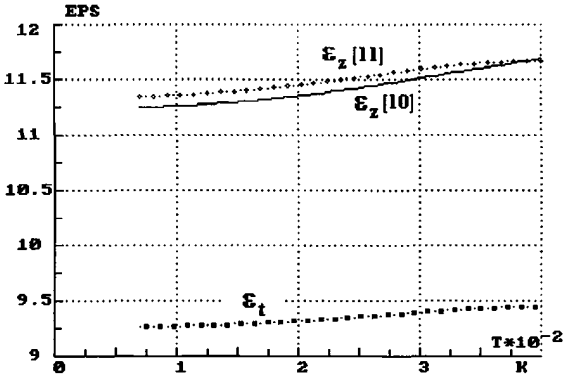


Fig.1. Sapphire relative permittivity components versus temperature.

The essential temperature dependence of a DDR self resonant frequency arises mainly due to increasing with temperature the both of sapphire dielectric permittivity tensor components, ε_z and ε_t, while thermal contraction of the resonator has secondary meaning.

Each particular resonator eigenmode differs from another modes with some details in the picture of the internal electromagnetic field and the part of its energy in surrounding space. The functions $\varepsilon_z(T)$ and $\varepsilon_t(T)$ are resemble but non-identical. Thus it is naturally to expect that different DDR modes have to have more or less different TCF values and this difference is getting more together with growth of the distinctive signs. The following material illustrates this basic idea.

The TCF values have been calculated through eigenmode frequencies temperature dependence using an IBM PC. The method is based on a single-mode representation of electromagnetic fields in partial regions of the DDR. The used computer program was an improved version of the program in [8] written by E.N.Ivanov. The next step of the algorithm development and its description one can find in [9].

0-7803-1945-1/94 $3.00 © 1994 IEEE

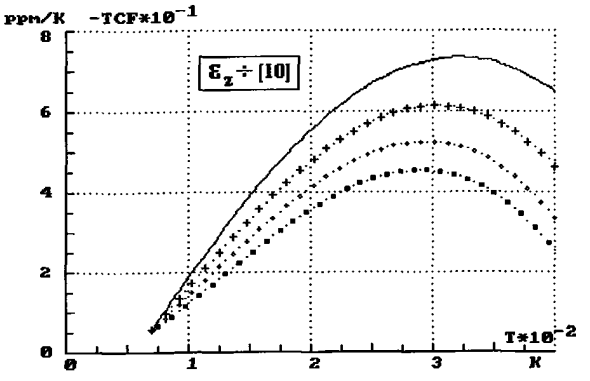


Fig.2. Sapphire DDR TCF(T).

Mode: HE_ 7 D/2h = 1.25 (———)
HE_ 7 = 2.50 (× × ×)
HE_ 7 = 3.75 (o o o)
HE_ 7 = 5.00 (▪ ▪ ▪)

To get reliable results one needs an accurate knowledge of sapphire permittivity constants as a function of temperature. Two sources of the initial data describing these functions were used: the "data 1" in accord with [10] and the "data 2" published last year [11]. The chosen data sets show practically the same dependencies with respect of the dielectric constant in the transverse to a disk axe dimension but differ rather distinguishably as far as the second dielectric constant (along the z-axe) is concerned (Fig.1). The polinomial fit used for the solid line yields:

$$\varepsilon_z(T) = 11.2727 - 7.5908\times10^{-4}T + 7.0650\times10^{-6}T^2$$

$$-5.5734\times10^{-9}T^3 + 2.1283\times10^{-12}T^4 \qquad (1)$$

HE-modes

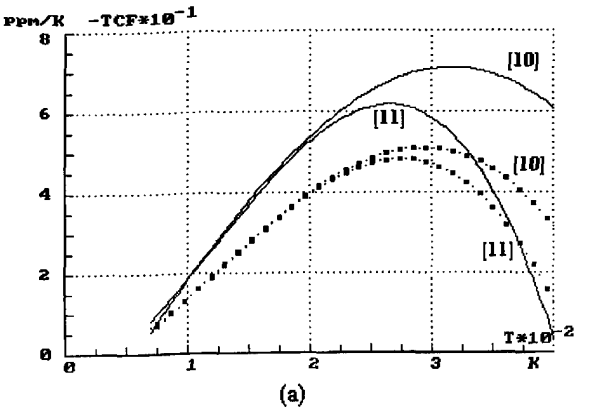
The first factor which role we investigate is the "disk format"

$$FT = D/2h, \qquad (2)$$

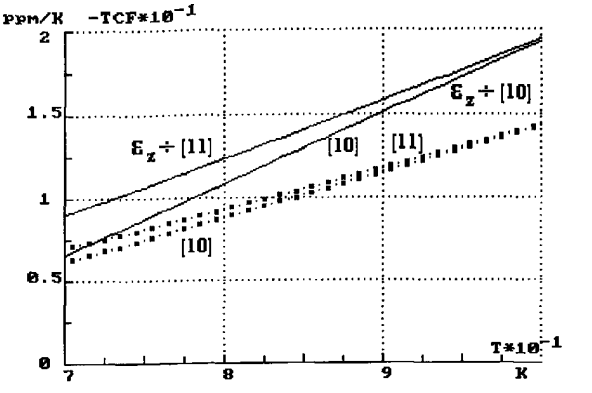
where D - a diameter, h - a height of the disk. According Fig.2 this parameter has strong influence on sapphire DDR TCF and the TCF absolute value decreases while format is getting more.

Quite unexpectedly the curves in Fig.2 reveal that the TCF temperature dependence is a non-monotone function within the specified temperature range but having a maximum round about 300 K. As for the TCF value it changes 6-8 times.

The comparison of the results got with the two chosen $\varepsilon_z(T)$ approximations (Fig. 3a) shows that the second one (i.e. [11]) leads to reduced TCF modulus and the maximum in the case takes place near 260-280 K. The difference in the TCF values for these approximations is especially pronounced for T >300K.



(a)



(b)

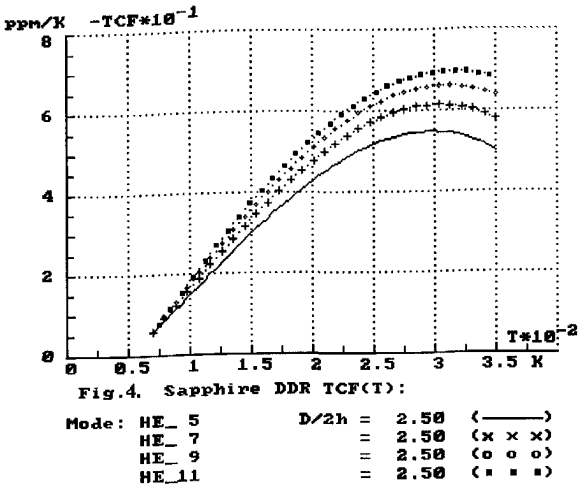
Fig.3. Sapphire DDR TCF(T):

Mode: HE_ 7 D/2h = 1.50 (———)
HE_ 7 = 4.00 (▪ ▪ ▪)

Fig. 3b gives the details of the curves behavior in the liquid nitrogen temperature range. One can see that here the TCF is more sensitive to temperature if we take the first $\varepsilon_z(T)$ case and small FT values.

Once more important parameter characterizing oscillations in a DDR is a azimuthal index, N. Its influence is illustrated by a set of curves in Fig.4. It is obvious that the more N the more TCF modulus. Some additional information on that subject can be derived from Fig.5 where we compare the curves calculated for two approximations with our experimental data [8]. The theoretical data here and before take into consideration also the thermal disk sizes change which impact is about 10%.

Fig.5 leads to the conclusion that approximation $\varepsilon_z(T)$ in accord with [10] gives better fit to the experimental data. It is useful to mark that for the azimuthal indexes less than 8 the two dependencies run very closely to each other. Thus one can not probably find this difference for the case of a big format. A lack of precise experimental data prevents to get more valuable inferences.



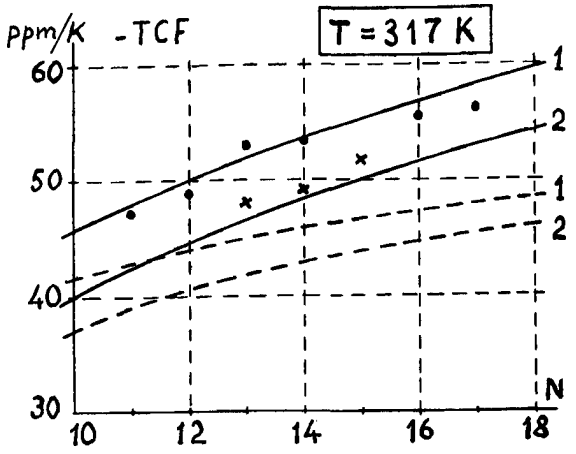


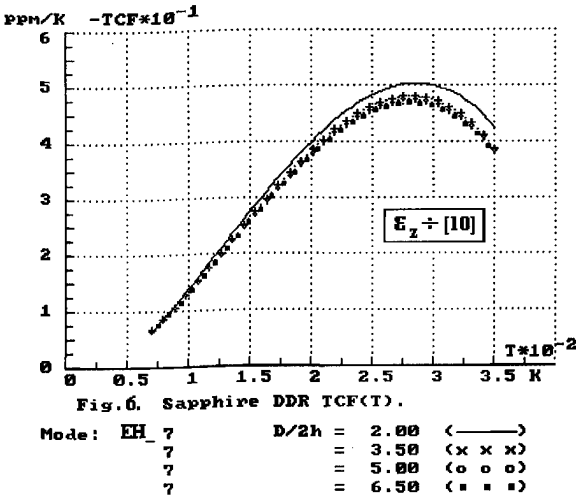
Fig.5. Sapphire DDR TCF versus azimuthal index. The solid lines correspond to $\varepsilon_z \div [10]$, the dashed ones - $\varepsilon_z(T) \div [11]$. The curves marked by a figure 1 and the experimental circles are drawn for FT = 6.37, two another curves and the experimental crosses - FT = 7.85.

EH-modes

For the case of EH-modes Fig. 6 has the same sense as Fig. 2 for HE-modes demonstrating the influence of a disk format on the function TCF(T). Again one can see that growth of the format makes TCF modulus less. But in comparison with the Fig. 2 it follows that the effect itself is much less pronounced. The curves maximize near 280 K.

Comparison of the two approximations (Fig.7) shows that for FT > 4 they do not differ practically and even with FT ~ 2 the difference is clear only at the high enough temperature.

At low end of our specified temperature range the two chosen $\varepsilon_z(T)$ functions give close results.



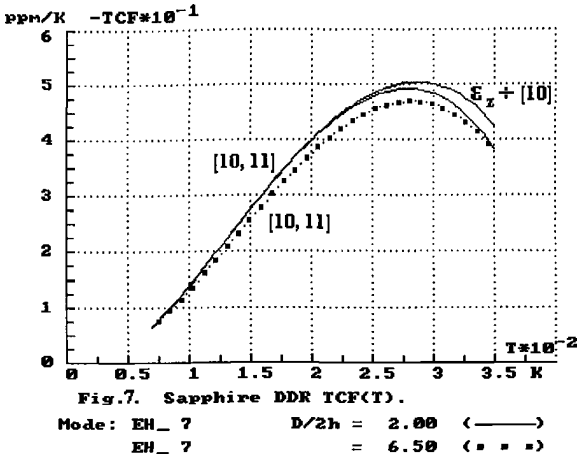
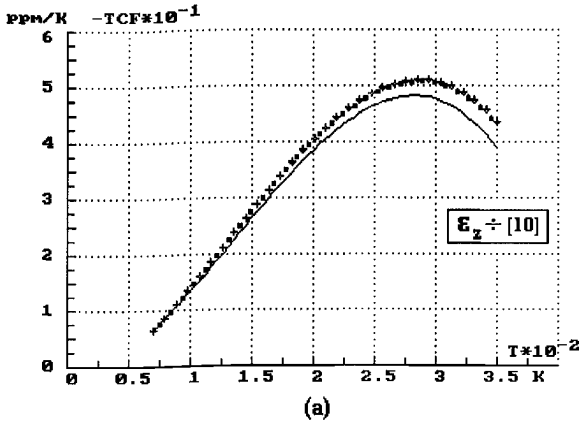


Fig. 8 give some insight into the azimuthal number influence. Once one can see the weak but qualitatively the same as for HE-modes effect. At the liquid nitrogen temperature, 77 K, all modes have practically identical TCFs (Fig. 8b).

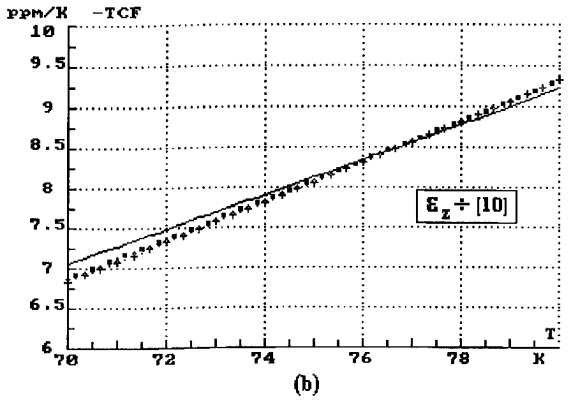
Summarizing we can say that the results for EH-modes resemble ones for HE-modes except all effects are somewhat weaker and more complicated in some details.

One particular practical application

The revealed TCF differences for different sapphire DDR modes permit to build a novel long term frequency stabilization method. Its main idea consists in that the AFC system watches now not for the output microwave frequency, as usually, but for chosen by us suitable frequency difference between the two DDR modes with different TCF. This idea was checked and patented [12].



(a)



(b)

Fig.8. Sapphire DDR TCF(T).

Mode: EH_ 6 D/2h = 2.50 (———)
EH_ 7 = 2.50 (× × ×)
EH_ 8 = 2.50 (o o o)
EH_ 9 = 2.50 (▪ ▪ ▪)

Conclusion

Our investigation revealed that due to sapphire crystall anisotropy the disk resonator TCF depends not only on temperature but on a disk configuration factor, a polarization of the basic slow wave (HE- or EH-type), and a mode number as well. In result we were able to suggest a novel technique to improve long term DDR-based oscillator stability.

References

[1] V.S.Dobromislov and V.F.Vzjatishev, "Whispering gallery" Mode Dielectric Resonators", Proceedings of Moscow Power Eng. Inst., Moscow, iss.161, pp.78-84, 1973 (in Russian).

[2] V.F.Vzjatishev, V.S.Dobromislov, V.L.Masalov et al., "On One Possibility of Super-High-Q Resonators Realization", Proceedings of Moscow Power Eng. Inst., Moscow, iss. 360, pp.51-57, 1978 (in Russian).

[3] E.N.Ivanov, I.N.Mukhtarov, and D.P.Tsarapkin, "High-Q Disk Dielectric Resonators", Radiotechnika i Electronika, 1983, vol.28, No.8, pp. 1658-1659 (in Russian).

[4] S.N.Bunkov, B.A.Vtorushin, V.N.Egorov et al., "Cooled Dielectric Resonators for Frequency Stabilization", Radiotechnika i Electronika, 1987, vol.31, No.5, pp. 1071-1081 (in Russian).

[5] S.K.Jones, D.G.Blair, and M.J.Buckingham, "Effect of Paramagnetic Impurities on Frequency of Sapphire-Loaded Superconducting Cavity Resonators", Elec. Letters, 1988, vol. 24, iss. 6, pp. 346-347.

[6] S.L.Abramov, E.N.Ivanov, and D.P.Tsarapkin, "Low noise microwave oscillator with a temperature stabilized disk dielectric resonator", Radiotechnika, 1988, No.11, pp. 81-83 (in Russian).

[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., vol. UFFC-37, pp. 339-346, Sept. 1990.

[8] D.P.Tsarapkin, S.L.Abramov, and E.N.Ivanov, "On Difference of Temperature Coefficients of Frequency of Leucosapphire Disk Dielectric Resonator Different Modes", Radiotechnika i Electronika, 1990, vol.35, No. 2, pp. 431-432 (in Russian).

[9] E. N. Ivanov, D.G.Blair, and V.I.Kalinichev, "Approximate Approach to the Design of Shielded Dielectric Disk Resonators with Whispering-Gallery Modes", IEEE Trans., vol. MTT-41, No. 4, pp. 632-638, April 1993.

[10] S.N.Bunkov, V.N.Egorov, and T.Y.Ri, "Sapphire Dielectric Permittivity Dependence on Temperature", in Digest of Conference on "Design of Dielectric Waveguide-and-Resonator Based Devices and Systems", 25-27 Oct.1988, Tbilisi, FSU, pp.253-254, 1988 (in Russian).

[11] C.A.Flory and R.C.Taber, "Microwave Oscillators Incorporating Cryogenic Sapphire Dielectric Resonators", in Proc. 47th Annual Freq. Contr. Symp., 1993, pp.763-773.

[12] D. P. Tsarapkin, S. L. Abramov, and E. N. Ivanov, "Microwave Source", S.U. Patent #1,605,890, int. cl. H 03 B 7/14, July 8, 1990.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

LOW PHASE NOISE SAPPHIRE DISK DIELECTRIC RESONATOR OSCILLATOR WITH COMBINED STABILIZATION

D.P.Tsarapkin

Radio Engineering faculty
Moscow Power Engineering Institute, Moscow 111250, RUSSIA

Abstract: This paper deals with a theoretical analysis of a low phase noise microwave transistor oscillator stabilized with a sapphire "whispering gallery" mode disk dielectric resonator which used both as a filter in the oscillator feedback circuit and as a dispersive element of a frequency discriminator in the AFC loop providing additional noise degeneration.

It is found that in many cases a noise floor of the AFC system is determined with residual flicker noise of a ferrite circulator used as a part of the discriminator circuitry. Replacing the circulator by a 3-dB hybrid and adding the auxiliary amplifier(s) at the phase detector input(s) open a possibility to reduce the AFC system noise floor as much as 30-35 dB and thus to get impressive improvement in the oscillator close-in phase noise.

Introduction

Progress in radar, microwave communication systems and some fundamental scientific experiments necessitates achievement of high degree of reference oscillator output signal spectral purity. The best nowadays microwave sources use bipolar transistors (BT) as active elements and a sapphire disk dielectric resonator (DDR) excited on a "whispering gallery" mode as a filter. Since the first successful application in oscillators in 1980 the sapphire DDRs have become the main means to generate microwave signals with spectral purity performance characterized in the middle of a centimetric band by $\pounds(1\ \text{kHz}) \approx -120$ dB/Hz at room temperature and even much lower levels with cryogenic cooling [1-3].

Mainly two DDR oscillator utilization schemes are in use now: a feedforward amplifier with the DDR included in the positive feedback circuit as a band-pass filter, and a commercial oscillator which is frequency locked to a DDR used as a frequency discriminator. Each of them has its merits and demerits. While the first one is simpler the second provides lower noise due to better diode noise characteristics in comparison with transistors.

The purpose of this paper is to consider the advanced "combined stabilization" configuration where a single DDR is used both as a positive feedback loop filter and a basic element of the AFC frequency discriminator with special attention to phase noise analysis and parameters optimization.

The idea of such a configuration appeared more than 20 years ago as a result of the further development of the works collected in [4]. That time this approach was used in TWT- and drift tube klystron oscillators. Later the same design was spread on semiconductor oscillators [5] and used successfully by A.J. Giles et al. [2] in a cryogenic DDR oscillator.

Use of the combined stabilization scheme reduces an initial level of phase noise which has to be additionally suppressed by the AFC system thus making easier the loop low frequency amplifier (LFA) operation. Another very important advantage is elimination of the frequency acquisition problem restricting a working range of stabilized sources with the only AFC loop [3, 6-8].

Oscillator Block Diagram

Figure 1 shows a schematic diagram of the oscillator under analysis which includes two basic parts: the positive feedback loop (its components are drawn with thick lines) and the AFC system. The feedback loop incorporates "main" amplifier (MA), i. e. a BT microwave sustaining stage(s), a DDR, a phase shifter, a 3-dB hybrid and a limiter. The last one is not a compulsory element but useful to keep the MA transistors in a quasi-linear region thus reducing their residual noise [9]. The 3-dB hybrid junction serves to split the whole MA output, P_{MA}, onto a load power, P_{out}, and a signal passing to the resonator. In general case this hybrid insertion loss, L_L, can have rather arbitrary value.

Three other blocks situated in fig.1 in the rf feedback loop have to do with the AFC system. They consist of the device picking out a wave reflected from the resonator (a "reflected wave extractor" - RWE) and two directional couplers. In the most of works on the topic the RWE is a ferrite Y-type circulator. As known this function can be fulfilled equivalently with a suitable reciprocal circuit. Our particular interest further will be the case where the mentioned 3-dB hybrid plays also a role of the RWE (a dashed output line from the hybrid symbol) thus providing a possibility to eliminate the circulator.

0-7803-1945-1/94 $3.00 © 1994 IEEE

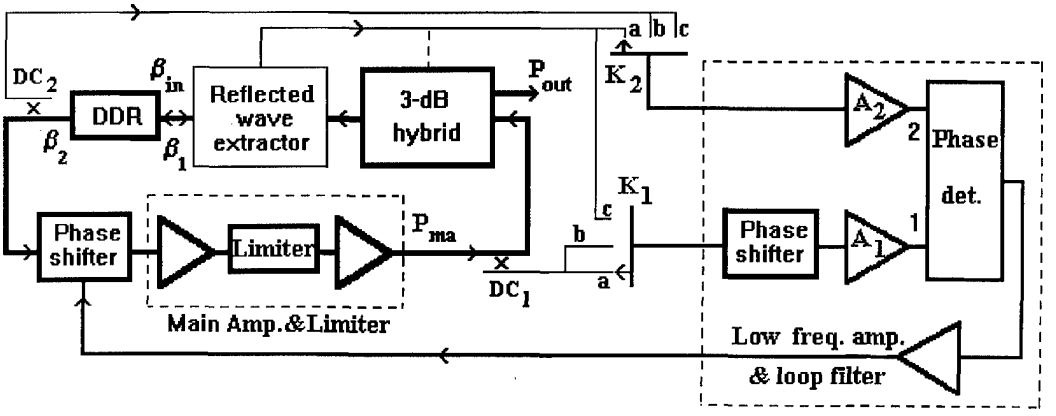


Figure 1. Block diagram of the DDR based oscillator with combined stabilization. The right-hand side dashed portion contains some elements associated with the resonator-based frequency discriminator: a double-balanced mixer (phase detector), an adjustable phase shifter, two auxiliary microwave limiting amplifiers and a low frequency amplifier & AFC loop filter. The K1, K2 keys allow to choose any of three standard frequency discriminator configurations (options "a", "b", "c").

The K_1, K_2 keys allow to choose any of three standard resonator-based frequency discriminator (FD) configurations. For the option "a" ("A"-type FD or FD_a) the phase detector (PD) processes a constant amplitude signal splitted out with the directional coupler DC_1 and gained (if necessarily) by G_1 dB due to presence of the auxiliary microwave limiting amplifier A_1 (PD input 1) and a reflected wave gained by G_2 dB due to the second auxiliary microwave amplifier A_2 (PD input 2). If the A_2 amplifier does not provide limiting then the amplitude of a signal at the second PD input is strongly frequency dependent and there is a real danger to damage PD during tuning. Use of two additional amplifiers instead of one proposed in [10] enlarges our ability to optimize the whole system. Another two cases exploit a part of energy of the incident wave and transmitted one (option "b", FD_b) or use the pair of reflected and transmitted waves (option "c", FD_c).

An uncooled sapphire DDR has very a poor frequency coefficient with respect to temperature and needs usually some kind of its center frequency stabilization. All such circuits are quite inertial as a rule and so not influence noise processes corresponded with offset frequenses more than ≈1 Hz. That is why we do not show any of them in fig. 1.

Oscillator Noise Model

Output phase noise arises from some primary noise sources filtered by both incorporated stabilizing mechanisms. Let us consider firstly phase noise in a free running oscillator to put into some definitions.

An equivalent initial phase perturbation, $\theta(t)$, at the input of a loop amplifier (the MA with its power gain, G_{MA}, in our case) is enhanced by a positive phase feedback and filtered by the resonator. According Leeson's model the closed loop response due to a stimulus $\theta(t)$ at the output of the amplifier is :

$$\varphi(t) = \theta(t) \times (1 + 1/q_r p), \quad (1)$$

where $p = j\xi = d/d\tau$ – a differentiation operator with respect to normalized time τ ; $\xi = f_m / B_0$; f_m – an offset (modulation) frequency from the average carrier frequency f, which considered to be equal a self resonator frequency f_0 for simplicity; $B_0 = f_0/2Q_0$ – a half bandwidth of the unloaded resonator; Q_0 – an unloaded resonator quality factor; Q_L – a loaded resonator quality factor; $q_r = Q_L/Q_0$ – a normalized resonator loaded Q-factor. The oscillator single side band phase noise spectral density at the amplifier output described by :

$$\pounds_{out}(f_m) = \tfrac{1}{2} S_{MA} [1+1/(q_r \xi)^2] = \tfrac{1}{2} S_{MA} [1+ (B_0 /q_r f_m)^2]. \quad (2)$$

At the amplifier input due to filtering by the resonator the free running oscillator phase noise is:

$$\pounds_{in}(f_m) = \tfrac{1}{2} S_{MA}(f_m) /(q_r \xi)^2 = \tfrac{1}{2} S_{MA}(f_m) \times (B_0 /q_r f_m)^2. \quad (3)$$

The amplifier phase noise spectral density, $S_{MA}(f_m)$, shows a 1/f ("flicker") characteristic empirically described by the corner frequency, f_c. For microwave BT f_c can range from 3 to 20 kHz approximately depending on the device type and a drive level. Thus for generic offsets non-exceeding 1 kHz the oscillator regime is disturbed namely by 1/f noise.

As was found recently [11] a spectral density of multiplicative

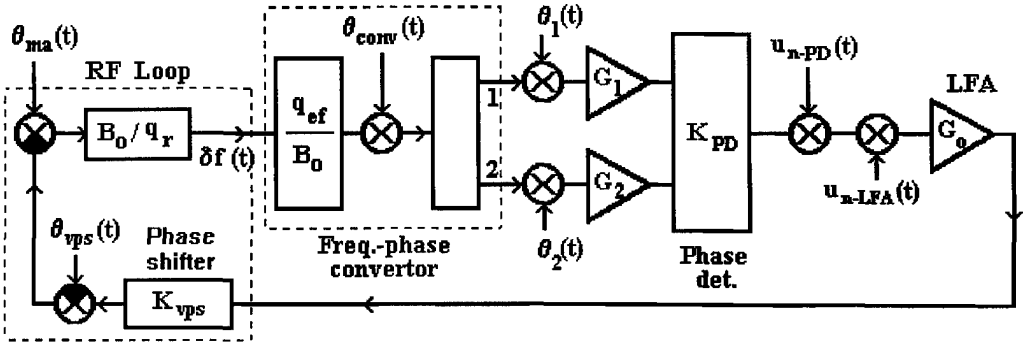


Figure 2. Block diagram describing phase noise of the analyzed oscillator with combined stabilization.

phase noise of a single microwave BT amplifier stage is approximately directly proportional to the stage power gain. So with oscillator noise performance improvement point of view it is profitable to reduce the MA gain and optimize the amplifier structure. For a given total multistage amplifier gain, G_{amp}, its phase noise reaches a minimum if the amplifier consists of

$$n = n_{opt} \approx 0.23 \times G_{amp}\,(dB) \qquad (4)$$

identical stages. Accordingly, each of the optimized stages has the gain

$$G_j = G_{amp} / n, \ [dB] \qquad (5)$$

and its phase noise spectral density may be described by the expression

$$S_j = S_{\varphi o} + G_j, \quad [dB/Hz] \qquad (6)$$

where the "level of support", $S_{\varphi o}$(dB/Hz), corresponds to a stage with 0-dB gain. The total amplifier phase noise spectral density results from separate stage contributions added on a root-sum-square basis:

$$S_{amp} = S_j + 10\lg n = S_{\varphi o} + G_{amp}/n + 10\lg n, \quad [dB/Hz]. \qquad (7)$$

This approximate noise model of a BT stage is true for f > 2...3 GHz where needed gain is got due to a transistor internal positive feedback.

Resonator coupling with incoming wave lines is characterized with the normalized coupling coefficients β_1 and β_2 (fig.1). It follows the resonator normalized input impedance, β_{in}, from the RWE side is:

$$\beta_{in} = \beta_1 / (1 + \beta_2). \qquad (8)$$

The theory shows that β_1, β_2, Q_L, G_{MA} and £ are interdependent.

Minimum sideband noise in the free running oscillator can be achieved near $\beta_1 = \beta_2 = ½$ when $q_r = (1 + \beta_1 + \beta_2)^{-1} = ½$ and resonator transition losses at resonance

$$L_r = (1 + \beta_1 + \beta_2)^2 \times (4\beta_1\beta_2)^{-1} = 4 = 6\ dB. \qquad (9)$$

Let f_o = 10 GHz, f_m = 1 kHz, $Q_{o\,DDR} = 2\times10^5$, $L_L = -3$ dB, a feedback loop excess gain on small signal L_{eg} = 3 dB, $S_{\varphi o}$ = -157 dB/Hz {$S_{\varphi o} = 2\times10^{-13}/f_m$ [rad²/Hz]}, $\beta_1 = \beta_2 = ½$. Then in accord with (2)-(9)

$$G_{MA} = L_r - L_L + L_{eg} = 12\ dB, \qquad (10)$$

n = 2, S_{MA}(1 kHz) ≈ -148 dB/Hz, B_o = 25 kHz, $q_r = ½$, £$_{in}$ ≈ £$_{out}$ = −117 dB/Hz. It is 3 dB less in comparison with a single stage MA and practically the same as for a three stages MA.

Now go back to the complete system. The oscillator phase balance equation has a form

$$\varphi_{ma} + \varphi_{DDR} + \varphi_{VPS} + \varphi_{AFC} = -2\pi m \quad (m = 0, 1, 2, ...). \qquad (11)$$

Here φ_{ma}, φ_{DDR} and φ_{VPS} are phase delays in the MA, the DDR and the variable phase shifter (VPS), accordingly, φ_{AFC} describes an additional phase shift due to the AFC system. Each of these terms may be thought as a sum of regular and noise parts, i.e.

$$\varphi_j(t) = \varphi_{jo} + \theta_j(t). \qquad (12)$$

Put for simplicity φ_{AFCo} = 0 at an average frequency of oscillations and eliminate index "o". Then this term if being unequal zero represents the phase correction providing by the AFC system. From (11)-(12) it follows that within a half bandwidth of the DDR a deviation of an instantaneous frequency from its average value is:

$$\delta f/f = \delta f_o/f + (\theta_{ma} - \varphi_{AFC} - \theta_{AFC}) / (2\,Q_{L\,DDR}). \qquad (13)$$

The first term of the (13) right-hand side takes into account the DDR self frequency instability. According [12] the upper limit to the DDR close-in residual phase noise does not exceed a mixer noise floor at least. So within some approach we can ignore this member.

The AFC loop itself includes a frequency-to-phase converter (simply "converter" further), auxiliary limiting amplifiers (ALA), a phase detector, a LFA which is a shaping filter simultaneously and the just mentioned VPS.

A conversion factor of the converter, K_{conv}[rad/Hz], depends on the FD type and a DDR quality factor. "Higher sensitivity is an advantage associated with utilization of resonator reflection response (compared to transmission response) in the discriminator circuit" [7]. This true idea is a consequence of the fact that the closer a resonator to exact matching (which implies $\beta_{in} \rightarrow 1$) the steeper a characteristic phase slope of the reflected signal [13]. At f_o

$$K_{conv} = q_{ef} / B_o \qquad (14)$$

where a normalized convertor Q-factor, q_{ef}, is :

$$q_{ef\text{-}a} = \mathrm{abs}\{[(1 - \beta_{in})(1 + \beta_2)]^{-1} - q_r\} \text{ (option "a") ; } (15)$$

$$q_{ef\text{-}b} = q_r \qquad \text{(option "b") ; } (16)$$

$$q_{ef\text{-}c} = q_{ef\text{-}a} + q_r \quad (\beta_{in} < 1) \qquad \text{(option "c") . } (17)$$

For $\beta_{in} \approx 1$, as easy to see, $q_{ef\text{-}a} \approx q_{ef\text{-}c} >> q_{ef\text{-}b}$. In a DDR-based oscillator it is reasonable to choose namely $\beta_{in} < 1$ because due to the DDR features its Q_o-factor decreases rapidly if β_1 or β_2 (or both of them) is getting sufficiently more than unity. Besides, the more β_1, β_2 the less the rf feedback loop Q_L. Thus in our case namely the FD_c provides the most K_{conv}.

Two main conditions of PD residual self noise minimization are optimization of mixer diodes power (usually 1...2 mW per a Schotty diode) and approximately equal power levels at FD inputs. If they are met the phase to voltage conversion factor for the PD at zero output, K_{PD}, is 0.5-0.9 V/rad [6, 7] and the noise floor of the X-band PD in terms of spectral density of output voltage fluctuations, $V_{PD}(f_m)$, might be as low as [6, 14]

$$V_{PD}(f_m) \approx 10^{-14} / f_m + 10^{-17} \quad [V^2/Hz] . \qquad (18)$$

If $\beta_{in} \rightarrow 1$ an amplitude of the reflected wave vanishes. To keep an optimum regime of the PD one needs either to increase the loop power or to insert the ALA [10]. The latter method proposed by G.J.Dick et al. looks especially promising.

Oscillator noise floor

The AFC system output contains a useful error signal $\varphi_{AFC} \propto \delta f$, which, in turn, proportional θ_{MA}, and a noise component θ_{AFC}. Their relation defines the AFC system phase noise floor and thus establishes the phase noise floor of the whole oscillator. The full AFC system output noise is given by (see Fig.2)

$$\theta_{AFC}(t) = [(\theta_{conv} + \theta_1 + \theta_2) K_{PD} + (u_{n\text{-}PD} + u_{n\text{-}LFA})] G_o K_{VPS} - \theta_{VPS} . \qquad (19)$$

The terms on the right-hand side of (19) describe residual phase noise (θ_{conv} - converter; $\theta_{1,2}$ - auxiliary amplifiers; $u_{n\text{-}PD}$ - PD; $u_{n\text{-}LFA}$ - LFA; θ_{VS} - variable phase shifter), the LFA gain (G_o) and a conversion factor of the VPS (K_{VPS} [rad/V]).

In accord with (13, 19) and Fig.2

$$\delta f(t) = [\theta_{MA}(t) - \theta_{AFC}(t)] \times (B_o / q_r) \times [1 + K_{PD} G_o K_{VPS} q_{ef} / q_r]^{-1} . \qquad (20)$$

Thus with the AFC loop closed

$$\pounds_{in}(f_m) = \tfrac{1}{2} [S_{MA}(f_m) + S_{AFC}(f_m)] \times (B_o / q_r f_m)^2 \times [1 + K_{PD} G_o K_{VPS} q_{ef} / q_r]^{-2} \qquad (21)$$

where the total AFC loop phase noise spectral density is :

$$S_{AFC}(f_m) = [(S_{conv} + S_1 + S_2) K^2_{PD} + V_{PD} + V_{LFA})] (G_o K_{VPS})^2 + S_{VPS} . \qquad (22)$$

It is suitable to include V_{LFA} into V_{PD} thus characterizing LFA and PD together and join S_{VPS} to S_{MA}. Then the only terms which distinguish (22) from the usual expression are S_{conv}, S_1, S_2.

From (21-22) it follows that with high enough AFC open loop gain the phase noise floor can reach

$$\pounds_{in\text{-}floor} = \tfrac{1}{2} (S_{conv} + S_1 + S_2 + V_{PD} K^{-2}_{PD}) \times (B_o / q_{ef} f_m)^2 . \qquad (23)$$

The earlier works consider PD residual phase noise as the main reason which influences a AFC system noise floor and ignore in fact all other noise sources. With our point of view namely this assumption led to the 6-10 dB discrepancy between calculated and measured phase noise levels of precise sources as marked in some papers [6, 7].

Circulator noise model

If FD has to use a reflection response one needs to build the RWE. The traditional way to get reflected power implies a presence of a ferrite circulator (Fig.3). Unfortunately this block as we shall show may add substantial noise power to a useful signal.

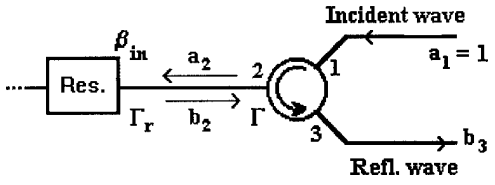


Figure 3. RWE circuitry associated with an Y-circulator.

One can find very a little data on circulator residual flicker noise. Based on works by N.N.Kolachevsky [15] the author arrived at a conclusion that the reasonable estimation for the respective power spectral density of the circulator self flicker noise could be $S_{c\text{-}fd}(f_m) \approx 3\times10^{-14\pm1} / f_m$ [rad^2/Hz] in forward direction which means -(165 ± 10) dB/Hz at 1 kHz offset. At first sight it is quite a negligible value. But in our case the FD employes a specific circulator regime where a relative useful signal due to $\beta_{in}\approx1$ has the same order of magnitude with respect to incident power as a circulator reverse isolation. This particular situation enhances the cyrculator noise influence many times.

Bearing in mind the specific symmetry of the Y-type circulator, the electromagnetic field on its ports may be represented by a sum of two waves (fast and slow) running inside the circulator in opposite directions plus a synphase wave. Assume the vector of an excitation in a form

$$\frac{1}{3}\begin{bmatrix}1\\ \exp(120^\circ)\\ \exp(-120^\circ)\end{bmatrix} + \frac{1}{3}\begin{bmatrix}1\\ \exp(-120^\circ)\\ \exp(120^\circ)\end{bmatrix} + \frac{1}{3}\begin{bmatrix}1\\ 1\\ 1\end{bmatrix} . \qquad (24)$$

Each of the column-vectors in (24) describes a distribution of one particular wave along the ports 1, 2, 3. Take further the reactions of the circuit on these vectors as

$$\mathbf{R}_i \approx [(1-\Delta_i + m_i) + j(\zeta_i + \theta_i)] \exp(j\varphi_i) \quad (i = 1, 2, 3) \quad (25)$$

where $\varphi_1 = -120^\circ$, $\varphi_2 = 120^\circ$, $\varphi_3 = 0$ - nominal values of phase shifts between the ports, Δ_i - losses, ζ_i - phase deviations from the nominal delays, $m_i(t)$ and $\theta_i(t)$ - amplitude and phase fluctuations of $\mathbf{R}_i$. The energy in this circulator moves against a hour-hand. As for an origin of the $\mathbf{R}_i$ fluctuations the main impact goes from random magnetic domain boundaries displacements [15].

The full reaction is a sum of all partial reactions. In the port 3 where one might expect a lack of power in an ideal circulator we have

$$6\times\mathbf{R}_{3\Sigma} = [(\Delta_1 + \Delta_2 - 2\Delta_3) - (m_1 + m_2 - 2m_3) + (\theta_2 - \theta_1)\sqrt{3}] + j\times[(\Delta_2 - \Delta_1)\sqrt{3} + (m_1 - m_2)\sqrt{3} - (\zeta_1 + \zeta_2) - (\theta_1 + \theta_2 - 2\theta_3)] . \qquad (26)$$

It follows from (26), by the way, necessity to tune a circulator to improve its reverse isolation which nevertheless stays restricted because all Δ_i have different values in general. But the most important thing is our found ability to deal with noise processes in a circulator. The "null" in the port 3 reflects deep interference of the waves involved. It is a dynamic phenomenon in principle. In accord with (26) the random phase shift depends on both amplitude and phase disturbances of the partial waves.

Basing on process physics one can postulate

$$\langle m^2_3\rangle < \langle m^2_1\rangle \ll \langle m^2_2\rangle, \quad \langle \theta^2_3\rangle < \langle \theta^2_1\rangle \ll \langle \theta^2_2\rangle \qquad (27)$$

because the slow wave indexed by "2" is "slow" via resonance in ferrite media and so the most sensitive to initial fluctuations. This permits to simplify the task taking into account only the slow wave fluctuations. Of course, this assumption and some next ones must be checked later on.

Let us consider now a working situation with a reflecting resonator in the arm 2. Using S-parameter technique the full relative signal in the port 3 can be put down as (look at Fig.3)

$$\mathbf{b}_3 = \mathbf{S}_{21}\mathbf{S}_{32}\Gamma + \mathbf{R}_{3\Sigma} . \qquad (28)$$

Here

$$\Gamma = \Gamma_r \times \exp(-j\,2\varphi_{cr}) \qquad (29)$$

- the resonator reflection coefficient, Γ_r, phase shifted with doubled phase delay between the port 2 and the resonator;

$$\Gamma_r(f_o) = -(1 - \beta_{in}) / (1 + \beta_{in}). \qquad (30)$$

Some additional tedious transformations come to the expression of the full circulator phase noise in the port 3. It is rather cumbersome containing terms caused by amplitude noise for which reason depends on φ_{cr}. Fortunately after final sorts we get the clear-cut approximate formula :

$$S_{cir}(f_m) \approx S_{c\text{-}fd}(f_m) \times (3 + 1/\Gamma^2). \qquad (31)$$

The integer in brackets arises from two successive passes of the signal through a circulator. Its value (three units instead of two) enhanced due to mutual correlation of phase perturbations of the incident and reflected by a resonator slow waves into a part of the ferrite media volume. The second term in brackets goes from the fact of constant presence of the interference noise in the port 3. Namely this term has deciding influence on the reflected signal phase noise increase if Γ^2 is getting less.

Examples

Let us discuss the main consequences following from the proposed heuristic noise model on examples of sources described in [6, 7]. Start with [6] where :

f_o = 10600 MHz, f_m = 1 kHz, FD ÷ FD_c,
$Q_o = 55\times10^3$, $Q_L = 20\times10^3$, q_{ef} = 1.09,
K_{PD} = 0.5 V/rad, $V_{PD}(f_m)$ = -167 dB(rad²/Hz),
$S_{floor\ (theory)}(f_m)$ = -122 dB/Hz,
$S_{floor\ (exp)}(f_m)$ = [-122 + (6...10)] dB/Hz.

The additional conditions of our simulation are :

$S_{\varphi o\ MA}(f_m)$ = -155 dB/Hz, $S_{\varphi o\ ALA}(f_m)$ = -160 dB/Hz,
L_L = -3 dB, L_{eg} = 3 dB, $L_r + L_{DC2}$ = 5.35 dB, $n_{MA} = n_{opt}$,
$L_{cir\text{-}refl}$ = 0.85 dB, $P_{PD\ (1\ diode)}$ = 1.5 mW, P_{MA} = 80 mW.

In this situation the calculated system phase noise floor 3-dB increased with respect of (23) changes from -121.4 to -105.2 dB/Hz as $S_{c\text{-}fd}$ (1 kHz) varies from -180 to -155 dB/Hz. The average experimental value, $S_{floor}(f_m)$ = -114 dB/Hz takes place with $S_{c\text{-}fd}$ = -164.5 dB/Hz. Without the AFC system the analyzed oscillator must have $£_{in} \approx$ -101 dB/Hz.

Now go to [7] where :

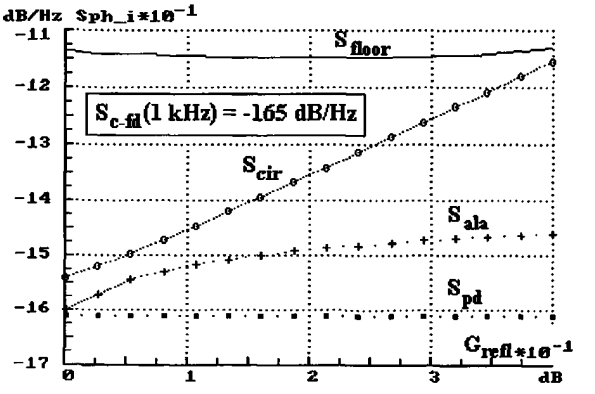
f_o = 10080 MHz, f_m = 100 Hz, FD ÷ FD_a,
$Q_o = 500\times10^3$, $Q_L = 300\times10^3$, q_{ef} = 0.8,
K_{PD} = 0.9 V/rad, $V_{PD}(f_m)$ = -146 dB(rad²/Hz),
$S_{floor\ (th)}(f_m)$ = -103 dB/Hz, $S_{floor\ (exp)} \approx$ -93 dB/Hz.

The additional conditions of our simulation are :

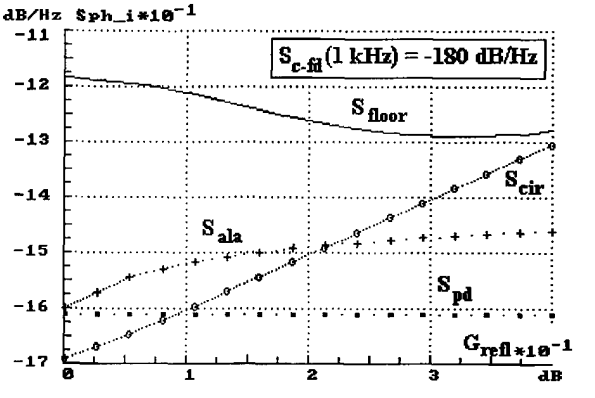
$S_{\varphi o\ MA}(f_m)$ = -145 dB/Hz, $S_{\varphi o\ ALA}(f_m)$ = -150 dB/Hz,
L_L = -3 dB, L_{eg} = 3 dB, $L_r + L_{DC2}$ = 9.6 dB, $n_{MA} = n_{opt}$,
$L_{cir\text{-}refl}$ = 1.2 dB, $P_{PD\ (1\ diode)}$ = 2.5 mW, P_{MA} = 100 mW.

In this situation the theoretical phase noise floor changes from -103 to -93.2 dB/Hz if $S_{c\text{-}fd}$(100 Hz) varies from -170 to -145 dB/Hz. Thus the experimental value $S_{floor}(f_m)$ is consistent here with $S_{c\text{-}fd} \approx$ -145 dB/Hz. It is interesting that with the AFC loop turned off the oscillator would have $£_{in}$ = -93.1 dB/Hz.

Investigate now potential improvements which may arise from adding the ALA to a reflect wave channel in these two cases. Start again with the case 1 assuming $S_{c\text{-}fd}$(1 kHz) = -165 dB/Hz firstly and -180 dB/Hz secondly. The results of computer modelling are shown in Fig.4a, b.



(a)



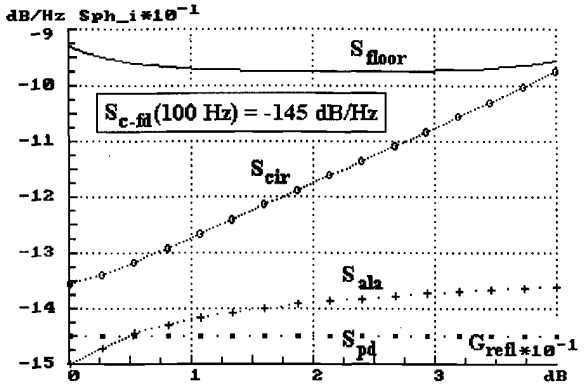
(b)

Figure 4. Influence of the reflected signal channel gain on the system phase noise floor. Modelling conditions in accord with example 1.

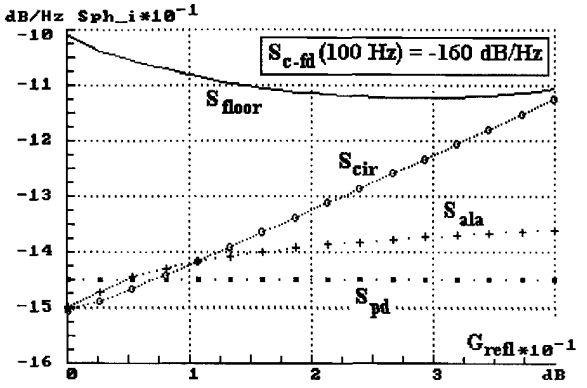
One can see that for the expectable circulator residual phase noise level this technique gives just symbolic result decreasing the system noise floor by 0.4 dB. And even with the 15-dB lower circulator noise the most improvement is only 7.5 dB.

Return to the example 2 and compute the same values assuming successively $S_{c\text{-}fd}$(100 Hz) = -145 and -160 dB/Hz. The results

of modelling are shown in Fig.5a, b. The possible phase noise decreasing is 4.3 and 10.2 dB accordingly.

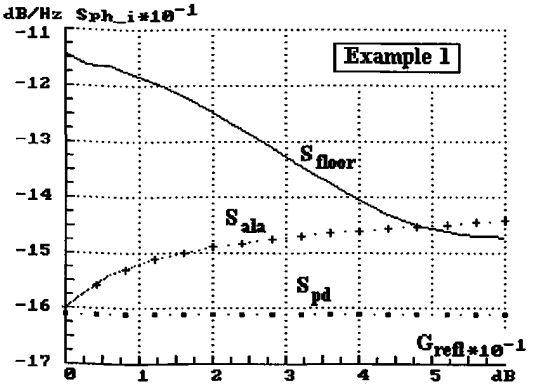


(a)

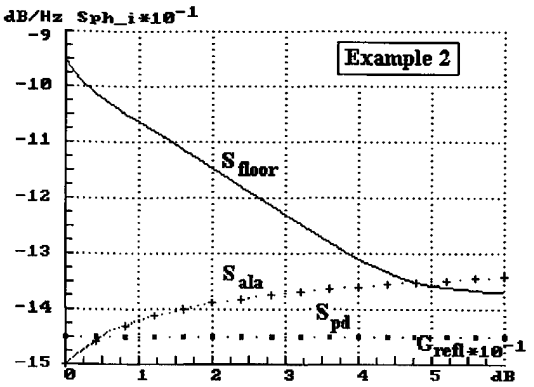


(b)

Figure 5. Influence of the reflected signal channel gain on the system phase noise floor. Modelling conditions in accord with example 2.



(a)



(b)

Figure 6. Influence of the reflected signal channel gain on the phase noise floor of the oscillator with a circulator replaced by a 3-dB hybrid. Modelling conditions in accord with: (a) - example 1, (b) - 2.

The only way to overcome radically the circulator self flicker noise impact on oscillator parameters is to change it with 3-dB hybrid junction or another suitable noiseless reciprocal circuit. The results of calculations made on that basis (Fig.6) reveal the (30...40)-dB oscillator phase noise improvement in comparison with the simplest case $G_{refl} = 0$.

Conclusion

It is found that use of a ferrite circulator in the frequency discriminator of a low phase noise microwave oscillator with the AFC system comes to unevitable degradation of the noise floor due to impact of the self circulator noise. The only way to overcome this drawback is to replace the circulator as a means of reflected wave extraction with a noiseless hybrid junction.

Then inserting an auxiliary low noise amplifier to gain the reflected signal one can expect to get the phase noise improvement not less than by 30...35 dB in accord with the results of computer modelling.

Thus the optimized X-band bipolar transistor oscillator using the sapphire DDR-based combined stabilization scheme might provide the £(1 kHz) value approaching -150 dB/Hz .

Acknowledgments

The author should like to thank D.M.Sazonov and especially B.A.Mishustin for their valuable help.

References

[1] S.L.Abramov, E.N.Ivanov, and D.P.Tsarapkin, "Low noise microwave oscillator with a temperature stabilized disk dielectric resonator", Radiotechnika, 1988, No.11, pp. 81-83 (in Russian).

[2] 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", in Proc. 43rd Annual Freq. Contr. Symp., 1989, pp.89-93.

[3] 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., vol. UFFC-37, pp. 339-346, September 1990.

[4] G.M.Utkin, M.V.Kapranov, L.A.Belov, et al., An automatic phase correction in amplifiers. Moscow: Soviet Radio, 1972 (in Russian).

[5] M. J. Bianchini, J. B. Cole, R. DiBiase, et al. "A single-resonator GaAs FET oscillator with noise degeneration", in 1984 IEEE MTT-S Digest, 1984, pp.270-273.

[6] F. L. Walls, C. M. Felton, and T. D. Martin, "High spectral purity X-band source", in Proc. 44th Annual Freq. Contr. Symp., 1990, pp.542-548.

[7] M. M. Driscoll and R. W. Weinert, "Low noise, microwave signal generation using cryogenic, sapphire dielectric resonators: An update", in Proc. 46th Annual Freq. Contr. Symp., 1992, pp. 157-162.

[8] D.G.Santiago and G.J.Dick, "Closed loop tests of the NASA sapphire phase stabilizer", in Proc. 47th Annual Freq. Contr. Symp., 1993, pp. 774-778.

[9] A. M. Darwish, A. K. Ezzeddine, H-L. A. Hung, and F. R. Phelleps, "A new phase reduction technique for MMIC oscillators", in 1984 IEEE MTT-S Digest, 1984, pp.463-466.

[10] G. J. Dick and D. G. Santiago, "Microwave frequency discriminator with a cryogenic sapphire resonator for ultra-low phase noise", in Proc. of 6th European Frequency and Time Forum, 1992.

[11] A.A.Borodulin and D.P.Tsarapkin, "Regime Impact on KT647A-2 Transistor Amplifier Phase Noise", Electronic Technique, Ser. "Microwave electronics", No 8(432), pp. 23-25, 1990 (in Russian).

[12] M.E. Tobar, A.J. Giles, S.Edwards, and J.Searls, "High-Q TE stabilized sapphire microwave resonators for low noise applications", in Proc. 47th Annual Freq. Contr. Symp., 1993, pp. 749-756.

[13] D.P.Tsarapkin, "An uncooled microwave oscillator with 1-million effective Q-factor", in Proc. 47th Annual Freq. Contr. Symp., 1993, pp. 779-783.

[14] V.F. Kroupa, "Noise properties of PLL systems", IEEE Trans., vol. COM-30, No 10, pp. 2244-2252, Oct. 1982.

[15] N.N.Kolachevsky, Fluctuation phenomena in ferromagnetic materials. Moscow: Nayka, 1985 (in Russian).

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

HIGH TUNING COEFFICIENT WHISPERING GALLERY MODES IN A SAPPHIRE DIELECTRIC RESONATOR TRANSDUCER

H. PENG and D.G. BLAIR

Department of Physics, University of Western Australia
Nedlands, WA 6009, Australia

Abstract. A sapphire dielectric resonator transducer utilizes high tuning coefficient whispering gallery modes in a double disk resonator. The high tuning coefficients and high Q-factors of the whispering gallery modes allow the transducer to achieve very high sensitivity for displacement measurements. In this report, we present an improved mode matching method which can determine the resonant frequencies of the whispering gallery modes in double cylindrical dielectric disk resonators with high accuracy. Two double sapphire disk resonators with different dimensions have been tested to verify the theory. The comparison of the theoretical and experimental results is described. The frequency tuning behaviour of the whispering gallery modes is investigated.

Introduction

A Sapphire Dielectric Resonator (SDR) transducer is a microwave parametric transducer which utilises the modulation of the resonant frequency of a double disk sapphire resonator due to mechanical motion.[1] The high tuning coefficients and high Q-factors of the whispering gallery modes in the sapphire dielectric resonator allow the transducer to achieve very high sensitivity for displacement measurements. The SDR transducer as shown schematically in Fig.1 can achieve tuning coefficients (df/dx) exceeding 2MHz/μm around 10GHz, where f is frequency and x is the gap spacing between two sapphire elements. The SDR transducer has a high electrical Q-factor, giving a very high value to the product Q(1/f)(df/dx) which determines the sensitivity of the transducer.

At the cryogenic temperature of 4.2K, the Q-factor of sapphire resonators can exceed 10^9.[2] Even at room temperature, the Q-factor is greater than 10^5. Thus the SDR transducer can be used as a sensitive room temperature transducer. Experimental tests of SDR transducers at room temperature with different configurations have shown that the sensitivity of the SDR transducers can approach 10^{-15}m/√Hz.[3][4]

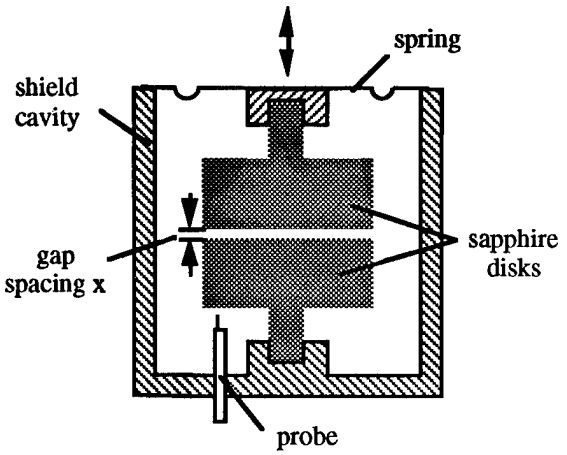


Fig.1 Schematic diagram of a double-disk sapphire dielectric resonator transducer.

The key characteristics of Sapphire Dielectric Resonators (SDR) for transducer applications - the tuning coefficient and the Q-factor - are essentially determined by the resonator geometry. For many practical applications, the tunability of dielectric resonators is a desirable property. In order to maintain the high-Q properties of the resonators, the tunable dielectric resonators employ configurations comprising two or more pieces of the dielectric instead of using

0-7803-1945-1/94 $3.00 © 1994 IEEE

conductors, such as a tuning screw placed at a location of strong fields surrounding the resonator. For transducer applications, the tuning coefficients and Q-factors of the resonators are required to be as large as possible.

Whispering Gallery (WG) modes have the highest Q-factors in sapphire resonators. Hence it is important to investigate the tuning behaviour of the WG modes. Accurate determination of the resonant frequencies from complete field solutions for the tunable dielectric resonators which comprise two or more pieces of dielectrics are very difficult, and there is a need to find simplified approaches. A variety of the tunable dielectric resonators with two or more pieces of dielectrics have been reported previously.[5-8] The methods of calculating the resonant frequency of the tunable resonators can be classified into two types. One is the mode matching method which involves solving the basic electromagnetic field equations with relevant boundary conditions.[6][9] This method can usually provide accurate solutions for some simple and symmetrical resonator configurations. The other is the coupling method which uses equivalent electrical circuits to model the coupled resonators.[7][10-11] The latter provides a simple way to calculate the resonant frequency tuning due to the interaction of two coupled modes. But this coupling calculation method is limited to coupled modes with simple coupling coefficients. (eg., linear or constant coupling coefficients) When the coupling coefficients are neither linear nor constant, the coupling calculation becomes very difficult.

This paper presents an improved mode matching method for determining the resonant frequencies of WG modes in double cylindrical dielectric disk resonators. The method has high accuracy and can also be applied to one-piece cylindrical dielectric resonators. This is an extension of Garault and Guillon's method [12-16] from one piece of isotropic dielectric to double pieces of anisotropic dielectrics. The method is applied to quasi TE and quasi TM WG modes with even or odd axial mode numbers. By taking account of the different wave numbers for TM and TE modes inside the dielectric due to the anisotropy, the method allows the calculation for the TM modes with high axial number to be more accurate than previous work[14]. The theoretical calculation of the resonant frequencies is consistent with experimental results obtained from two double disk resonators with different dimensions, even for high axial mode numbers. This analysis provides a way of understanding the tuning behaviour of the high tuning coefficient WG modes. In addition, the analysis could have potential applications to the design of high-Q dielectric resonators and filters with large frequency tuning range.

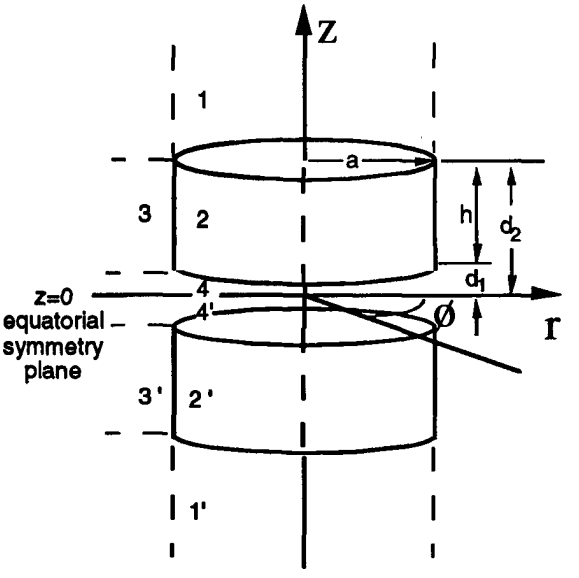


Fig.2 A cylindrical double disk resonator with a defined cylindrical coordinates

Theory

A cylindrical double disk resonator model with a defined cylindrical coordinates is illustrated by Fig.2. To simplify the model, the two dielectric disks are assumed to be identical and are placed in open space. The dielectric is anisotropic and its dielectric c-axis is assumed to be parallel to the z direction. The permittivity of the dielectric parallel the z-axis is defined as ε_z. The permittivity perpendicular to the z-axis defined as ε_r and ε_ϕ are assumed to have the same value, $\varepsilon_r=\varepsilon_\phi$. The resonator is divided into 8 regions which are indicated by 1, 2, 3, 4, 1', 2', 3' and 4'. Owing to its symmetry configuration, only four regions (1, 2, 3

and 4) above the symmetry plane are considered in the analysis.

WG modes have the property that their energy is mainly confined to the region near the boundary of the dielectric. They are hybrid type modes, however at first approximation we can consider that there are two types of WG modes: quasi TM modes and quasi TE modes. Each mode is described by three mode numbers m, n and p describing the number of field variations along the azimuthal, radial and axial directions respectively.

The determination of the resonant frequencies of the WG modes requires solution of Maxwell's equations with appropriate boundary conditions. The propagation characteristics of the electrical and the magnetic fields depend on the configuration of the double disk resonator. Inside the dielectric the fields appears to be periodic, but the fields are evanescent in region 1, 3 and 4 outside the dielectric. Assuming the arguments for the z dependence between z components E_z and H_z in region 2 and 3 are equal, the z components E_z and H_z for the resonator can be written as

$$E_{z1} = A_E J_m(k_E r) \cos(m\phi) \exp(-\alpha_E z) \quad (1a)$$

$$E_{z2} = B_E J_m(k_E r) \cos(m\phi) [A_1 \sin(\beta z) + B_1 \cos(\beta z)], \quad (1b)$$

$$E_{z3} = C_E K_m(k_3 r) \cos(m\phi) [A_1 \sin(\beta z) + B_1 \cos(\beta z)], \quad (1c)$$

$$E_{z4} = D_E J_m(k_E r) \cos(m\phi) \begin{Bmatrix} \sinh(\alpha_E z) \\ \cosh(\alpha_E z) \end{Bmatrix}, \quad (1d)$$

$$H_{z1} = A_H J_m(k_H r) \sin(m\phi) \exp(-\alpha_H z), \quad (1e)$$

$$H_{z2} = B_H J_m(k_H r) \sin(m\phi) [A_2 \sin(\beta z) + B_2 \cos(\beta z)], \quad (1f)$$

$$H_{z3} = C_H K_m(k_3 r) \sin(m\phi) [A_2 \sin(\beta z) + B_2 \cos(\beta z)], \quad (1g)$$

$$H_{z4} = D_H J_m(k_H r) \sin(m\phi) \begin{Bmatrix} \cosh(\alpha_H z) \\ \sinh(\alpha_H z) \end{Bmatrix} \quad (1h)$$

where

$$k_E^2 = \varepsilon_z k_0^2 - \beta^2 \varepsilon_z/\varepsilon_r \quad (2a)$$

$$k_H^2 = \varepsilon_r k_0^2 - \beta^2, \quad (2b)$$

$$k_3^2 = \beta^2 - k_0^2, \quad (2c)$$

$$k_E^2 = k_0^2 + \alpha_E^2, \quad (2d)$$

$$k_H^2 = k_0^2 + \alpha_H^2. \quad (2e)$$

Here m is azimuthal mode number, β the axial propagation constant, k_0 the free space wave number which has $k_0^2 = \omega^2\mu_0\varepsilon_0$, k_3 the radial propagation constant outside the dielectric, k_E the radial propagation constant inside the dielectric for the electrical field, k_H the radial propagation constant inside the dielectric for the magnetic field, α_E and α_H the axial decay constants outside the dielectric for the electrical field and the magnetic field respectively.

The other transverse components can be obtained from the components E_z and H_z using following equations:

$$H_\phi = \frac{1}{k_i^2}\left(-j\omega\varepsilon_0\varepsilon_i \frac{\partial E_z}{\partial r} + \frac{1}{r}\frac{\partial^2 H_z}{\partial z \partial \phi}\right) \quad (3a)$$

$$E_\phi = \frac{1}{k_i^2}\left(j\omega\mu \frac{\partial H_z}{\partial r} + \frac{1}{r}\frac{\partial^2 E_z}{\partial z \partial \phi}\right) \quad (3b)$$

$$H_r = \frac{1}{k_i^2}\left(j\omega\varepsilon_0\varepsilon_i \frac{1}{r}\frac{\partial E_z}{\partial \phi} + \frac{\partial^2 H_z}{\partial z \partial r}\right) \quad (3c)$$

$$E_r = \frac{1}{k_i^2}\left(-j\omega\mu \frac{1}{r}\frac{\partial H_z}{\partial \phi} + \frac{\partial^2 E_z}{\partial z \partial r}\right) \quad (3d)$$

where $k_i^2 = k_H^2$ in region 1, 2 and 4; $k_i^2 = -k_3^2$ in region 3; $\varepsilon_i = \varepsilon_r$ in region 2 and $\varepsilon_i = 1$ in region 1, 3 and 4.

For radial match, the tangential components of the E and H fields at r=a between region 2 and region 3 must

be equal. This leads to a transcendental equation as follows:

$$\left(\frac{K_m'(\xi)}{\xi K_m(\xi)}+\frac{\varepsilon_r \lambda_E J_m'(\lambda_E)}{\lambda_H^2 J_m(\lambda_E)}\right)\left(\frac{K_m'(\xi)}{\xi K_m(\xi)}+\frac{J_m'(\lambda_H)}{\lambda_H J_m(\lambda_H)}\right)$$

$$=\frac{m^2(\xi^2+\lambda_H^2)(\varepsilon_r\xi^2+\lambda_H^2)}{\xi^4\lambda_H^4} \qquad (4)$$

where $\xi=k_3 a$, $\lambda_H=k_H a$ and $\lambda_E=k_E a$. When $k_3^2<0$, ξ turns to be pure imaginary, the Bessel function K_m in equation (4) is replaced with a Hankel function of the second kind $H_m^{(2)}$.

For axial match, the transverse electric (TE) modes and the transverse magnetic (TM) modes in region 1, 2 and 4 are considered. The components E_z and H_z are assumed zero for quasi TE and quasi TM modes respectively. From satisfying the axial boundary conditions that the transverse electric and magnetic fields be continuous at the planer interfaces $z=d_1$ and $z=d_2$, four transcendental equations are obtained:

TE_{mnp}, p even;

$$(-\alpha_H/\beta + \tan(\beta d_2))\,(1 - \alpha_H/\beta \tanh(\alpha_H d_1)\tan(\beta d_1)) = (1 + \alpha_H/\beta \tan(\beta d_2))\,(\alpha_H/\beta \tanh(\alpha_H d_1) + \tan(\beta d_1)), \qquad (5a)$$

p odd;

$$(-\alpha_H/\beta + \tan(\beta d_2))\,(1 - \alpha_H/\beta \coth(\alpha_H d_1)\tan(\beta d_1)) = (1 + \alpha_H/\beta \tan(\beta d_2))\,(\alpha_H/\beta \coth(\alpha_H d_1) + \tan(\beta d_1)), \qquad (5b)$$

TM_{mnp}, p even;

$$(-\varepsilon_r\alpha_E/\beta + \tan(\beta d_2))\,(1 - \varepsilon_r\alpha_E/\beta \tanh(\alpha_E d_1)\tan(\beta d_1)) = (1 + \varepsilon_r\alpha_E/\beta \tan(\beta d_2))(\varepsilon_r\alpha_E/\beta \tanh(\alpha_E d_1) + \tan(\beta d_1)), \qquad (5c)$$

p odd;

$$(-\varepsilon_r\alpha_E/\beta + \tan(\beta d_2))\,(1 - \varepsilon_r\alpha_E/\beta \coth(\alpha_E d_1)\tan(\beta d_1)) = (1 + \varepsilon_r\alpha_E/\beta \tan(\beta d_2))(\varepsilon_r\alpha_E/\beta \coth(\alpha_E d_1) + \tan(\beta d_1)), \qquad (5d)$$

where m, n and p are integers. The above six transcendental equations (4-5) can be numerically solved using a computer. The resonant frequencies of WG modes for different double disk resonators, such as two disks with variable gap spacing, can be theoretically obtained. Hence the frequency tuning due to the change of gap spacing can be obtained from the mode matching method. The analysis for the double dielectric disk resonator extends to an ordinary one-piece cylindrical resonator when the gap spacing x=0.

Experimental and theoretical results

Two different double disk sapphire resonators were tested in open space at room temperature to verify the theory. Resonator A has a radius $a=a_1=a_2=15.81$mm and height $h_1=14.42$mm and $h_2=14.44$ for the two disks respectively. Resonator B has a radius $a=a_1=a_2=10.88$mm, and height $h_1=10.58$mm and $h_2=10.69$mm. The average height values of the disks $h=(h_1+h_2)/2$ are used to enable calculation in the symmetric configuration. The sapphire dielectric constants used for the calculations for the two resonators are the same as given in paper [17], $\varepsilon_z=11.589$ and $\varepsilon_r=\varepsilon_\phi=9.395$.

The WG modes of the double disk resonators were experimentally investigated when the gap spacing between two disks varied. Fig. 3 shows the experimental and theoretical results of the TM_{710} mode in resonator A as the gap spacing varies. The measurements are in good agreement with theoretical values. As the gap spacing increases, the difference between theoretical and experimental values increases. But the difference is still less than 0.4% at x=2mm. Figure 3(a) represents a typical resonant frequency tuning curve for the high tuning coefficient WG modes. Figure 3(b) shows that the Q-factor of mode TM_{710} in resonator A is nearly independent of the gap spacing variations. The degradation of the Q-factor at the gap spacing of ~0.1mm is due to the interaction of the TM_{710} mode with the TE_{511} mode.

Table 1 and Table 2 show more theoretical and experimental results of WG modes in resonator A and

Table 1:
Resonant frequencies of the WG modes in resonator A (a=15.8mm, h=14.4mm)

gap spacing	0		0.1 mm		1 mm	
modes	theory	experi	theory	experi	theory	experi
TM_{710}	8.910	8.933	9.101	9.110	9.312	9.341
TM_{711}	9.319	9.370	9.319	9.372	9.318	9.377
TM_{712}	10.076	10.142	10.314		10.838	10.799
TM_{713}	10.969	11.059	10.969	11.071	10.961	11.156
TE_{710}	10.856	10.828	10.895	10.848	11.050	10.977
TE_{711}	11.235	11.207	11.235	11.230	11.210	11.290

Table 2:
Resonant frequencies of the WG modes in resonator B (a=10.9mm, h=10.6mm)

gap spacing	0		0.1 mm		1 mm	
modes	theory	experi	theory	experi	theory	experi
TM_{510}	9.905	9.949	10.210	10.237	10.571	10.625
TM_{511}	10.657	10.682	10.657	10.685	10.653	10.696
TM_{610}	11.420	11.456	11.736	11.758	12.023	12.066
TM_{611}	12.081	12.095	12.081	12.099	12.080	12.111

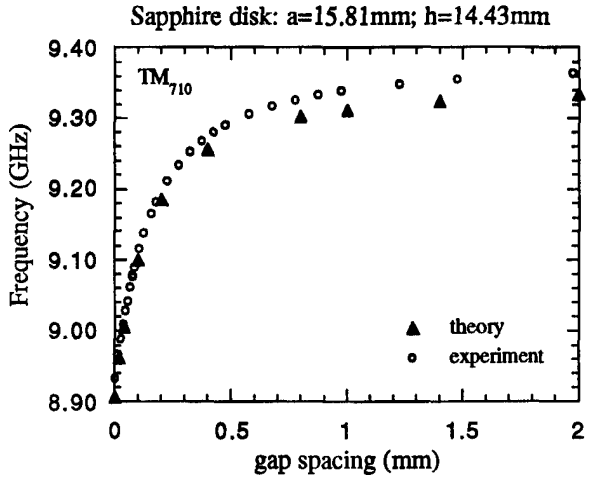
resonator B at three different gap spacings. Experimental results and theoretical calculation show that the gap spacing variations has a strong influences on some TM_{mnp} modes whose axial number p is zero or an even integer. As the gap spacing increases, the TM modes with zero or even p have a transition towards TM modes with axial numbers p+1 which are also compatible with the boundary conditions for one-piece disk resonator. This is why the TM modes with odd axial numbers have less dependence on the gap spacing. Compared with the high tuning coefficient TM modes, the resonant frequencies of TE modes show less dependence on the variations of the gap spacing. Like the TM modes, the discrepancy between experimental and theoretical results for TE modes increases with the increase of the gap spacing. In reality, both the TM and TE modes are hybrid modes and a more sophisticated theory is required for double resonators with a large gap spacing.

When the double resonator becomes a single resonator (x=0), experiment and calculation agree to high accuracy, a few tenths of a percent as shown in Table 1 and 2, even for the TM modes with high axial numbers ($p \geq 2$). The errors between theory and measurements for the single resonators is probably due to uncertainties in permittivity and dimensions. The TM and TE modes have different wave numbers inside the dielectric due to the anisotropy, which explains the discrepancy between theoretical and experimental results for TM modes with high axial numbers in paper.[14]

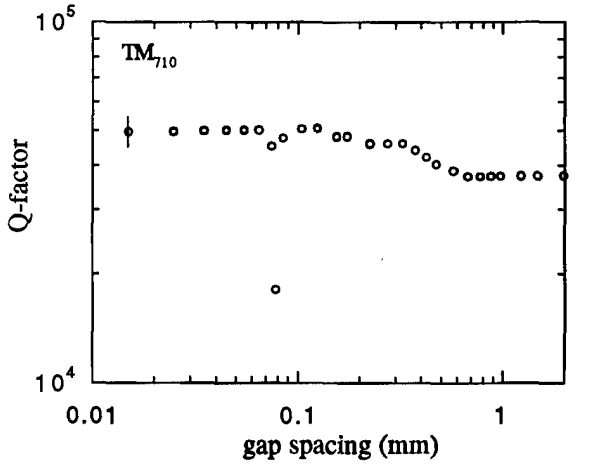
Conclusion

A mode matching method for determining the resonant frequencies of the WG modes in double disk resonators has been shown to give frequencies accurate to a few tenths of a percent. The theory gives improved accuracy for calculation of the TM modes with high

axial mode number (p≥2) in single resonators. Experiments and theory agree that the high tuning coefficient WG modes in the double disk resonators are TM modes with a zero or even axial mode number.



(a)



(b)

Fig.3 (a) Theoretical and experimental resonant frequencies of mode TM_{710} as a function of the gap spacing x between the two disks. (b) Experimental Q-factor of mode TM_{710} versus the gap spacing.

Acknowledgments

The authors wish to thank D. Newman and A. Gorham for their assistance on machining sapphire disks. This research work was supported by the Australia Research Council and the Overseas Postgraduate Research Scholarships .

References

[1] D.G. Blair, E.N. Ivanov and H. Peng, "Sapphire Dielectric Resonator Transducers", *J. Phys. D: Appl. Phys.* 25, pp1110-15, 1992.

[2] A. Luiten, A.G. Mann and D.G.Blair, "Ultrahigh Q-factor Cryogenic Sapphire Resonator", *Electronics Letters,* 29, p879-81, 1993.

[3] H. Peng, D.G. Blair and E.N. Ivanov, "Sapphire Dielectric Resonator Transducer For Highly Sensitive Measurements", *Proceedings of International Conference on Fluid Mechanics and Theoretical Physics,* pp625-29, Peking University Press, 1993.

[4] H. Peng, D.G. Blair and E.N. Ivanov, "An Ultra High Sensitivity Sapphire Transducer for Vibration Measurements", To be published in *J. of Phys. D: Applied Phys.* 27, 1994.

[5] A. Karp, H. J. Shaw and D. K.Winslow, "Circuit properties of microwave dielectric resonators" *IEEE Trans. Microwave Theory Tech.*, vol. MTT-16, pp818-28, 1968

[6] S. Fiedziuszko and A. Jelenski, "Double Dielectric Resonator", *IEEE Trans. Microwave Theory Tech.*, vol. MTT-19, pp779-80, 1971

[7] K. A. Zaki and C. Chen, "Coupling of non-axially symmetric hybrid modes in dielectric resonators", *IEEE Trans. Microwave Theory Tech.*, vol. MTT-35, pp1136-42, 1987

[8] D.G. Blair and A. Sanson, "High Q tunable sapphire loaded cavity resonator for cryogenic

operation," *Cryogenics,* vol. 29, pp1045-49, 1989

[9] S. W. Chen and K. Zaki, "A novel coupling method for dual-mode dielectric resonators and waveguide filters", *IEEE Trans. Microwave Theory Tech.*, vol. MTT-38, pp1885-92, 1990

[10] M.E. Tobar and D.G. Blair, "A generalised equivalent circuit applied to a tunable superconducting resonator" *IEEE Trans. Microwave Theory Tech.*, vol. MTT-39, pp1582-94, 1991.

[11] M.E. Tobar, "Effects of spurious modes in resonant cavities," *J. phys. D: Appl Phys.*, 26, pp2022-6, 1993.

[12] Y. Garault and P. Guillon, "Higher accuracy for the resonance frequencies of dielectric resonators," *Electron. Lett.*, vol. 12, no. 18, pp 475-6, 1976

[13] D. Kajfez and P. Guillon, Editors, *Dielectric Resonators*, Artech House, Inc., p65, 1986.

[14] M.E. Tobar and A.G. Mann, "Resonant frequencies of higher order modes in cylindrical anisotropic dielectric resonators," *IEEE. Trans. MTT.*, vol. 39, no. 12, pp2077-81, 1991.

[15] E.N. Ivanov, D.G. Blair and V.I. Kalinichev, "Approximate approach to the design of shielded dielectric disk resonators with whispering gallery modes", *IEEE. Trans. MTT.*, vol. 41, no. 4, pp632-8, 1993.

[16] Y. Kobayashi and T. Senju, " Resonant modes in shielded uniaxial-anisotropic dielectric rod resonators", *IEEE. Trans. MTT.*, vol. 41, no. 12, pp2198-205, 1993.

[17] R. Shelby and J. Fontanella, " The low temperature electrical properties of some anisotropic crystals", J. Phys. Chem. Solids, vol. 41, pp 69-74, 1980

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Ceramic Dielectric Resonator Oscillator Aging

Muhammad Mizan

Army Research Laboratory, E&PS Directorate, AMSRL-EP-MA, Fort Monmouth, NJ 07703

ABSTRACT

Two Ceramic Dielectric Resonator Oscillators (CDROs), one operating at center frequency of 1.5 GHz and the other one operating at 8.3 GHz is being investigated for its frequency aging characteristic. The oscillators were ovenized in a Sigma temperature chamber with aging temperature set at +65°C. Datas are taken with an automated test system. The frequency aging for the 1.5 GHz CDRO is found to be ±0.5 ppm/3 months and 40 ppb/hour. Aging is better than ±3 ppm/two month for the X-band CDRO. The oscillators were also evaluated for its single side band phase noise characteristic. The measured phase noise levels for the L-band oscillator were -100 dBc/Hz and -130 dBc/Hz at 100 Hz and 1 kHz carrier offset frequencies, respectively. The X-band CDRO's phase noise levels were -66 dBc/Hz and -93 dBc/Hz, at 100 Hz and 1 kHz carrier offset frequencies, respectively.

BACKGROUND

Although the L- and X-band ceramic dielectric resonator oscillators were investigated for its long term stability, in reality the oscillators were designed for low phase noise, for application in high dynamic range radar, and digital communication systems. In the process of evaluating the short term stability it was observed that frequency fluctuations exhibited by the CDROs were exceptionally good. So it was decided to investigate the aging characteristic for these CDROs even though it wasn't designed for such application.

The oscillators were designed and fabricated using a Murata Erie's dielectric resonator in a patented dielectric resonator oscillator structure which provided a loaded Q of about 25,000 at L-band and 14,000 at X-band for $TE_{01\delta}$ mode of operation. The L-band oscillator uses a Bipolar Junction Transistor(BJT) amplifier and the X-band oscillator utilizes medium power Field Effect Transistor (FET) amplifier. The amplifiers were tested for its residual phase noise to ensure acceptable 1/f noise levels. Positive feedback loop oscillator configuration is used in which the dielectric resonator is employed as the frequency determining element. The design is very simple and costs effective. The design consists of a loop amplifier, power divider, and a resonator enclosed in a metal cavity. Amplitude limiting is achieved by the non-linearities in the loop amplifier. Detailed information on how to design ultra stable frequency sources can be found in reference [1-4].

The oscillators were evaluated for phase noise, frequency vs temperature behavior, and frequency aging characteristics. For the aging measurement, the frequency sampling interval was set to 10 minute, initially, subsequently changed to 30 minute interval few days after the experiment started.

MEASURED RESULTS

The CDROs were first evaluated for its phase noise performance. The measurement was performed on a HP3047A noise measurement system. A high overtone bulk acoustic resonator based oscillator that was capable of generating frequencies from 635 MHz to 10.320 GHz was used for downconverting both L and X-band CDROs to an intermediate frequency below 160 MHz which was then measured against HP8662A frequency synthesizer. The use of this measurement technique limited the absolute phase noise measurement to less than 1 kHz carrier offset frequencies for 1.5 GHz CDRO and to about 10 kHz carrier offset frequencies for the 8.3 GHz CDRO due to

U.S. Government work not protected by U.S. copyright

synthesizer phase noise. The measured phase noise for both the oscillators are shown in Figure 1 and 2. The L-band CDRO exhibited phase noise levels of -100 dBc/Hz at 100 Hz from the carrier. At 1 kHz carrier offset frequency the phase noise level should have been -130 dBc/Hz, but unfortunately it was masked by the synthesizer noise. The 8.3 GHz CDRO's phase noise levels were -66 dBc/Hz and -93 dBc/Hz at 100 Hz and 1 kHz from the carrier, respectively.

The oscillators were also evaluated for its center frequency variation as a function of temperature. A programmable Tenny Jr. temperature chamber was used to perform this operation. The output frequencies were recorded while the temperature in the chamber was varied from -60°C to +80°C for L-band oscillator and from -50°C to +60°C for X-band oscillators. The L-band oscillator exhibited parabolic temperature behavior with zero temperature coefficient occurring at +20°C. This is illustrated in Figure 3. The maximum frequency drift was about 40 parts per million over the stated wide temperature range. The frequency vs temperature plot for X-band CDRO is shown in Figure 4. The worst frequency variation is about 65 parts per million (ppm). The frequency-temperature behavior of the L-band CDRO, shown in Figure 3, is significantly different from X-band CDRO, shown in Figure 4. This is mainly because the resonator is constructed of different materials.

Finally, the oscillators are being tested for the frequency aging characteristic. Test setup is shown in Figure 5. The oscillators were ovenized in a Sigma temperature chamber and stabilized for about 4 hour at the aging temperature of 65°C before taking data. At this temperature the frequency vs temperature curve for the 8.3 GHz CDRO is reasonably flat, but not so for the L-band CDRO, it has a slop of about 0.4 ppm/°C (60°C-80°C). Frequency aging for 1.5 GHz CDRO is shown in Figure 6. Since we did not monitored the oven temperature we can not say for sure that the frequency drift is due to the variation in oven temperature or the oscillator itself. We are currently investigating the actual cause of the reported frequency variation. In any case the total frequency variation over a period of three months is less than ±0.5 ppm. The aging performance for the 8.3 GHz oscillator is shown in Figure 7. Frequency varied from 8.323925 GHz to 8.323950 GHz. The frequency deviation is about 25 kHz. These results indicate that ceramic dielectric resonator oscillators are highly stable frequency sources and are suitable for commercial and military systems use.

CONCLUSION

Aging characteristic of Ceramic dielectric resonator based oscillators are reported in this paper. The CDROs operating at 1.5 GHz and 8.3 GHz exhibited excellent long term as well as short term stability. The long term stability could be improved by using proper resonator and innovative circuit design technique.

REFERENCES

[1] M. Mizan, "An X-band, 2.5 Watt Continuous Wave Dielectric Resonator Oscillator for Future Military Systems", Proceeding of the 1992 IEEE Frequency Control Symposium, pp. 409-414, Catalog No.92 CH3083-3

[2] D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum", Proceedings of the IEEE, Vol. 54, No. 2, pp. 329-330, February 1966

[3] M. Mizan, T. Higgins, and D. Sturzebecher, "Ultra-Stable, Low Phase Noise Dielectric Resonator Stabilized Oscillators for Military and Commercial Systems", NASA Conference Publication 3189, Volume 2, Technology 2002 , pp. 289-295, Baltimore, MD, December 1-3, 1992.

[4] M. J. , T. E. Parker and G. K.Montress, "Frequency Stability of L-Band, Two Port Dielectric Resonator Oscillator," IEEE Transactions on MTT, Vol. MTT35, No. 12, Dec. 1987, pp. 1334-1339.

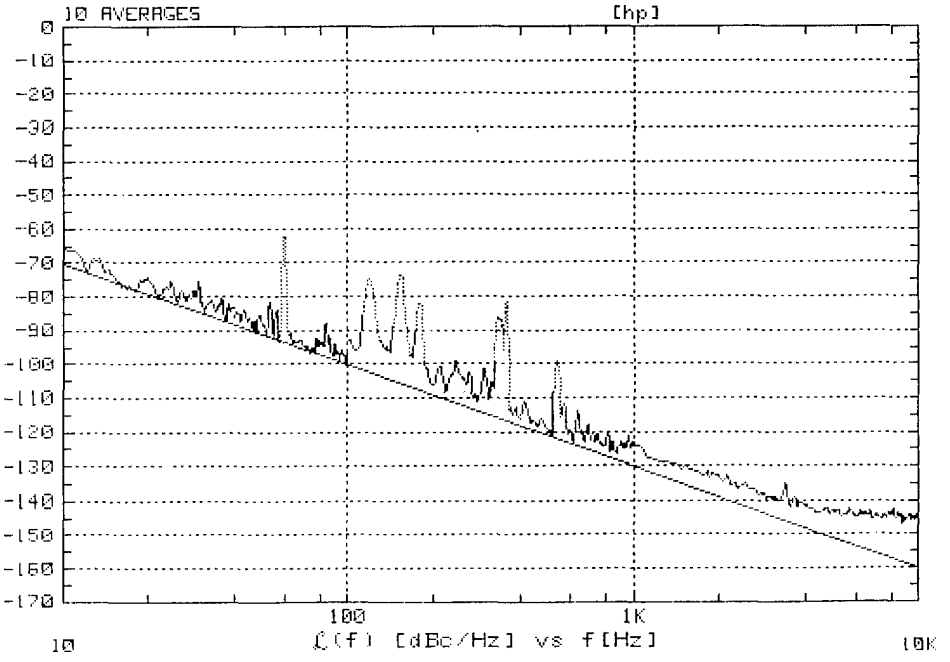


Figure 1. Phase Noise of 1.5 GHz CDRO

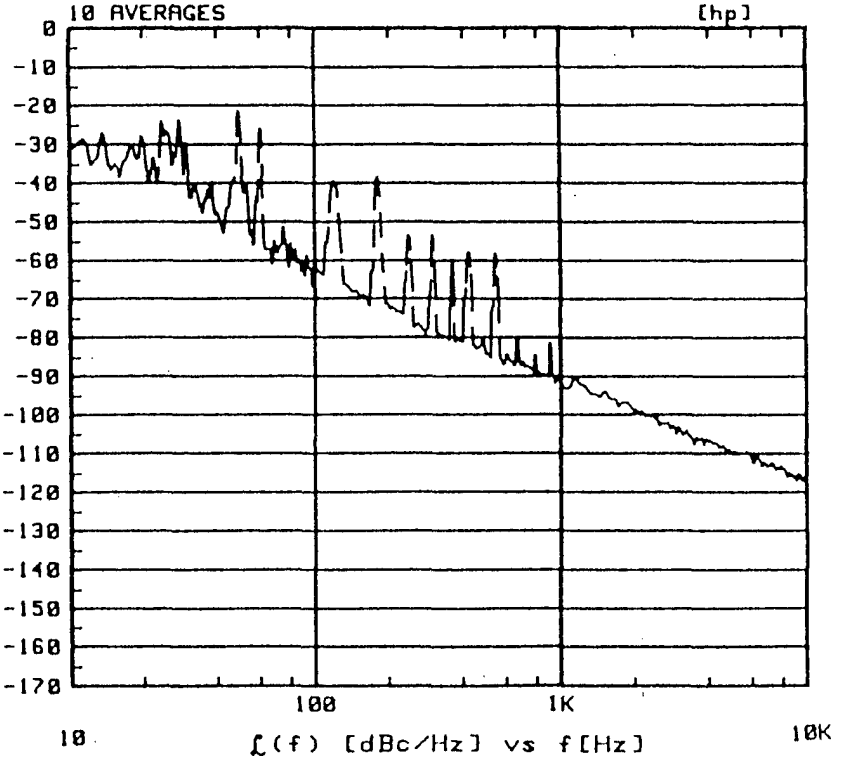


Figure 2. Phase Noise of 8.3 GHz CDRO

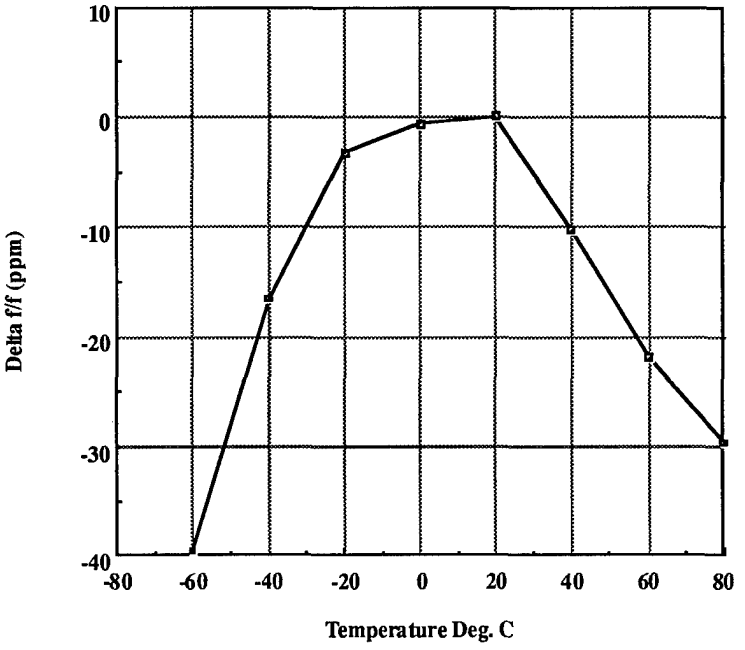


Figure 3. Frequency vs Temperature Behavior of L-band CDRO

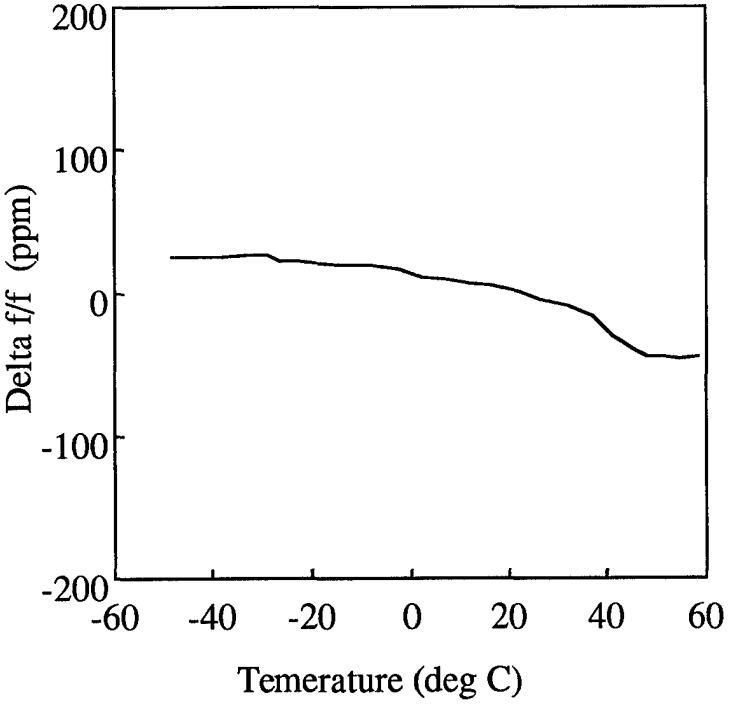


Figure 4. Frequency vs Temperature Characteristic, X-band CDRO

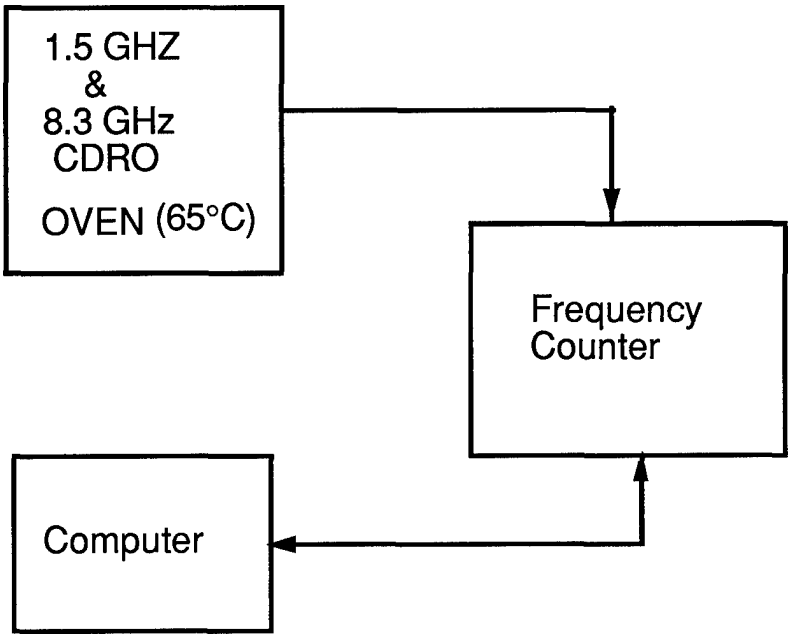


Figure 5. CDRO Aging Test Setup

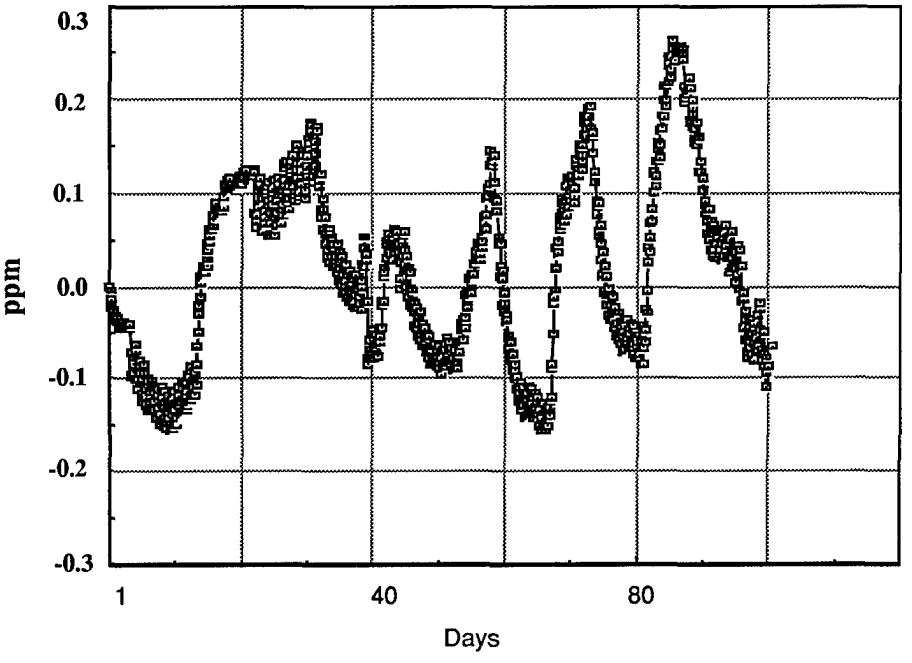


Figure 6. Long term stability of 1.5 GHz CDRO

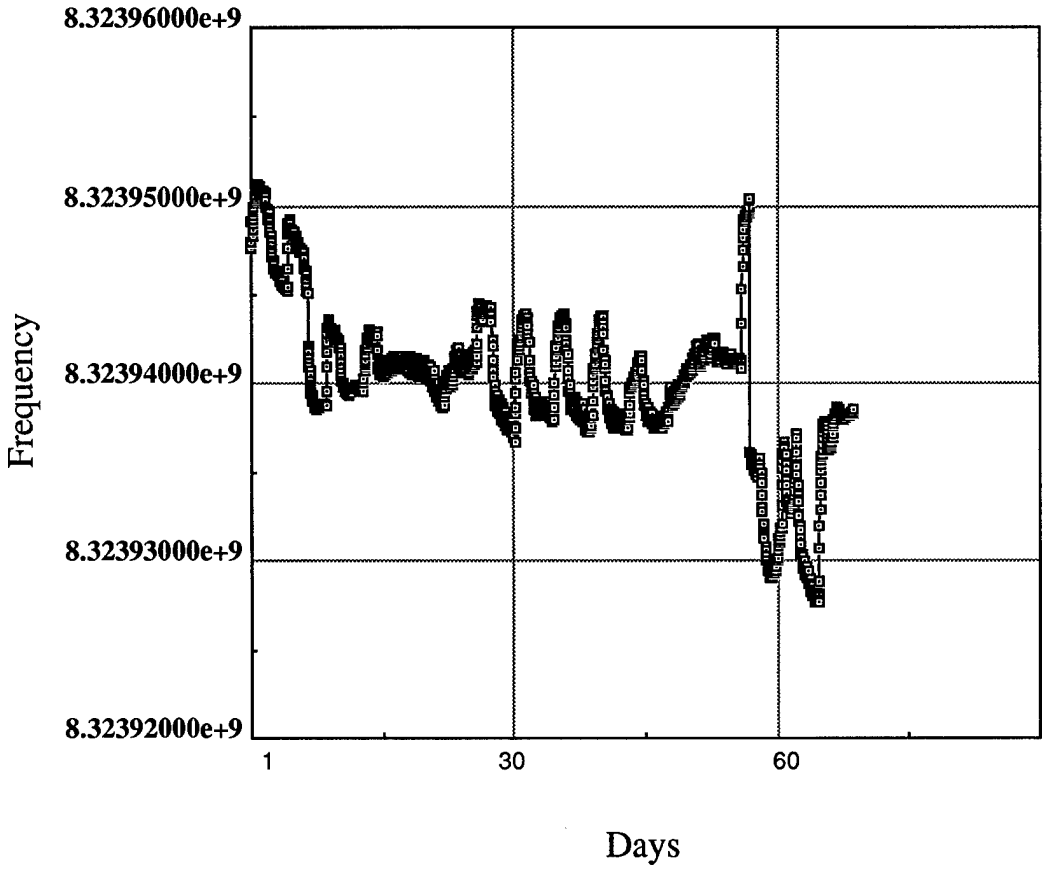


Figure 7. X-band CDRO Aging Performance

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Development of a Low Noise L-Band Dielectric Resonator Oscillator

William J. Tanski
United Technologies Research Center, MS-31
411 Silver Lane
East Hartford, CT 06108

Abstract

This paper describes the design and performance of a dielectric resonator oscillator near 1.3 GHz. Unique features of this oscillator are the very low single side-band phase noise levels achieved, the effectiveness of the vibration isolation system implemented and the parabolic frequency-temperature characteristic attained. A phase noise level of -132 dBc/Hz at 1 KHz offset and a noise floor of -175 dBc/Hz were measured under both quiescent and vibrating conditions. Phase noise data are presented under sinusoidal and random vibration conditions and a vibration sensitivity of $3*10^{-10}/g$ was measured at 25 Hz vibration frequency. The temperature sensitivity was 0.4 ppm/^{0}C at 60^0 C, about 7^0 from the measured turnaround temperature of 67^0 C.

Introduction

A very low phase noise oscillator was required for the development of a low cost radar exciter. The required phase noise levels are (offset frequency and level in dBc/Hz): 10 Hz, -72; 100 Hz, -102; 1 KHz, -132; 10 KHz, -153; 100 KHz, -166; and a noise floor of -173 dBc/Hz. An unsuccessful search was made to determine if oscillators could be purchased meeting these specifications. The only devices which came close were low 1/f phase noise SAW oscillators developed by Parker & Montress [1]. The 500 MHz Raytheon devices achieved about -140 dBc/Hz at 1 KHz offset and would just meet the -132 dBc/Hz required after multiplication. SAW oscillators with this level of performance were considered high risk for this application since these oscillators had never been manufactured in quantity or at low cost.

The DRO was selected for this application because prior work [2] indicated that the noise performance and cost goals could probably be achieved and the space, weight, thermal and vibration requirements were compatible with the use of a DRO.

Design

The oscillator schematic is shown in Fig. 1. All the components, except for the dielectric resonator, are designed or specified to have 50 input and output impedances. In order to perform phase noise measurements, two DRO's were used and one had a voltage variable phase-shifter added to the loop. The phase shifter was found to contribute no measurable phase noise to the oscillator.

The dielectric resonator (or puck as it is commonly called) was selected to be Type 8515 ($E_r = 36$) low loss ceramic manufactured by Trans-Tech. The puck was in the shape of a cylinder sized according to the manufacturers formula to be about 1.64" in diameter and 0.66" high. The cavity was initially designed to be twice the diameter of the puck and three times as high. However, measurements indicated that the maximum (unloaded) Q attainable was only about 70% of the material limit (about 30,000) at 1.3 GHz. The second cavity design was three times the puck diameter and five times as high and this size yielded unloaded Q values in the range of 90% of the material limit. Both brass and aluminum cavities were fabricated yielding similar Q values, and the cavities did not have any interior coating. Two iterations on the size of the puck were required to produce a resonance frequency within 100 KHz of the frequency desired. Further trimming of the resonance frequency was performed by slight modifications of the amount of dielectric material (adding small chips or removing a small amount of material from the puck) and by the use of a small tuning stub. For testing purposes the frequency could be changed by several MHz by the use of a large (2" diameter) tuning stub).

The puck was centered in the cavity on a fused quartz cylindrical stand-off 0.75" in diameter.

0-7803-1945-1/94 $3.00 © 1994 IEEE

Other stand-off materials were tested but the cavity losses increased as indicated by decreasing Q values. Electrical coupling was performed using .060" diameter tube probes, 1.5" long, positioned near the bottom of the cavity and oriented circumferentially. The level of coupling was adjusted by varying the length and position of the probes. Once the desired two port insertion loss (7 dB) and loaded Q (15,000) were obtained, the probes were epoxied in place on quartz stand-offs.

The cavity resonance selected for use is the $TE01_d$ mode and oscillation on this mode was ensured by the addition of a low pass filter to the loop.

The amplifier selected for use was the HP/Avantek UTO-2026 which is a two stage, silicon bipolar transistor, design. This amplifier has a power output level (1 dB compression point) of +20.5 dBm allowing the oscillator to yield the desired output power of +16 dBm. The amplifier gain is 15 dB which is sufficient to yield 3 dB excess gain, open loop, compensating for the 7 dB cavity loss, the 3 dB coupler loss, plus an additional 2 dB in filter, connector, and cable losses. Each amplifier was tested to determine the level of 1/f phase noise and it was found that only about 1/3 of the amplifiers available met the required -135 dBc/Hz level at 1 Hz offset frequency. The amplifier 1 Hz noise levels varied from -135 to -125 dBc/Hz. The amplifier noise figure of 6.5 dB sets the noise floor for the oscillator but the noise figure does not correlate with the 1/f noise.

Equation 1, developed by Parker & Montress [2] to calculate the spectral density of phase noise in surface acoustic wave oscillators, was used to guide the design of the resonator and to aid in specifying the amplifier performance of this DRO.

$$(1)\quad S_f(f) = [a_R F_0^4 + a_E(F_0/(2Q_L))^2]/f^3 + [(2GF_0KT/P_0)(F_0/(2Q_L))^2]/f^2 + (2\, a_R\, Q_L F_0^3)/f^2 + a_E/f + 2GF_gKT/P_0$$

In Eq. 1, f is the frequency offset from the carrier, F_0 is the carrier frequency, Q_L is the resonator loaded Q value, a_R is the resonator 1/f noise constant, a_E is the amplifier 1/f noise constant, G is the amplifier gain, Fg the amplifier noise figure, P_0 the output power level, K is Boltzmann's constant, and T the absolute temperature. It is found experimentally that the dielectric resonator does not contribute any measurable 1/f phase noise to the circuit and the constant, a_R, in Eq. 1 may thus set to zero. For the DRO then, Eq. 1 reduces to

$$(2)\quad S_f(f) = [a_E(F/(2Q_l))^2]/f^3 + [(2GF_gKT/P_0)(F/(2Q_l))^2]/f^2 + a_E/f + 2GF_gKT/P_0$$

Equation 2 permits calculation of the required resonator loaded Q_L, found to be 15,000, for a specified amplifier performance as given by a_E, G, Fg, P_0 and oscillator spectral density level. All the amplifiers met the gain and noise figure requirement to establish the noise floor (the last term in Eq. 2). The other elements in the oscillator loop (coupler, voltage-variable phase shifter when used and low pass filter) were found to contribute no measurable phase noise.

Quiescent Phase Noise Performance

The phase noise was measured using an HP 3047A test set and two DRO's, since no other available oscillators had comparably low noise levels. One of the DRO's had a voltage variable phase shifter [3] added to the loop in order to perform the frequency adjustments required. Representative quiescent 1/f phase noise levels are shown in Fig. 2. Table 1 below lists the required, predicted (from Eq. 2) and measured phase noise levels:

Table 1
DRO Single Side-band Phase Noise (dBc/Hz)

Offset Frequency	Required	Calculated	Measured
10 Hz	-72	-72.4	-72
100 Hz	-102	-102.3	-103
1000 Hz	-132	-131.8	-132
10 KHz	-153	-158	-159
100 KHz	-166	- 172	-173
Floor	-173	-173	-175

Phase Noise under Vibration

The DRO was required to meet the specified noise levels under a detailed set of sinusoidal and random vibrations at 1 Khz offset frequency and above. An example of the phase noise under vibration at 25 Hz is shown in Fig. 3. The calculated G sensitivity is $4*10^{-9}$ for the expected 25 Hz spur, but there is also relatively broad band

vibration induced noise in the range of 2 KHz offset which had to be eliminated.

A considerable amount of effort was expended to find vibration isolators which would eliminate the vibration induced noise above 1 KHz offset. Tests were conducted with the DRO mounted on a plate, isolated from the vibration test system by a set of three isolators. The DRO was vibrated in the Z (cylindrical) axis as well as the X and Y axes perpendicular to Z. Test data are shown only for the Z axis. Tests were conducted on a variety of isolator types including plain elastomeric pads of various compositions, manufactured elastomeric isolators, spring type mechanical isolators, and a wire rope device. One type of elastomeric pad (known as Magic Rubber from a footwear manufacturer) was found to provide acceptable isolation but this material could not be properly traced. Wire rope isolators (Aeroflex type CA2L-708) were also found to effectively eliminate the unwanted vibration responses and these were used for an extensive series of tests. Examples of the effectiveness of the CA2L-708's are given in Figs. 4 and 5. In Fig. 4, the excitation is also a 25 Hz sine wave (slightly higher in amplitude than in Fig. 3) and it is seen that the 25 Hz spur is reduced by 20 dB in comparison with the unisolated response of Fig. 3. Further, the broadband response seen in the range of 2 KHz offset on Fig. 3 is completely eliminated. The isolated G sensitivity was measured to be $3.1*10^{-10}$/g. Tests in which the oscillator was vibrated in the X and Y direction showed that the wire rope isolators were equally effective in suppressing the effects of vibration in all three axes. The isolated response of a DRO under random vibration is shown in Fig. 5 and it is noted that there are no vibration induced signals above 100 Hz offset. The resonance of the aluminum DRO-isolator system was measured to be about 10 Hz using a series of sine wave vibrations, and this resonance is the reason for the large peak at about 10 Hz seen in Fig. 5. It is noted that the brass DRO weighed about 3 times as much as the aluminum DRO and the mechanical resonance frequency seen in Fig. 5 would be reduced by a factor of 1.7 (to about 6 Hz) using brass.

Temperature Performance

In an operational system the DRO will require temperature stabilization at the high temperature operational limit of 60^0C in order to maintain the output frequency within allowed limits. A brass cavity DRO was set up to produce its maximum power output (+16 dBm) at 60^0C by adjusting the loop phase shift with the device at the high temperature. The frequency-temperature characteristic was found to have a parabolic shape with the turn-around temperature at 67^0C. At the the 60^0 C operating temperature the frequency sensitivity is 0.4 ppm /^{0}C. Additional adjustments can be made to further reduce the temperature sensitivity at 60^0 C.

Summary

The design and performance of a low noise DRO operating at 1.3 GHz has been described. The oscillator meets all the prescribed operational characteristics including noise levels under quiescent and vibrating conditions, and it has an acceptable frequency-temperature characteristic with a turn-around at the maximum operational temperature. A vibration isolation system is described which significantly reduces the vibration sensitivity of the DRO at the vibrational frequencies as well as offset frequencies well above the vibrational range.

Acknowledgements

The author wishes to thank Drs. Gary Montress and Thomas Parker for their helpful discussions and Mr. Sherwood Sheades for his work in the fabrication and testing of the oscillators used in this project.

References

(1) T. E. Parker & G. K. Montress, "Spectral purity of acoustic resonator oscillators", Proc. 46th Annual Symp. on Frequency Control, 1992, pp. 340-348

(2) G. K. Montress, & T. E. Parker, "Design techniques for achieving state-of-the-art oscillator performance", Proc. 44th Annual Symp. on Frequency Control, 1990, pp. 522-535

(3) G. K. Montress, "A high performance hybrid circuit electronic phase shifter " Proc. of RF Expo East, 1989, pp. 231-236

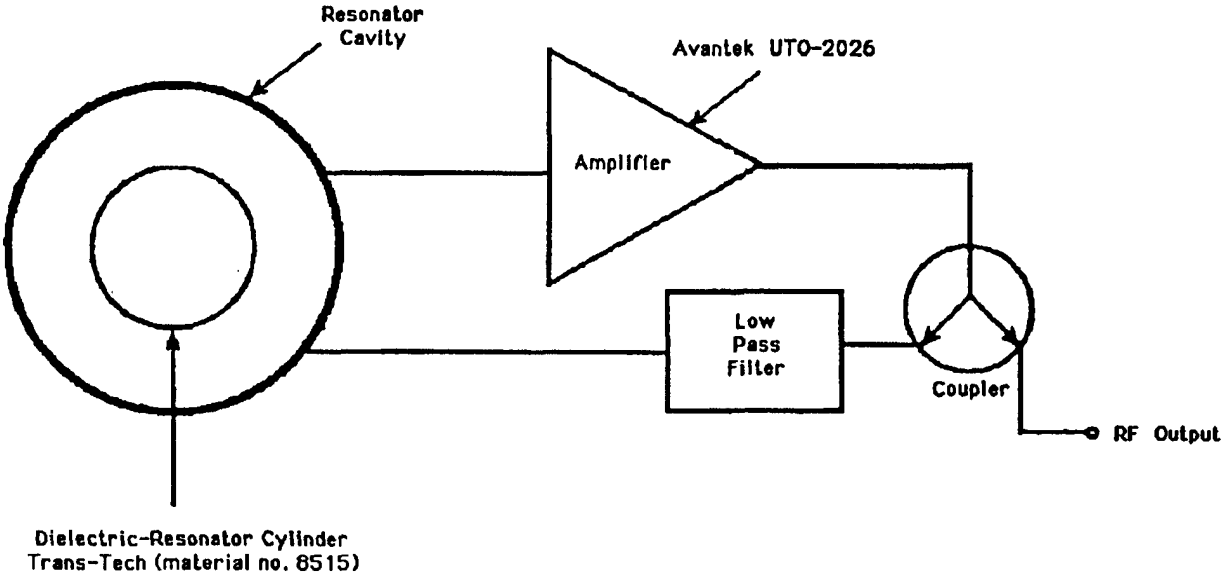


Figure 1
Schematic Diagram of the Dielectric Resonator Oscillator

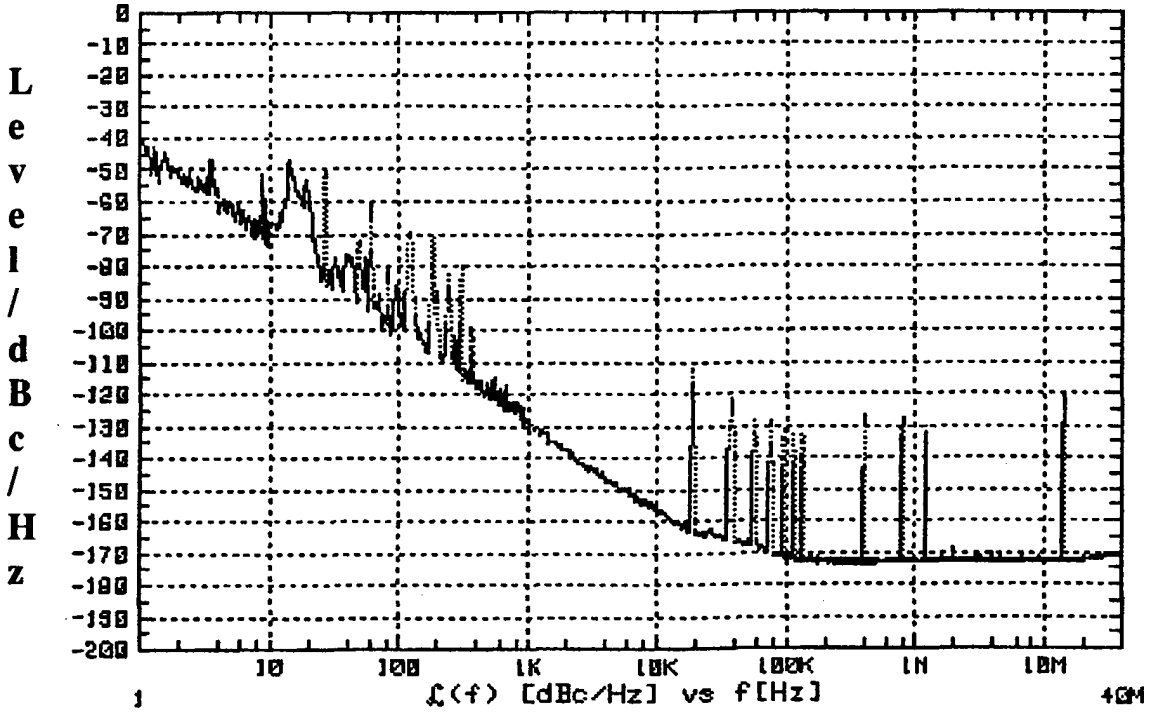


Figure 2
Dielectric resonator oscillator pair single side-band phase noise without vibration.
(Note: subtract 3dB from the measured points to obtain the noise level for a single oscillator)

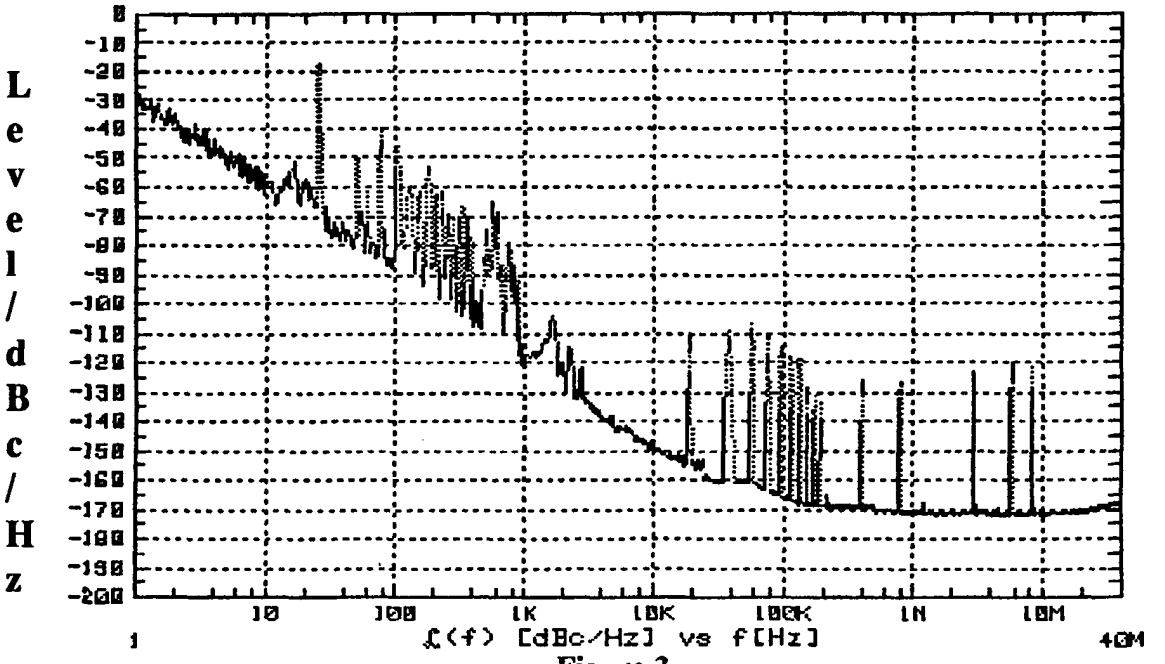


Figure 3

DRO single side-band phase noise under 25 Hz sine wave vibration of 40 mils double amplitude, without vibration isolation. The G sensitivity is 4 e-9. The vibration induced response in the neighborhood of 2 KHz should be noted.

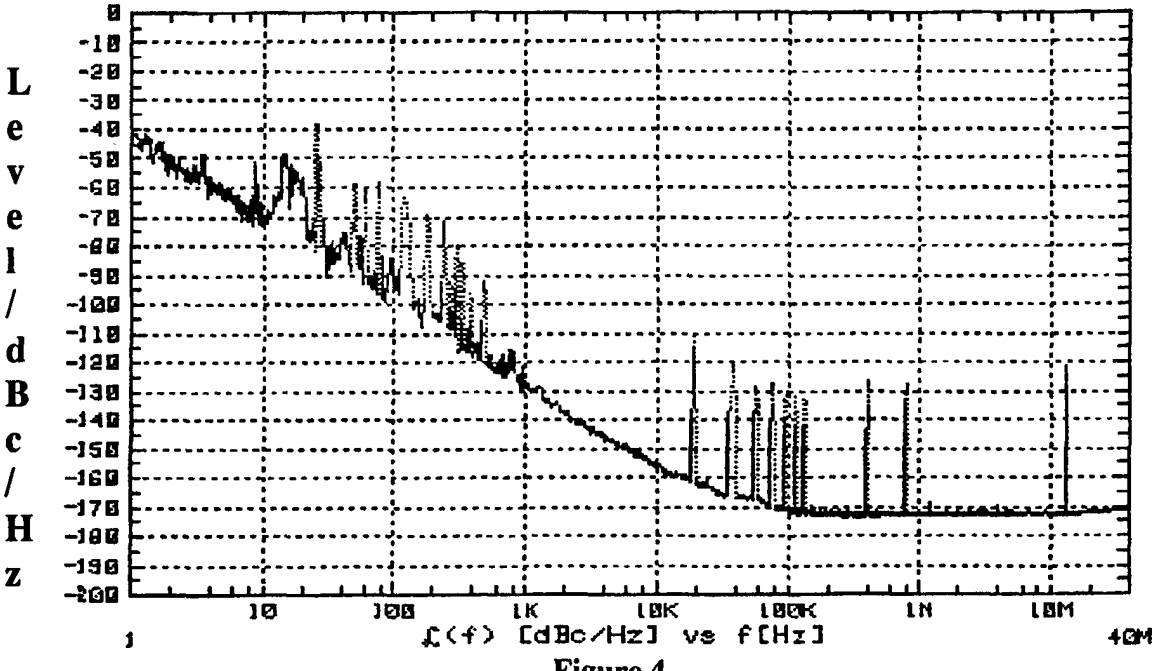


Figure 4

DRO single side-band phase noise under 25 Hz sine wave vibration of 48 mils double amplitude, with Aeroflex wire rope vibration isolation. The G sensitivity is 3.1e-10. The vibration induced response around 2 KHz seen in Fig. 3 has been suppressed.

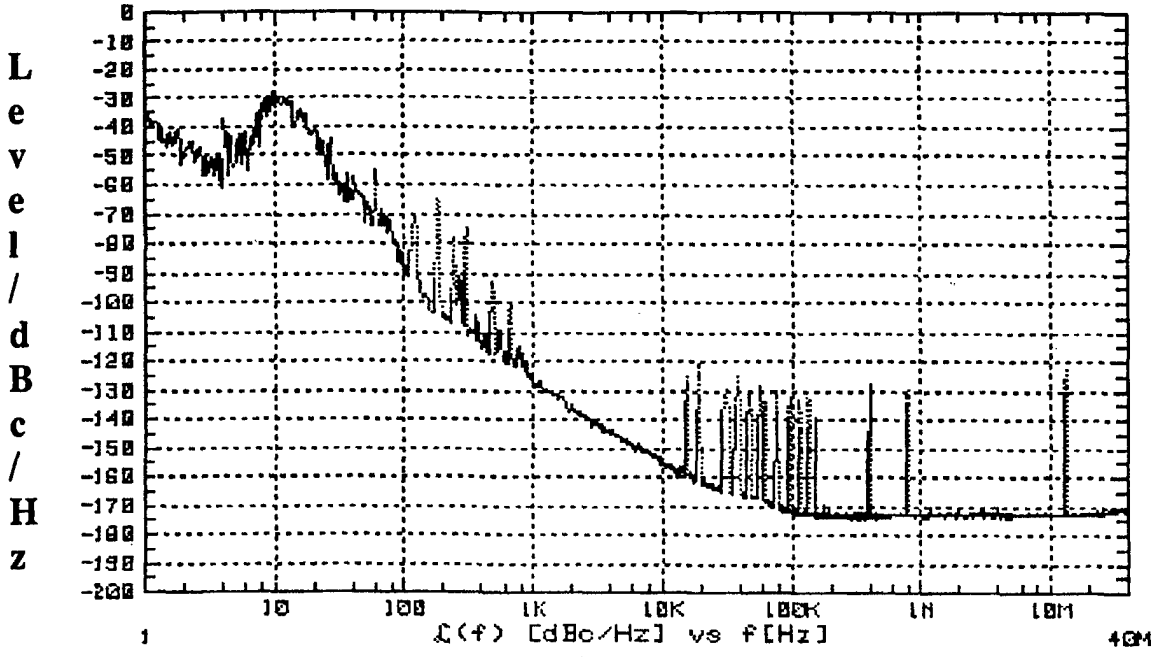


Figure 5

DRO single side-band phase noise under random vibration (0.8e-4 g2/Hz, 10 to 1000 Hz, 0.25g RMS), with Aeroflex wire rope vibration isolation. There are no vibration induced responses above 100 Hz.

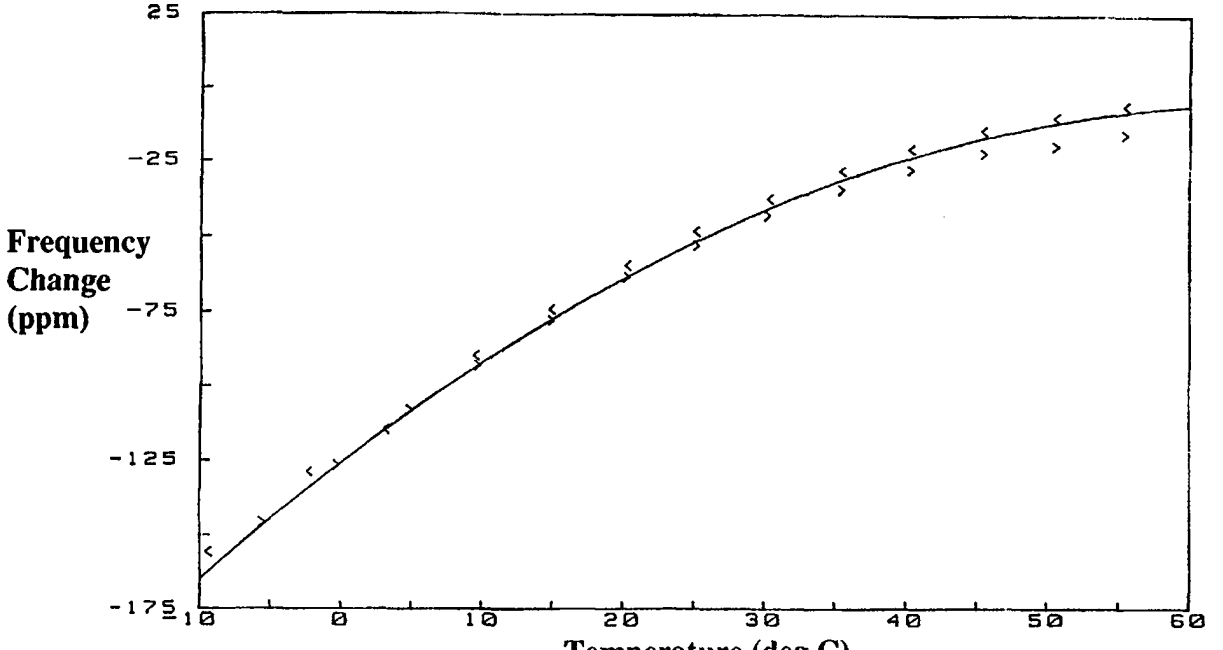


Figure 6

Frequency-Temperature characteristic of a 1.28 GHz DRO set for maximum power output at 60 deg C. The temperature sensitivity is 0.8 ppm/deg C^2 at the 60 deg C turnover point.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Supporting Structures Effects on High Q Dielectric Resonators for Oscillator Applications[†]

Ji-Fuh Liang, Chi Wang, Qin-Fen Sun and Kawthar A. Zaki
University of Maryland, Electrical Engineering Department,
College Park, MD 20742

R. W. Weinert
Westinghouse Science and Technology Center
1310 Beulah Road, Pittsburgh, PA 15235

ABSTRACT

The effects of the supporting structures on unloaded Q and resonant frequency of high Q sapphire dielectric resonators (DR) are investigated. Ring and post support structures are analyzed for TE_{01}, TE_{02} and whispering gallery modes, which are usually used in high Q sapphire DR loaded cavities for ultra low noise oscillator applications. Field distribution plots show how the fields are disturbed by the supporting structures. The effects of the supports on the unloaded Q of the cavities and the spurious spectrum are presented. The results provide valuable informations for ultrahigh Q cavity design.

I. INTRODUCTION

Cooled, ultrahigh Q sapphire DR TE_{01}, TE_{02} [1], [2], [3] and whispering gallery modes (higher order azimuthal hybrid modes) [4], [5], [6] find important applications in low noise, microwave signal generation. These modes have a common feature that the electromagnetic fields are highly concentrated in the DR. This feature combined with the extremely low loss tangent of sapphire, enables the realization of ultrahigh Q resonators. Typical unloaded Q's of the TE_{01} and TE_{02} fall in the range of 10^5 to 10^6 where superconductors might be used in parts of the enclosure. For whispering gallery modes, 10^7 to 10^9 unloaded Q's might be achieved without using superconductors. To achieve the highest possible unloaded Q of the cavity, the DR should be placed at a sufficiently far distance from the conducting enclosure to reduced the losses by the enclosure. However, support is necessary to hold the DR firmly at the center of the enclosure. The degradation of the unloaded Q and the resonant frequency (including spurious) shift from the idealized case are unavoidable. Investigations of these effects have not been reported before, and their contributions are not usually taken into account in the cavity design.

In this paper, a rigorous analysis method is applied to study the effects of supports. A general configuration of cylindrical cavities which includes the supporting structures under consideration is analyzed. Results for resonant frequencies, unloaded Q and spurious spectrum are presented. Field distributions are plotted to show how the fields inside the cavity are perturbed by the support. The results provide important information that can be used to optimize ultrahigh Q cavity designs.

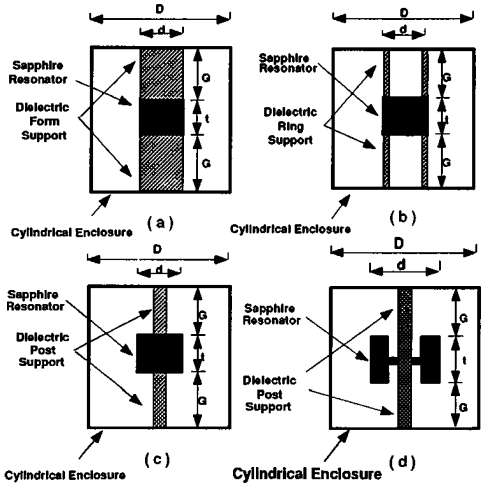


Fig. 1 Sapphire DR loaded cavity with (a) form support (b) ring support (c) post support (d) support structure for ring resonator.

II. Configuration and Analysis

Typical structures of sapphire resonators including the supports are shown in Fig. 1. Fig. 1(a) and (b) provide better mechanical performance. Fig. 1(a) should be used only for low dielectric support material while for (b) and (c) the supporting materials can be the same as the DR (i.e. sapphire) because it occupies only a small portion of the cavity. Fig. 1. (b) and (c) for TE mode operation will be studied. Fig. 1.(d) is the proposed configuration for ring resonators which is favored for spurious free region design,

[1†] Partially supported by ARPA contract # MDA972-88-C-0050

0-7803-1945-1/94 $3.00 © 1994 IEEE

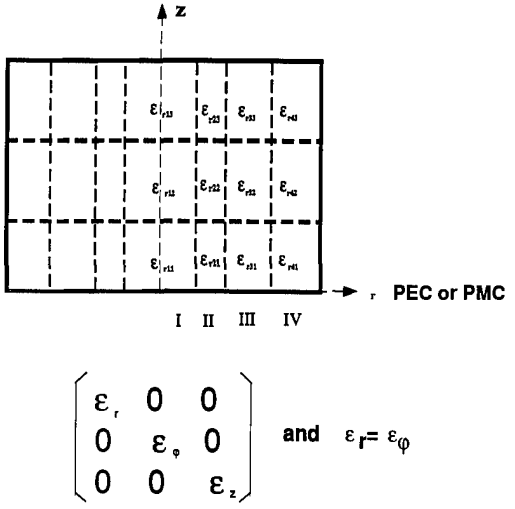


Fig. 2 Configuration of the electromagnetic boundary condition for analysis

especially for whispering gallery modes where the spectral density of the low order spurious is crucial.

Rigorous analysis method, i.e. mode-matching technique, as described in [7], [8] are extended to include the effects of the supporting structures and anisotropic characteristics of the materials.

The configuration used for analysis is shown in Fig. 2. This configuration can be adapted to Fig. 1(a), (b), (c) and (d) by properly assigning the dielectric constant of each region. In the radial direction of Fig. 2, there are four regions I, II, III and IV, while in the z-direction of each region, there are three layers with different dielectric constant. The fields in each region are expanded in terms of the modes of a three-layer inhomogeneous radial line. The mode fields of the inhomogeneous radial waveguide can be expressed analytically and the propagation constant in the radial direction is solved by matching the boundary conditions at the dielectric interfaces. Then the boundary conditions in the r-direction at the interfaces between each region are forced to be satisfied. This procedure leads to a characteristic equation. Solving the equation will give the resonant frequency and the fields expansion coefficients for each region.

Once, the fields expansion coefficients in each region are obtained, the fields are readily computed by the superposition of the mode fields. The unloaded Q of the cavity is computed by [9]:

$$Q_u = \frac{W_s \omega_o}{P_{l,c} + P_{l,d}} \tag{1}$$

where W_s is the stored energy of the cavity and is expressed as:

$$W_s = W_e + W_m = \frac{1}{4}\int_V \mu |H_o(r)|^2 dv + \frac{1}{4}\int_V \epsilon_o \epsilon_r |E_o(r)|^2 dv \tag{2}$$

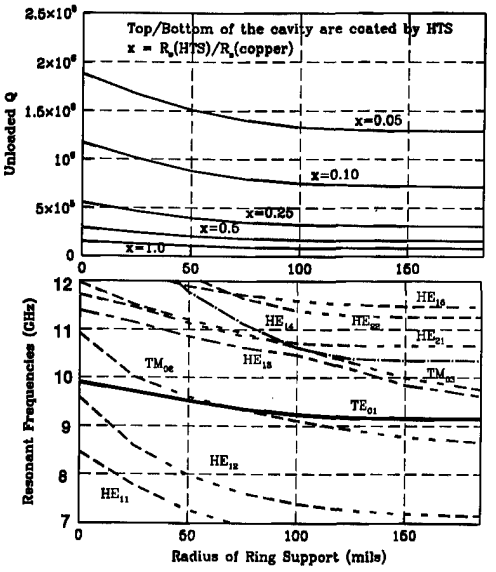


Fig. 3 Degradation of the unloaded Q and the resonant frequency shifting due to the ring support (DR thickness=0.2362"; DR radius=0.185"; Cavity radius=0.8128", Cavity length=0.551")

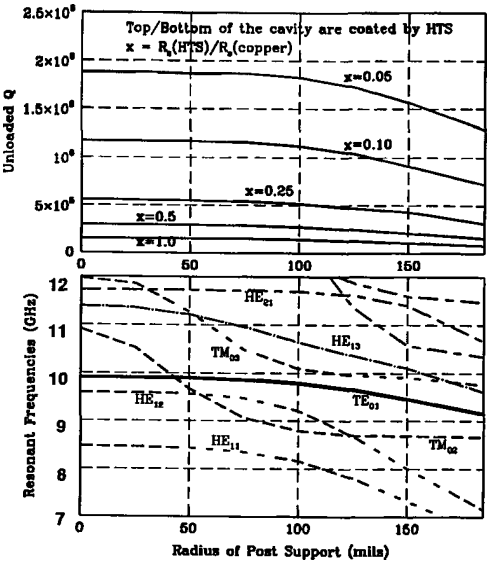


Fig. 4 Degradation of the unloaded Q and the resonant frequency shifting due to the post support (DR and cavity (dimensions: same as in Fig. 3).

The losses of the cavity are due to enclosure and dielectric losses. The enclosure power loss is:

$$P_{l,c} = \frac{1}{2}\int_S R_s |H_o(r)|^2 ds \tag{3}$$

The dielectric power loss is:

$$P_{l,d} = \frac{1}{2}\int_V \tan\delta_l \cdot (\omega_o \epsilon_r) \cdot |E_o(r)|^2 dv \tag{4}$$

Where $H_o(r)$, $E_o(r)$ are the magnetic and electric fields, R_S is the surface resistance of the conductor enclosure and $\tan\delta_t$ is the loss tangent of the dielectric material. Since the fields $E_o(r)$, $H_o(r)$ are expressed in terms of the linear superpositions of the radial waveguide mode fields, the integrations of each mode field in eqs. (2-4) can be carried out analytically. Details of the computations of the stored energies, conductor losses and dielectric losses have been well documented in [10], [11].

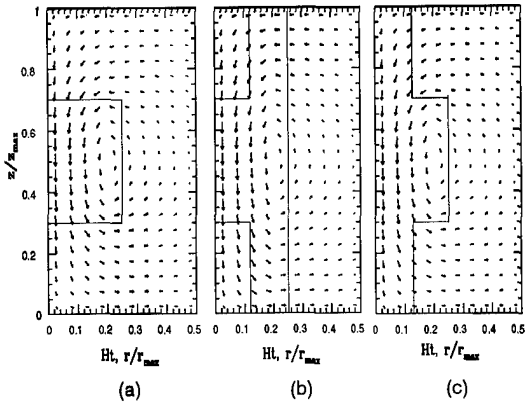


Fig. 5 Field distributions of sapphire loaded cavity (a) without support (b) with ring support (a) with post support

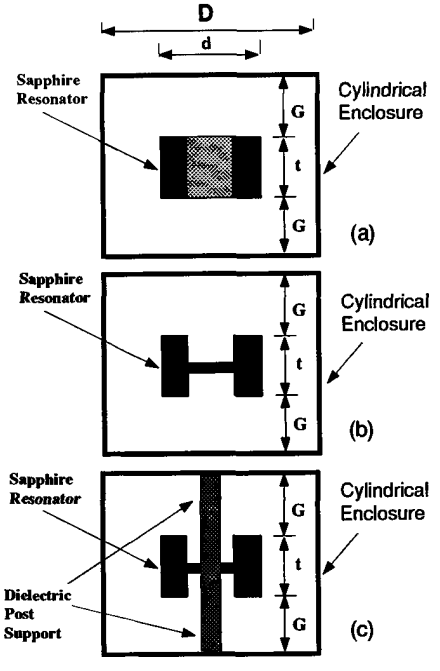


Fig. 6 Sapphire DR loaded cavity (a) ring resonator without support (a) ring resonator with web (c) ring resonator with web and post support

III. Results

Fig. 3 shows the effects of ring support on cavity resonant frequency (including spurious) and unloaded Q for a TE_{01} mode sapphire loaded cavity at 77^oK. The dimensions of the cavity and DR are determined as in [1], [2]. Fig. 4 shows the case of post support. It is clear that there are much more degradation of the unloaded Q and shift of the resonant frequencies for the case of a ring support than a post support. Field distributions in the r-z plane for the cavity without support, with ring and with post supports are shown in Fig. 5. The fields are disturbed more by the ring support and the fields close to the encloser are much stronger compared to the case of post support. The results of Fig. 3 and 4 indicate that the spurious is a strong function of location and dimensions of the support and should be taken into account for spurious free window design.

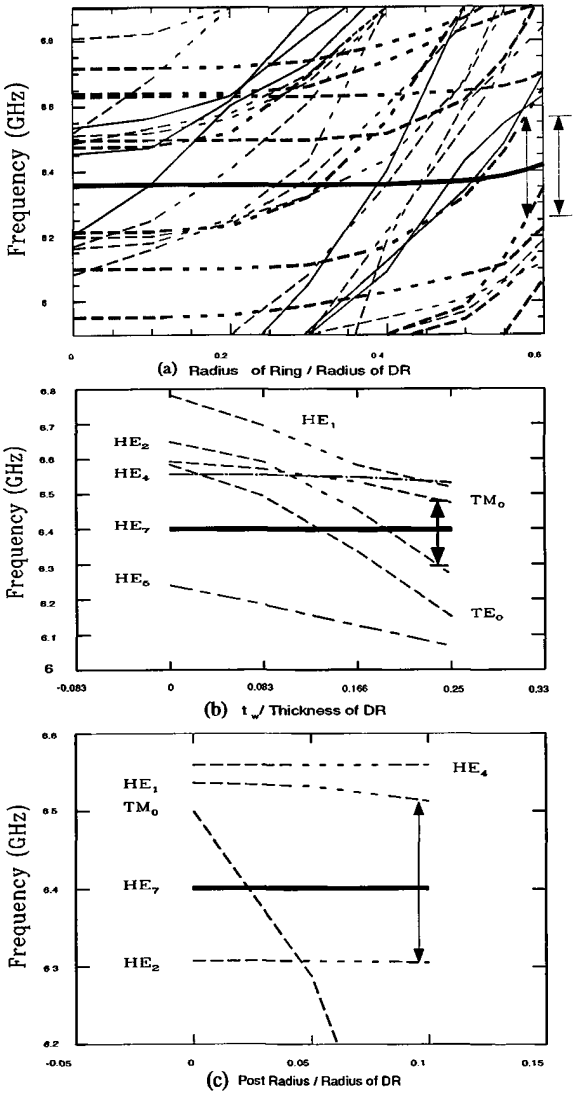


Fig. 7 (a) Mode chart of a sapphire as a function of ring diam -eter (b) relocations of the spurious due to web (c) relocation of the spurious due to the post.

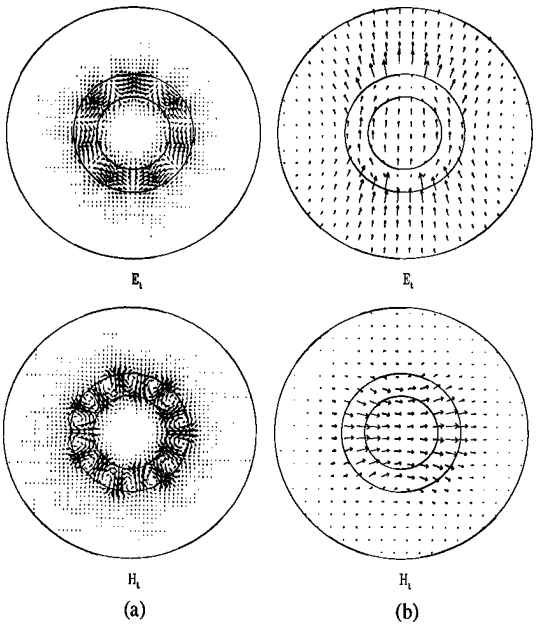


Fig. 8 Field distributions of (a) a high order (n=7) hybrid mode (whispering gallery mode) (b) a low order (n=1) hybrid mode.

The supporting structure effects on higher order modes (ϕ variation n=7) sapphire ring DR in a cavity are shown in Fig. 6 and 7. Fig. 6(a) shows the structure of a ring resonator, Fig. 6(b) is the ring DR with a layer of web and the ring resonator with both web and post is shown in Fig. 6(c). Fig. 7(a) shows a mode chart of a ring sapphire DR as in Fig. 6(a) as a function of ring diameter. A spurious free window $\pm 150MHz$ wide is found with ring radius at 60% of the outer radius of DR. Adding a layer of web at the center of the ring DR will disturb the spurious and again, a $\pm 100MHz$ window is found with suitable choice of W_t, as in Fig. 7(b). A post radius of some range, does not affect the window as shown in Fig. 7(c). Adding the supporting structure has negligible effects on desired resonant frequency and unloaded Q. Fig. 8 shows the $r-\phi$ plane field distributions of a high order (n=7) hybrid mode (i.e. whispering gallery mode) and a low order hybrid (n=1) mode. The field of the whispering gallery mode is almost completely confined inside the DR compared to the low order modes. The figures explain why the whispering gallery mode has negligible conductor losses.

IV. CONCLUSION

The degradation of unloaded Q, offset of resonant frequency and perturbation of the field distributions in an ultra high Q sapphire dielectric loaded cavity are analyzed by rigorous methods. The field distribution with and without support are plotted to show how the fields are disturbed by supports. The results of this paper provide valuable information for cooled, ultra high Q sapphire dielectric loaded cavity design.

Reference

[1] M. M. Driscoll, J. T. Haynes, R. A. Jelen, R W. Weinert, J. R. Gavaler, J. T. Talvacchio, G. R. Wagner, K. A. Zaki and X-P Liang, "Cooled, ultrahigh Q, sapphire dielectric resonators for low-noise microwave signal generation," *IEEE Trans. on Ultrasonics, Ferroelectric, and Frequency Control*, vol. Vol. 39, pp. 405-411, No. 3, MAY 1992.

[2] Kawthar A. Zaki and Xiao-Peng Liang, "Higher order modes, high Q dielectric resonators for oscillator applications," *Proceedings of IEEE Frequency Control Symposium*, May 1992, pp. 163-166.

[3] Zhi-Yuan Shen, Charles Wilker, Philip Pang and William L. Holstein, "High Tc superconductor-sapphire microwave resonator with extremely high-Q values up to 90 K," *IEEE MTT-S International Microwave Symposium Digest*, pp. 193-196, 1992.

[4] David G. Santiago and G. John Dick, "Microwave frequency discriminator with a cooled sapphire resonator for ultra-low phase noise," *Proceedings of IEEE Frequency Control Symposium*, May 1992, pp. 176-182.

[5] G. John Dick and John Saunders, "Measurements and analysis of a microwave oscillator stabilized by a sapphire dielectric ring resonator for ultra-low noise," *IEEE Trans. on Ultrasonics, Ferroelectric, and Frequency Control*, vol. Vol. 37, pp. 339-346, No. 5, Sept.. 1990.

[6] D. G. Blair and S. K. Jones, "A high-Q sapphire loaded superconducting cavity resonator," *J. Phys. D: Appl. Phys.* , vol. 20, pp. 1559-1566, 1987.

[7] K. A. Zaki and C. Chen, "New results in dielectric loaded resonators," without iris *IEEE Trans. Microwave Theory Tech.*, vol. MTT-34, pp. 815-826, Dec. 1985.

[8] S.-W. Chen and K. A. Zaki, "Dielectric ring resonators loaded in waveguide and on substrate," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-39, pp. 2069-2076, Dec. 1991.

[9] R. F. Harrington, "Time-Harmonic Electromagnetic Fields," New York, NY: McGraw-Hill Book Company, Inc., 1961.

[10] H. C. Chang and K. A. Zaki, "Unloaded Q's of axially asymmetric modes of dielectric resonators," *IEEE MTT-S International Microwave Symposium Digest*, pp. 1231-1236, 1989.

[11] X. P. Liang, H. C. Chang and K. A. Zaki, "Unloaded Q's of hybrid mode of dielectric resonators on grounded or suspended substrates," *Journal of Electromagnetic waves and Applications*, Vol. 5, No. 3, pp. 281-299, 1991.

EXPERIMENTAL TEST AND APPLICATION OF A 2-D FINITE ELEMENT CALCULATION FOR WHISPERING GALLERY SAPPHIRE RESONATORS*

David G. Santiago, Rabi T. Wang and G. John Dick
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

and

Roberto A. Osegueda, Joseph H. Pierluissi, Luis M. Gil,
Arturo Revilla and Gustavo J. Villalva
The University of Texas at El Paso
College of Engineering
El Paso, TX 79968

Abstract

This paper discusses the demonstrated accuracy and utility of a 2-D finite element methodology for whispering gallery mode sapphire resonators. The mode solutions obtained by the software compared with experimental results for a wheel-shaped sapphire resonator give an error in mode frequency of less than .55%. We also show parts per million agreement with analytical solutions for simple geometries such as an empty coaxial resonator. The CYRES 2D FEM software package developed at The University of Texas at El Paso has proven invaluable for the analysis and identification of modes and mode families for resonators of various geometries. The software also shows promise as a tool for optimization of new resonator designs. Current uses include design of optimum sized dielectric resonators for minimized wall losses, and new resonator geometries for temperature compensated resonators. The operational characteristics of the software and the general methodology for use of the software as a laboratory and design tool are discussed.

Introduction

Recently, whispering gallery mode (WGM) resonators consisting of a sapphire dielectric element in a metallic container, as in Figure 1, have made possible new capabilities for microwave oscillator phase noise and frequency stability [1,2,3,4]. With high azimuthal mode numbers, these resonators isolate radio-frequency energy to the dielectric element and away from the metallic container, thus providing extraordinarily high quality factors (Q's >10 million). Design and analysis of such a resonator with a certain frequency and Q requires a finite element method (FEM) model with higher resolution and accuracy than any previously available.

Modeling Challenges

The characteristics and exceptional properties of sapphire whispering gallery mode resonators make them particularly challenging to model using finite element analysis. The finite element model must allow for widely disparate field magnitudes in order to resolve the hybrid fields that exist in WGM resonators. The electromagnetic fields of whispering gallery modes are well confined to the sapphire element [5]. This creates areas of high field strength in the sapphire and weak field areas outside the sapphire, presenting a substantial burden for any calculational technique. Accurate modeling of these critical areas requires a high FEM node density to provide sufficient resolution. Because even small wall losses are important for preserving the very high Q's, accurate calculation of the low field strengths at the container wall by the FEM model is essential. Furthermore, the coupling ports for our WGM resonator are located in the evanescent region, and so an accurate determination of these weak field strengths is necessary to determine resonator coupling.

FEM software must eliminate spurious modes without losing any of the many modes found in a WGM resonator. The anisotropy of the sapphire dielectric must also be considered for proper FEM modeling. When the z-axis is aligned with the sapphire crystal c-axis, the resonator is anisotropic in two dimensions with $\varepsilon_z \approx 11$ and $\varepsilon_r = \varepsilon_\phi \approx 9$. A three-dimensional

* The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

0-7803-1945-1/94 $3.00 © 1994 IEEE

FEM allowing full treatment of sapphire's anisotropic dielectric constant, would require such a large number of elements as to be impractical. Approximate analytical methods are useful for some geometries, but a new approach would be required for every change in geometry. A two-dimensional finite element approach, however, allows easy treatment of any cylindrically symmetric resonator geometry.

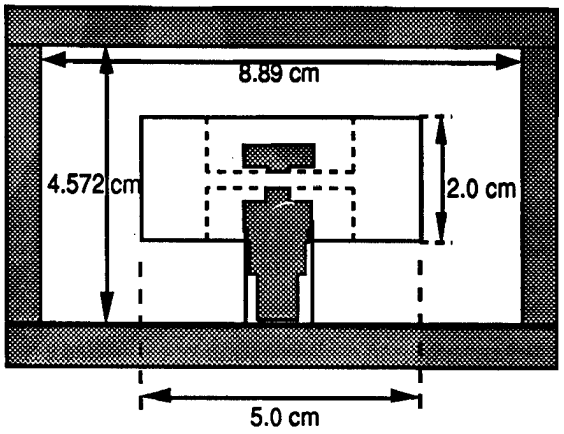


Figure 1. Sapphire Resonator in Containment Can

CYRES 2D

Because the dielectric constant for sapphire shows cylindrical symmetry, a two dimensional treatment is allowed for the important case where its crystal c axis is aligned with a physical axis of axisymmetry. While ruling out most anisotropic dielectric configurations, this approach makes possible the first two-dimensional finite element treatment for sapphire "whispering gallery" resonators. For the different modes of a given circularly symmetric geometry, the CYRES 2D finite element package, under development at the University of Texas at El Paso, allows determination of resonant frequencies and visualization of the electromagnetic fields [6,7].

A finite element mesh of rectilinear or curved elements defines the resonator in the r-z plane. Maxwell's equations for the anisotropic case are solved for the cross section. To complete the 3D solution, the method assumes a sinusoidal dependence in azimuth. The 2D solution is projected around the resonator by multiplication with the azimuthal sinusoidal dependence. These field solutions provide the resonant frequencies and cavity Q's(determined by wall losses) of each mode. A penalty term in the vector Helmholtz equations is used to suppress spurious solutions from appearing in the final list of resonant modes.

Validation

During development the FEM solutions were compared with several analytically determined solutions to test the software. Upon delivery to JPL from UTEP the software underwent additional validation testing to determine the FEM model's accuracy limits for different mesh densities. Figure 2 shows the fractional frequency

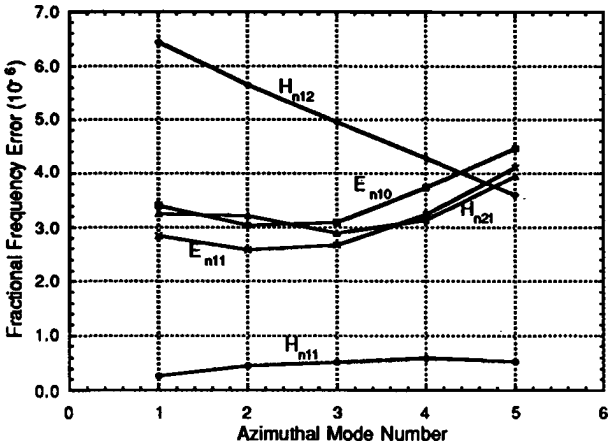


Figure 2. FEM Data Compared to Analytic Solution for an Empty Coaxial Resonator

error of the CYRES mode solutions for an empty coaxial resonator compared to the analytic exact solution. As seen in the figure, the largest error is about 6.5 parts per million.

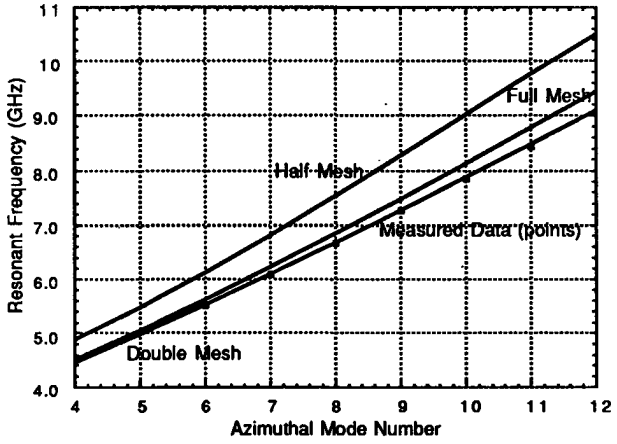


Figure 3. FEM calculated WGM frequencies for different mesh densities compared to laboratory measurements

High mesh densities for FEM calculations produce higher accuracy results with greater resolution. The benefits of superior accuracy are achieved at the expense of computational complexity and speed. Typically, our chosen quadrilateral element "full mesh" contains about 200 nodes. Calculation of the resonant mode solutions is also performed at half this mesh density (half mesh)

and double this density (double mesh), which respectively reduce and increase the number of nodes by a factor of four. Figure 3 shows the half, full and double mesh density FEM calculated resonant frequencies for the WGM resonator of Figure 1. The points which fall along the double mesh line are the laboratory measurements of the resonant frequencies.

By choosing a full mesh density of sufficient accuracy we can use the half mesh and double mesh calculations to extrapolate FEM solutions to an infinite mesh density, thereby providing excellent modeling accuracy for a modest computational expense. Figure 4 shows the fractional frequency difference of the three mesh densities and the measured data from the

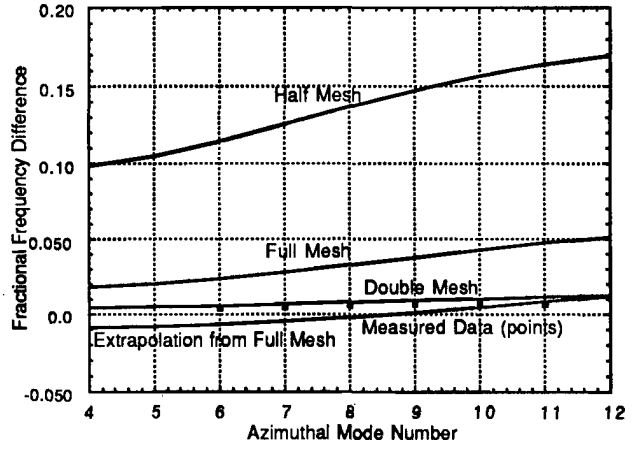


Figure 4. Difference comparison of FEM calculated frequencies for different mesh densities to laboratory measurements (zero equals extrapolation to infinite mesh)

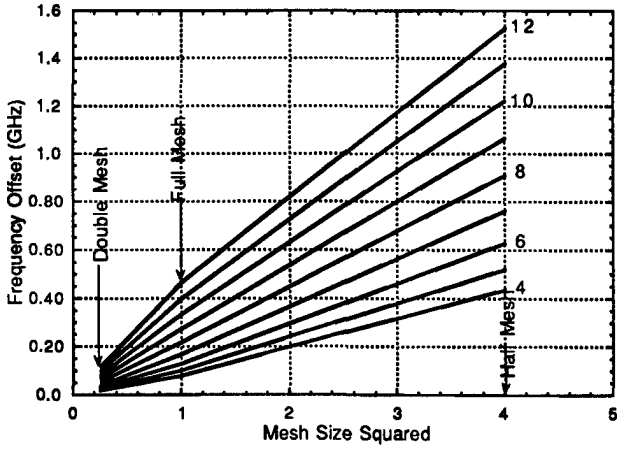


Figure 5. Calculated FEM Frequency offset vs. Mesh Density

extrapolation to infinite mesh density. A quadratic extrapolation to infinite mesh based on full and double meshes defines the zero on the y-axis. The measured data points are also plotted and fall no more than 0.5% above the extrapolated solution. Also shown is the "Extrapolation from Full Mesh" that uses only the half and full mesh solutions to extrapolate to an infinite mesh solution.

Higher order modes intrinsically require higher resolution calculations to provide accuracy equivalent to lower order mode solutions. Figure 5 shows the offset from the infinite mesh mode frequency solutions of the different mesh densities for different order modes. The nearly straight line plots lend credence to the quadratic extrapolation procedure used for Figure 4.

Laboratory Tool

The CYRES FEM modeling software has been predominately used as a laboratory tool for mode identification and analysis. Current laboratory research procedures for research resonators make full use of the software package. Use of the software requires the material properties and dimensions of the radial cross section be written in an input file. The PC based pre-processor generates a mesh of specified density defining the resonator and its boundary conditions. The mesh information feeds the FEM processor running on a Sun workstation or Cray supercomputer. Output for each azimuthal mode number includes a list of cavity resonant frequencies and wall loss Q's, and a file of vector magnetic field values. The field file is post-processed on a PC to graphically display the magnetic and electric field vectors in three different resonator cross sections. The field visualization is used to label the modes in the software's list solutions.

In the laboratory, modes are found by sweeping the input frequency and logging the frequency, Q, and coupling coefficient for each resonant mode. This list is then preliminarily matched by frequency with the finite element data. Because of the arrangement of our coupling ports it is not possible to experimentally find all of the modes indicated by the FEM calculations. Analysis of the electromagnetic visualization of the resonator cross section identifies the experimental modes for the geometry. Experimentally, high Q (>5 million) whispering gallery modes are found to have weak coupling to the output port due to the low field strength near the can wall. Low Q modes are often found to be strongly coupled to the output port, thereby indicating a relatively large field strength at the can wall. A strongly coupled high Q mode indicates a well behaved first order high azimuthal number whispering gallery mode which can reach design Q's (>12 million) when properly (critically or slightly weakly) coupled to the output port.

Using this information, the modes are identified and grouped into whispering gallery mode families by comparing the characteristics of the measured resonant modes to the those of the FEM modes.

As a laboratory analysis tool the software can suggest performance deficiencies of a particular mode, indicate possible design improvements, and identify possible mode competition.

Design Tool

More important than the software's ability to model existing resonators, may be its design capabilities for new resonators. The FEM solution's have shown to properly and accurately model every resonator we have constructed. It is expected that the software's performance will be similar for new resonator designs of cylindrical symmetry in dielectric and geometry.

New resonator geometries are being developed for improved performance including enhanced Q's [8], stability, coupling control and, spurious mode suppression. The software can model the performance of the actual resonator. In this way, the effects of geometrical modifications on the resonators' performance can be tested without construction of experimental units. A well confined, high Q mode can be chosen around which the resonator system will be designed. The appropriate cavity for low wall losses and proper coupling can also be designed in the process. Other modes near the design frequency can be studied and selected for suppression or use as diagnostic signals.

New materials can also be implemented if they have at most a two dimensional anisotropy. Also the consequences of introducing tuning elements, probes or other perturbations can be studied. Another very important use is quantifying the predictions of the many qualitative models used for preliminary designs of resonators. For example, the software may show that the capacitive model used to design many of our resonator tuning elements improperly predicts the magnitude of the tuning in a new design.

Conclusion

The accuracy of the CYRES 2D FEM software methodology has been validated by comparisons with analytical and experimental results. The software has proven to be an invaluable tool in the laboratory for mode identification and analysis. These results demonstrate the accuracy of the CYRES 2D software as an design tool for optimization of new resonators.

References

[1] 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).

[2] G. J. Dick, D. G. Santiago and R. T. Wang, "Temperature-Compensated Microwave Sapphire Oscillator at Temperatures Above 80 Kelvin," *These Proceedings.*

[3] A. G. Mann, et al, "Ultra-Stable Cryogenic Sapphire Dielectric Microwave Resonators," *Proc. 1992 IEEE Frequency Control Symposium*, 167-171 (1992).

[4] M. E. Tobar, E. N. Ivanov, R. A. Woode, J. H. Searls, and S. Edwards, "Low Noise Microwave Oscillators Based on High-Q Temperature Stabilized Sapphire Resonators," *These Proceedings.*

[5] G. J. Dick and D. M. Strayer, "Measurements and Analysis of Cryogenic Sapphire Dielectric Resonators and DRO's," *Proc. 41st Ann. Symposium on Frequency Control*, 487-491 (1987).

[6] R. A. Osegueda, J. H. Pierluissi, L. M. Gil, A. Revilla, G. J. Villalva, G. J. Dick, D. G. Santiago and R. T. Wang, "Azimuthally-Dependent Finite Element Solution to the Cylindrical Resonator," to be published in *Proc. 10th Ann. Rev. of Progress in Appl. Computational Electromagnetics (ACES)* (1994).

[7] R. A. Osegueda, A. Revilla, J. H. Pierluissi, and G. J. Dick, "2-D Finite Element Program to Compute Resonant Frequencies and 3-D Visualization of Electromagnetic Modes in Cylindrical Resonators," *These Proceedings.*

[8] E. N. Ivanov, V. I. Kalinichev and D. G. Blair, "Sapphire Loaded Microwave Resonators with Enhanced Quality Factors," *These Proceedings.*

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

STRESS INDUCED FREQUENCY SHIFT OF DIELECTRIC RESONATORS WITH MAGNETIC WALL

J. S. Yang
Department of Civil Engineering and Operations Research
Princeton University, Princeton, New Jersey 08544, U.S.A.

X. Y. Wu
Department of Civil and Environmental Engineering
Carleton University, Ottawa, Ontario, Canada K1S 5B4

Abstract

This paper presents a perturbation analysis on the stress or strain induced frequency shift through piezooptic or photoelastic effect in dielectric resonators with open circuit boundary or magnetic wall. Based on the availability of the frequency and mode when stress is not present, prediction on first order stress induced frequency shift is obtained for a resonator of arbitrary shape and with general dielectric permittivity tensor. The result is in fact also valid for resonators with short circuit boundary or electric wall, or resonators with partial open circuit boundary and partial short circuit boundary.

I. Introduction

When a dielectric resonator is subjected to steady forces or accelerations, strains are produced. The strains, through the nonlinear piezooptic or photoelastic effect,[1] induce changes in the impermeability tensor which in turn affect the velocity of electromagnetic waves and hence cause frequency shift.

Stress induced frequency shift were computed[2] for TE modes in a circular disk dielectric resonator in free space based on a set of two-dimensional equations[3] for guided EM (electromagnetic) waves in dielectric plates and on solutions for frequencies and modes of the resonator when there is no stress present.[4] Stress effect on frequency for anisotropic dielectric plate waveguides has also been studied.[5]

Dielectric resonators with open circuit boundary are easier to be dealt with than those in free space. Since our may concern is the stress effect, it is informative to study resonators with open circuit boundary and it is hoped that this will reveal some basic features of the stress effect on dielectric resonators. Historically, resonators with open circuit boundary were among the earliest approximations for resonators in free space.[6]

In the present paper, a perturbation method is employed to calculate the frequency shift for dielectric resonators with open circuit boundary. Based on the availability of the frequency and mode when stress is not present (which is true for resonators with magnetic wall in several cases), prediction of first order frequency shift due to strain or stress is obtained for a resonator of arbitrary shape and with general dielectric permittivity tensor. The result is in fact also valid for resonators with short circuit boundary or electric wall, or resonators with partial open circuit boundary and partial short circuit boundary. For an initially isotropic rectangular resonator with open circuit boundary, frequency shift of TE and TM modes under unidirectional compression, vertical acceleration, and horizontal acceleration, are calculated.

II. Governing Equations

In the general formulation of the problems with piezooptic effect, the governing equations of the EM fields are coupled to the elastic fields by strains through the piezooptical effect in the dielectric constitutive relations. The problem in general is nonlinear. One class of problems can be formulated as the superposition of small, incremental elastic and EM fields on finite, static initial elastic fileds. For this kind of problems, the governing equations become[7]
For the static initial elastic fields

$$\begin{aligned} T_{ij,i} + \rho f_j &= 0 \\ S_{ij} &= \frac{1}{2}(u_{j,i} + u_{i,j}) \\ T_{ij} &= c_{ijkl} S_{kl} \end{aligned} \tag{1}$$

0-7803-1945-1/94 $3.00 © 1994 IEEE

For the incremental EM fields

$$\begin{aligned} &\epsilon_{ijk}H_{k,j} = D_{i,t}\ , \quad B_{i,i} = 0 \\ &\epsilon_{ijk}E_{k,j} = -B_{i,t}\ , \quad D_{i,i} = 0 \\ &D_i = \epsilon_o\epsilon_{ij}(\mathbf{S})E_j\ , \quad B_i = \mu_o H_i \end{aligned} \tag{2}$$

We note that (1) are the equations of elasticity for anisotropic solids, where T_{ij} is the stress tensor, S_{ij} the strain, u_j the displacement vector, ρ the mass density, c_{ijkl} the elastic stiffness tensor and f_j the body force per unit mass. And (2) are the governing equations of EM field in dielectrics where E_i and H_i are the electric and magnetic filed vectors, D_i electric displacement, B_i the magnetic induction, ϵ_{ijk} the unit alternating tensor, ϵ_o and μ_o the dielectric permittivity and magnetic permeability of free space, and ϵ_{ij} the relative permittivity tensor which depends on the strain S_{ij} because of the piezooptic effect. The above equations are are not coupled to the equations for incremental elastic fields hence those equations are not included here.

The solution procedure is that the static initial elastic fileds are solved first from (1), then the incremental EM fields are solved from (2) with the known initial strain fileds as variable coefficients of the equations. In the following, we will assume that the initial elastic fields have been obtained, and concentrate on solving the incremental EM fields.

For a finite dielectric resonator, let the spatial region occupied by the resonator be V, the boundary of V be S, the unit outward normal of S be n_i, then the governing equations and boundary conditions are

$$\begin{aligned} &\epsilon_{ijk}H_{k,j} = D_{i,t}\ , \quad B_{i,i} = 0 \quad \text{in} \quad V \\ &\epsilon_{ijk}E_{k,j} = -B_{i,t}\ , \quad D_{i,i} = 0 \quad \text{in} \quad V \\ &D_i = \epsilon_o\epsilon_{ij}E_j\ , \quad B_i = \mu_o H_i \quad \text{in} \quad V \\ &n_i D_i = 0\ , \quad \epsilon_{ijk}n_j H_k = 0 \quad \text{on} \quad S \end{aligned} \tag{3}$$

where E_i and H_i are the electric and magnetic intensity vectors, D_i and B_i are the electric and magnetic flux vectors, ϵ_{ijk} the permutation tensor, ϵ_o the electric permittivity of free space ϵ_{ij} the relative dielectric permittivity tensor, and μ_o the magnetic permeability of free space.

For time harmonic motions, let

$$\begin{aligned} &E_i(\mathbf{x},t) = -iE_i(\mathbf{x})e^{i\omega t} \\ &D_i(\mathbf{x},t) = -iD_i(\mathbf{x})e^{i\omega t} \\ &H_i(\mathbf{x},t) = H_i(\mathbf{x})e^{i\omega t} \\ &B_i(\mathbf{x},t) = B_i(\mathbf{x})e^{i\omega t} \end{aligned} \tag{4}$$

Then (3) becomes

$$\begin{aligned} &\epsilon_{ijk}H_{k,j} = \omega D_i\ , \quad B_{i,i} = 0 \quad \text{in} \quad V \\ &\epsilon_{ijk}E_{k,j} = \omega B_i\ , \quad D_{i,i} = 0 \quad \text{in} \quad V \\ &D_i = \epsilon_o\epsilon_{ij}E_j\ , \quad B_i = \mu_o H_i \quad \text{in} \quad V \\ &n_i D_i = 0\ , \quad \epsilon_{ijk}n_j H_k = 0 \quad \text{on} \quad S \end{aligned} \tag{5}$$

We note that in (5), for time harmonic motions, the two curl equations imply the two divergence equations. Therefore only the two curl equations need to be kept. For the same reason, the boundary condition on the normal component D_i may also be dropped. Using the constitutive equations $(5)_3$ to eliminate D_i and B_i, we have

$$\begin{aligned} &\epsilon_{ijk}H_{k,j} = \omega\epsilon_o\epsilon_{ij}E_j \quad \text{in} \quad V \\ &\epsilon_{ijk}E_{k,j} = \omega\mu_o H_i \quad \text{in} \quad V \\ &\epsilon_{ijk}n_j H_k = 0 \quad \text{on} \quad S \end{aligned} \tag{6}$$

The inverse of ϵ_{ij} is denoted by β_{ij} and is called the relative dielectric impermeability tensor. The presence of strain field S_{kj}, through the piezooptic effect, causes a change in β_{ij} such that

$$\beta_{ij} = \beta^o_{ij} + \Delta\beta_{ij} \tag{7}$$

where β^o_{ij} is the initial relative impermeability tensor when strain field is not present, and $\Delta\beta_{ij}$ is the strain induced change. For an initially isotropic medium, we have the following Pockel's formulation[8]

$$\Delta\beta_{ij} = p_{ij(kl)}S_{kl} \tag{8}$$

where a pair of parenthesis in the subscripts means symmetrization of the indices in the parenthesis.

For an initially anisotropic medium, Nelson and Lax found that the Pockel's formulation is incomplete and showed that[8]

$$\Delta\beta_{ij} = p_{ij(kl)}S_{kl} + p_{ij[kl]}R_{kl} \tag{9}$$

where R_{kl} is the rotation tensor of the deformation of the medium, a pair of square brackets in the subscripts means anti-symmetrization of the indices in the brackets, and $p_{ij[kl]}$ is related to β^o_{ij} as[8]

$$p_{ij[kl]} = \frac{1}{2}(\beta^o_{il}\delta_{jk} + \beta^o_{lj}\delta_{ik} - \beta^o_{ik}\delta_{jl} - \beta^o_{kj}\delta_{il}) \tag{10}$$

The above formulations are for initially homogeneous media. For an initially inhomogeneous medium, since the translation of a differential continuum element at a spatial location also causes a change of β^o_{ij} at that location, another term needs to be added to represent the translation effect. Hence we propose the following general formulation

$$\Delta\beta_{ij} = p_{ij(kl)}S_{kl} + p_{ij[kl]}R_{kl} - \beta^o_{ij,k}u_k \tag{11}$$

where u_i is displacement.

Corresponding to (7), for ϵ_{ij} we have

$$\epsilon_{ij} = \epsilon^o_{ij} + \Delta\epsilon_{ij} \tag{12}$$

where ϵ^o_{ij} is the initial relative permittivity tensor when strain field is not present, and $\Delta\epsilon_{ij}$ is the strain induced change.

According to the measured values listed in Landolt-Börnstein,[9] β^o_{ij} are in the order of 1 to 10^{-2} and p_{ijkl} in the order of 10^{-1}. Strains due to the applied forces and accelerations in many experimental measurements are usually in the order of 10^{-5}. Therefore $\Delta\beta_{ij}$ is in the order of 10^{-6} and much smaller than β^o_{ij}. Hence $\Delta\epsilon_{ij}$ can be expressed by $\Delta\beta_{ij}$ approximately as follows.[8] Neglecting second order terms and noting $\epsilon^o_{ij}\beta^o_{jk} = \delta_{ik}$ in

$$(\epsilon^o_{ij} + \Delta\epsilon_{ij})(\beta^o_{jk} + \Delta\beta_{jk}) = \delta_{ik} \tag{13}$$

we obtain

$$\Delta\epsilon_{ij}\beta^o_{jk} + \epsilon^o_{ij}\Delta\beta_{jk} = 0 \tag{14}$$

and then contracting with ϵ^o_{kl} and using $\beta^o_{jk}\epsilon^o_{kl} = \delta_{jl}$ we have

$$\Delta\epsilon_{il} = -\epsilon^o_{ij}\Delta\beta_{jk}\epsilon^o_{kl} \tag{15}$$

III. Perturbation Solution

Since the strain induced change in ϵ_{ij} is small, ϵ_{ij} can generally be represented by

$$\epsilon_{ij} = \epsilon^o_{ij} + \sigma\epsilon^*_{ij} \tag{16}$$

where σ is a small, dimensionless parameter, and $\sigma\epsilon^*_{ij} = \Delta\epsilon_{ij}$.

Corresponding to (16), the following perturbation expansion can be made

$$\begin{aligned} \omega &= \omega^o + \sigma\omega^* + O(\sigma^2) \\ E_i &= E^o_i + \sigma E^*_i + O(\sigma^2) \\ H_i &= H^o_i + \sigma H^*_i + O(\sigma^2) \end{aligned} \tag{17}$$

Substitution of (16) and (17) into (6) gives

$$\begin{aligned} &\epsilon_{ijk}H^o_{k,j} + \sigma\epsilon_{ijk}H^*_{k,j} \\ &\quad = \omega^o\epsilon_o\epsilon^o_{ij}E^o_j + \sigma\omega^*\epsilon_o\epsilon^o_{ij}E^o_j \\ &\quad + \sigma\omega^o\epsilon_o\epsilon^*_{ij}E^o_j + \sigma\omega^o\epsilon_o\epsilon^o_{ij}E^*_j + O(\sigma^2) \quad \text{in} \quad V \\ &\epsilon_{ijk}E^o_{k,j} + \sigma\epsilon_{ijk}E^*_{k,j} \\ &\quad = \omega^o\mu_oH^o_i + \sigma\omega^*\mu_oH^o_i \\ &\quad + \sigma\omega^o\mu_oH^*_i + O(\sigma^2) \quad \text{in} \quad V \\ &\epsilon_{ijk}n_jH^o_k + \sigma\epsilon_{ijk}n_jH^*_k + O(\sigma^2) = 0 \quad \text{on} \quad S \end{aligned} \tag{18}$$

Collecting coefficients of terms of like powers of σ on both sides in (18), we have the following perturbation problems of successive orders.

Zeroth order problem

$$\begin{aligned} \epsilon_{ijk}H^o_{k,j} &= \omega^o\epsilon_o\epsilon^o_{ij}E^o_j \quad \text{in} \quad V \\ \epsilon_{ijk}E^o_{k,j} &= \omega^o\mu_oH^o_i \quad \text{in} \quad V \end{aligned} \tag{19}$$

$$\epsilon_{ijk} n_j H_k^o = 0 \quad \text{on} \quad S$$

First order problem

$$\begin{aligned}
&\epsilon_{ijk} H^*_{k,j} = \omega^* \epsilon_o \epsilon^o_{ij} E^o_j + \omega^o \epsilon_o \epsilon^*_{ij} E^o_j \\
&\quad + \omega^o \epsilon_o \epsilon^o_{ij} E^*_j \quad \text{in} \quad V \\
&\epsilon_{ijk} E^*_{k,j} = \omega^* \mu_o H^o_i + \omega^o \mu_o H^*_i \quad \text{in} \quad V \\
&\epsilon_{ijk} n_j H^*_k = 0 \quad \text{on} \quad S
\end{aligned} \tag{20}$$

The zeroth order problem can be solved exactly in a few cases.[6] Now we proceed to solve the first problem. Multiplying $(20)_1$ by E^o_i and $(20)_2$ by H^o_i and adding the resulting equations, we have

$$\begin{aligned}
&E^o_i \epsilon_{ijk} H^*_{k,j} - \omega^o E^o_i \epsilon_o \epsilon^o_{ij} E^*_j \\
&\quad + H^o_i \epsilon_{ijk} E^*_{k,j} - \omega^o \mu_o H^o_i H^*_i \\
&\quad = \omega^* E^o_i \epsilon_o \epsilon^o_{ij} E^o_j + \omega^o E^o_i \epsilon_o \epsilon^*_{ij} E^o_j + \omega^* \mu_o H^o_i H^o_i
\end{aligned} \tag{21}$$

(21) can be written as

$$\begin{aligned}
&H^*_i (\epsilon_{ijk} E^o_{k,j} - \omega^o \mu_o H^o_i) \\
&\quad + E^*_i (\epsilon_{ijk} H^o_{k,j} - \omega^o \epsilon_o \epsilon^o_{ij} E^o_j) \\
&\quad - (E^*_i \epsilon_{ijk} H^o_k)_{,j} + (E^o_i \epsilon_{ijk} H^*_k)_{,j} \\
&\quad = \omega^* (E^o_i \epsilon_o \epsilon^o_{ij} E^o_j + \mu_o H^o_i H^o_i) + \omega^o E^o_i \epsilon_o \epsilon^*_{ij} E^o_j
\end{aligned} \tag{22}$$

Integrating (22) over V, we obtain

$$\begin{aligned}
&\int_V [H^*_i (\epsilon_{ijk} E^o_{k,j} - \omega^o \mu_o H^o_i) \\
&\quad + E^*_i (\epsilon_{ijk} H^o_{k,j} - \omega^o \epsilon_o \epsilon^o_{ij} E^o_j)] dV \\
&\quad + \int_S (-E^*_i \epsilon_{ijk} n_j H^o_k + E^o_i \epsilon_{ijk} n_j H^*_k) dS \\
&\quad = \omega^* \int_V (E^o_i \epsilon_o \epsilon^o_{ij} E^o_j + \mu_o H^o_i H^o_i) dV \\
&\quad + \int_V \omega^o E^o_i \epsilon_o \epsilon^*_{ij} E^o_j dV
\end{aligned} \tag{23}$$

The integrals on the left hand of (23) vanish because of equations for the zeroth order problem (19) and the boundary condition for the first order problem $(20)_3$. Hence, we have, for ω^*

$$\omega^* = -\omega^o \frac{\int_V E^o_i \epsilon_o \epsilon^*_{ij} E^o_j dV}{\int_V (E^o_i \epsilon_o \epsilon^o_{ij} E^o_j + \mu_o H^o_i H^o_i) dV} \tag{24}$$

or, for $\Delta\omega = \sigma\omega^*$

$$\frac{\Delta\omega}{\omega^o} = -\frac{\int_V E^o_i \epsilon_o \Delta\epsilon_{ij} E^o_j dV}{\int_V (E^o_i \epsilon_o \epsilon^o_{ij} E^o_j + \mu_o H^o_i H^o_i) dV} \tag{25}$$

(25) is the equation from which the frequency shift can be calculated. Once the solution to the zeroth order problem is known, the right hand side of (25) is known. We note that as ϵ tends to zero, (25) is asymptotically equivalent to the result by a one trial function Rayleigh-Ritz method with the mode when stress is not present as the trial function in a variational formulation of the problem [10]. In (25), the frequency shift is linear to $\Delta\beta_{ij}$ and hence linear to stress or strain. Therefore superposition is allowed within the first order perturbation theory. With this we can examine the effect of each individual stress or strain component separately. From the derivation above, it can be seen that (25) is in fact also true for resonators with short circuit boundary or electric wall on which $n_i B_i = 0$ and $\epsilon_{ijk} n_j E_k = 0$, or resonators with partial open circuit boundary and partial short circuit boundary.

For an initially homogeneous and isotropic material, $\epsilon^o_{ij} = \hat{n}^2 \delta_{ij}$ and $\Delta\epsilon_{ij} \approx -\hat{n}^4 \Delta\beta_{ij}$. Hence

$$\frac{\Delta\omega}{\omega^o} \approx \frac{\int_V \epsilon_o \hat{n}^4 E^o_i \Delta\beta_{ij} E^o_j dV}{\int_V (\epsilon_o \hat{n}^2 E^o_i E^o_i + \mu_o H^o_i H^o_i) dV} \tag{26}$$

IV. Modes of a Rectangular Resonator

For a rectangular dielectric resonator with magnetic wall, there exist two types of modes[6]

TE_{lmn} mode

$$\begin{aligned}
E_x =& -\omega^o \mu_o \frac{m}{2b\pi} \\
&\sin\frac{l\pi}{2a}(x-a)\cos\frac{m\pi}{2b}(y-b)\cos\frac{n\pi}{2c}(z-c) \\
E_y =& \omega^o \mu_o \frac{l}{2a\pi} \\
&\cos\frac{l\pi}{2a}(x-a)\sin\frac{m\pi}{2b}(y-b)\cos\frac{n\pi}{2c}(z-c)
\end{aligned}$$

$$E_z = 0$$

$$H_x = \frac{ln}{4ac} \cos\frac{l\pi}{2a}(x-a)\sin\frac{m\pi}{2b}(y-b)\sin\frac{n\pi}{2c}(z-c)$$

$$H_y = \frac{mn}{4bc} \sin\frac{l\pi}{2a}(x-a)\cos\frac{m\pi}{2b}(y-b)\sin\frac{n\pi}{2c}(z-c)$$

$$H_z = -[(\frac{l}{2a})^2 + (\frac{m}{2b})^2] \sin\frac{l\pi}{2a}(x-a)\sin\frac{m\pi}{2b}(y-b)\cos\frac{n\pi}{2c}(z-c)$$

$$(\omega^o)^2\mu_o\epsilon_o\hat{n}^2 = (\frac{l\pi}{2a})^2 + (\frac{m\pi}{2b})^2 + (\frac{n\pi}{2c})^2 \quad (27)$$

TM_{lmn} mode

$$E_x = \frac{ln}{4ac} \sin\frac{l\pi}{2a}(x-a)\cos\frac{m\pi}{2b}(y-b)\cos\frac{n\pi}{2c}(z-c)$$

$$E_y = \frac{mn}{4bc} \cos\frac{l\pi}{2a}(x-a)\sin\frac{m\pi}{2b}(y-b)\cos\frac{n\pi}{2c}(z-c)$$

$$E_z = -[(\frac{l}{2a})^2 + (\frac{m}{2b})^2] \cos\frac{l\pi}{2a}(x-a)\cos\frac{m\pi}{2b}(y-b)\sin\frac{n\pi}{2c}(z-c)$$

$$H_x = \omega^o\epsilon_o\hat{n}^2\frac{m}{2b\pi} \cos\frac{l\pi}{2a}(x-a)\sin\frac{m\pi}{2b}(y-b)\sin\frac{n\pi}{2c}(z-c)$$

$$H_y = -\omega^o\epsilon_o\hat{n}^2\frac{l}{2a\pi} \sin\frac{l\pi}{2a}(x-a)\cos\frac{m\pi}{2b}(y-b)\sin\frac{n\pi}{2c}(z-c)$$

$$H_z = 0$$

$$(\omega^o)^2\mu_o\epsilon_o\hat{n}^2 = (\frac{l\pi}{2a})^2 + (\frac{m\pi}{2b})^2 + (\frac{n\pi}{2c})^2 \quad (28)$$

The above solution is the zeroth order solution when there are no mechanical fields and will be used in the next section to calculate frequency shift, or stress sensitivity.

In the following, we will calculate the relative frequency shift

$$\frac{\Delta f}{f_o} = \frac{f - f_o}{f_o} = \frac{\omega - \omega^o}{\omega^o} = \frac{\Delta\omega}{\omega^o} \quad (29)$$

where $f = \omega/2\pi$ and $f_o = \omega_o/2\pi$ are the frequencies of resonance of the disk with and without initial strains, respectively.

V. Stress Sensitivity

The effect of initial stresses on the frequency shift will be studied for three cases of loading on the disk: (1) unidirectional compression, (2) steady vertical acceleration, and (3) steady horizontal acceleration. In the last two cases, the bottom face of the disk is bounded to a rigid base for support.

For numerical computations, we consider a square dielectric resonators of gallium arsenide (GaAs) for which the material properties are given below.[11]

$$\hat{n}^2 = 12.5\ , \quad \rho = 5.3 \times 10^3 kg/m^3$$

$$c_{11} = 119\ , \quad c_{12} = 53.8$$

$$c_{44} = 59.4 \times 10^9 N/m^2 \quad (30)$$

$$p_{11} = -0.165\ , \quad p_{12} = -0.14$$

$$p_{44} = -0.072$$

(1) Under unidirectional compression

This case reveals the effect of T_{11} or T_{33}. The stress distribution is assumed to be

$$T_{11} = -\frac{F}{4ab}\sin^2\psi$$
$$T_{33} = -\frac{F}{4ab}\cos^2\psi \quad (31)$$
$$T_{31} = -\frac{F}{4ab}\cos\psi\sin\psi$$
$$\text{all other}\quad T_{ij} = 0$$

where $F/4ab = 1MPa$ (or $10^6 N/m^2$) and then the strain field is obtained from the anisotropic stress-strain relations.

For a square resonator of length and width $2a = 2c = 15mm$, thickness $2b = 6mm$ subjected to unidirectional compression force $F = 9.173kg$ which gives an average compression stress $F/4ab = 1MPa$ (or $10^6 N/m^2$), the frequency shift $\Delta f/f_o$ are computed as a function of the azimuth angle ψ of the force F for TE_{110}, TE_{111}, TM_{011}, TM_{101} and TM_{111} modes as shown in Fig. 2. We note that the curves for TE_{110} and TE_{111} modes coincide.

Maximum frequency shift as a function of the

length-to-thickness ratio a/b are computed for the same constant average stress $F/4ab = 1MPa$ and for the same two TE modes and three TM modes as shown in Fig. 3. We note that the curves for TE_{110}, TE_{111}, and TM_{011} modes coincide.

(2) Under Vertical Acceleration

This case reveals the effect of T_{22}.

When the disk is subjected to a steady acceleration of magnitude G in the x_2 direction and its bottom face is bounded to a rigid base. Neglecting the small variations at the edge, the stresses in the resonator are[2]

$$T_{22} = -\rho G(b - x_2)$$
$$\text{all other} \quad T_{ij} = 0 \tag{32}$$

For the same square resonator $2a = 2b = 15mm$, thickness $2b = 6mm$, and $G = 1g$, where $g = 9.81m/sec^2$, $\Delta f/f_o$ vs. a/b curves are computed for the same modes and the results are plotted in Fig. 4. We note again that the curves for TE_{110}, TE_{111}, and TM_{011} modes coincide.

(3) Under Horizontal Acceleration

This case reveals a combined effect of T_{21} and T_{23}.

We now consider the resonator subject to a horizontal acceleration G which makes an angle ψ with the x_3 axis. Neglecting the small variations at the edge, we have the following stress field[2]

$$T_{23} = -\rho G(b - x_2)cos\psi$$
$$T_{21} = -\rho G(b - x_2)sin\psi \tag{33}$$
$$T_{11} = T_{22} = T_{33} = T_{13} = 0$$

In a similar manner, for $G = 1g$, frequency shift of the same TE and TM modes are calculated for various values of ψ and a/b.

It is found that in this case the frequency shift of TE modes are always zero. This is because the nonvanishing stress components T_{21} and T_{23} only contribute to $\Delta\beta_{21}$ and $\Delta\beta_{23}$ for gallium arsenide and $\Delta\beta_{21}$ and $\Delta\beta_{23}$ are multiplied in (26) by the vanishing E_2^o $(= E_z^o)$ component of the electric field.

The frequency shift of TM modes are found to be negligibly small $(\frac{f-f_o}{f_o}10^{10} \approx 10^{-8})$ and are not plotted.

References

1. W. P. MASON, *Crystal Physics of Interaction Process*, Academic Press, New York, pp. 165-169, 1966.
2. P. C. Y. LEE, J. S. YANG AND A. BALLATO, "Stress sensitivity of resonances of TE modes in circular disk dielectric resonators," *Proc. of the Symposium on Electromagneto-Elastic Materials and Structures, the First ASCE-ASME-SES Joint Mechanics Meeting*, Charlottesville, VA, June 1993.
3. P. C. Y. LEE AND J. S. YANG, "Two-dimensional equations for guided EM waves in dielectric plates surrounded by a vacuum," *J. Appl. Phys.*, (accepted for publication).
4. P. C. Y. LEE AND J. S. YANG, "Vibrations of circular disk dielectric resonators," *J. Appl. Phys.*, (accepted for publication).
5. P. C. Y. LEE AND J.-D. YU, "Stress sensitivity of guided EM waves in anisotropic dielectric plates," *Proc. of the 1993 IEEE Frequency Control Symposium*, Salt Lake City, June, 1993.
6. H.-Y. YEE, *An Investigation of Microwave Dielectric Resonators*, Internal Memorandum, Microwave Laboratory, W. W. Hansen Laboratories of Physics, Stanford University, Stanford, California, 1963.
7. A. C. ERINGEN AND G. A. MAUGIN, *Electrodynamics of Continua*, Vol. I, pp. 287-288, Springer-Verlag, New York Inc., 1990.
8. E. DIEULESAINT AND D. ROYER, *Elastic Waves in Solids*, John Wiley and Sons, New York, pp. 383-389, 1980.
9. H. LANDOLT AND R. BÖRNSTEIN, *Landolt-Börnstein New Series*, Group III, Vol. 1, K. H. Hellwege Ed., Spring-Verlag, Berlin, 1966.
10. A. D. BERK, , "Variational principles for electromagnetic resonators and waveguides," *IRE Trans. on Antennas and Propagation*, AP-4, 104-111, 1956.
11. J. F. ROSENBAUM, *Bulk Acoustic Wave, Theory and Devices*, Artech House, Norwood, MA, pp. 451-453, 1988.

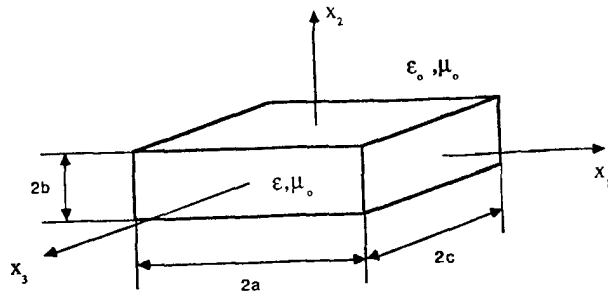


Fig. 1 A rectangular dielectric resonator.

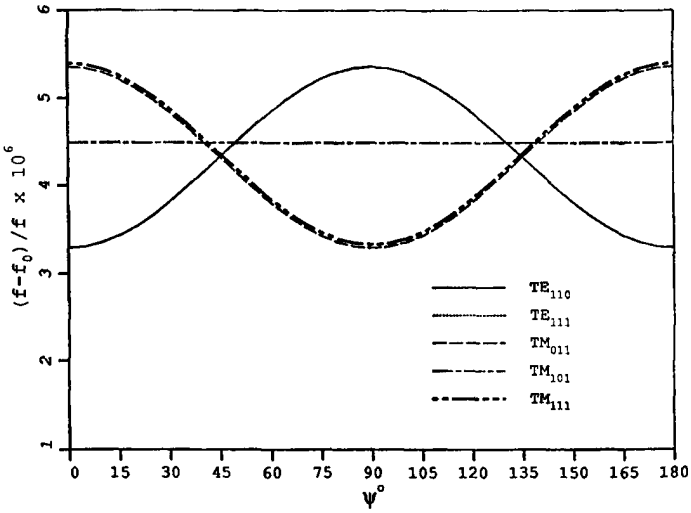


Fig. 2 Frequency shift as a function of the azimuth angle ψ for TE_{110}, TE_{111}, TM_{011}, TM_{101} and TM_{111} modes in a gallium asenide square resonator with $F/4ab = 1MPa$ and $a/b = 2.5$.

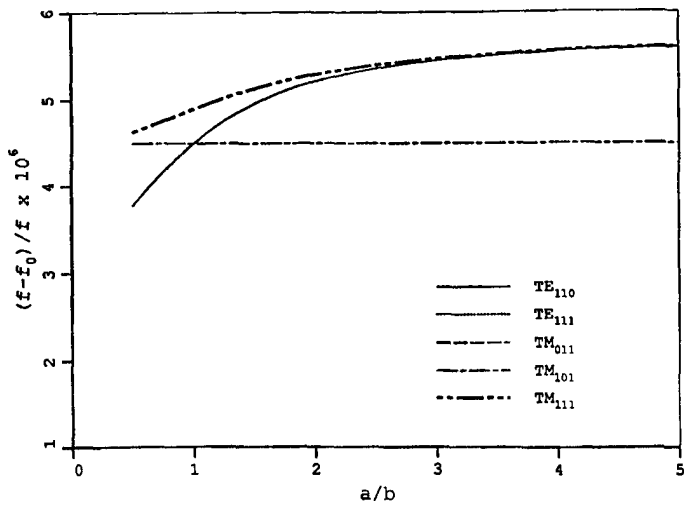


Fig. 3 Maximum frequency shift as a function of a/b for TE_{110}, TE_{111}, TM_{011}, TM_{101} and TM_{111} modes and for $\frac{F}{4ab} = 1MPa$.

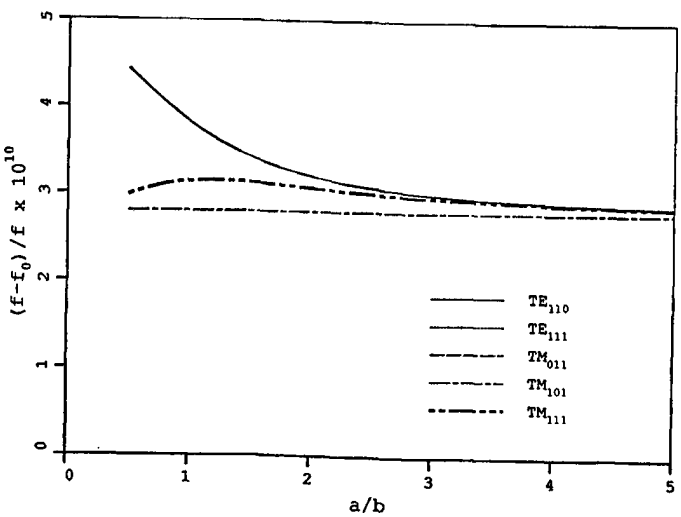


Fig. 4 $|\Delta f/f_o|$ vs. a/b for TE_{110}, TE_{111}, TM_{011}, TM_{101}, and TM_{111} modes and for vertical acceleration of $G = 1g$.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

2D FINITE ELEMENT PROGRAM TO COMPUTE RESONANT FREQUENCIES AND 3D VISUALIZATION OF ELECTROMAGNETIC MODES IN CYLINDRICAL RESONATORS

ARTURO REVILLA, ROBERTO A. OSEGUEDA, JOSEPH H. PIERLUISSI and GUSTAVO J. VILLALVA
The University of Texas at El Paso. College of Engineering.
El Paso, TX. 79968.
and
G. JOHN DICK
Jet Propulsion Laboratory.
California Institute of Technology.
Pasadena, CA. 91009

Abstract - A two-dimensional finite element program called CYRES to compute the electromagnetic resonant frequencies and corresponding modes of cylindrical cavity resonators loaded with anisotropic dielectric is presented. Half of the longitudinal plane of the resonator is analyzed with a two-dimensional mesh assuming a sinusoidal dependence in azimuth. The finite element package consists of PC-based pre- and post-processors, as well as the finite element processor implemented in UNIX. The preprocessor allows for the automatic graphical generation of the finite element mesh. The finite element processor computes the resonant frequencies along with the electromagnetic modes. The postprocessor is used to display the magnetic and electric fields at three different sectional planes of the cylindrical resonator. The display is in the form of color contours representing magnitudes of the field vector, and arrows indicating the three-dimensional orientation of the vector. Visualization is available on longitudinal, transverse and circumferential planes. The numerical results obtained are in excellent agreement with closed form solutions for the empty cavity and coaxial resonator, and with experimental resonant frequency measurements made on a shielded dielectric-loaded sapphire resonator.

I. INTRODUCTION

Although the analytical determination of resonant modes and frequencies of the metallic cylindrical cavity has a well established history, a solution for the cavity partially filled with an anisotropic dielectric generally requires a complex, three-dimensional numerical analysis. Approximate analytical means of analyzing the dielectric resonator have been proposed throughout the years [1,2], and with some degree of accuracy the theoretical estimates have agreed well with experimental results. However, because of the inherent shortcomings of the approximate analytical models, numerical methods have continued to receive a great deal of attention during the past years [3,4].

Recently, so called "whispering gallery" resonators consisting of a sapphire dielectric element in a metallic container have made possible new capabilities for microwave oscillator low phase noise and frequency stability [5,6]. With high azimuthal mode numbers, these resonators isolate radio-frequency energy to the dielectric element and away from the metallic container, thus providing extraordinary low losses and high quality factors (Q's). However, these widely disparate field magnitudes pose a challenge for any methodology to accurately calculate conductive losses due to small evanescent fields at the wall of the containing can. In particular, a three-dimensional finite element method allowing full treatment of sapphire's anisotropic dielectric constant, would require such a large number of nodes as to become impractical. Analytical methods are unattractive, with new approaches required for every geometrical configuration change. A two-dimensional finite element approach, however, would allow easy treatment of any cylindrically symmetric resonator geometry.

Because the dielectric constant for sapphire shows cylindrical symmetry, a two dimensional treatment is allowed for the important case where its crystal c-axis is aligned with a physical axis of symmetry. In terms of the field intensities, the problem is governed by the three-component vector Helmholtz equation which can be treated as an axisymmetric problem only for modes with no azimuthal (or ϕ) dependence. Such zero-order modes can be obtained from a two-dimensional approach to the cavity in the r-z plane using a variety of techniques which yield reasonable accuracy. Higher order solution for isotropic dielectrics are still obtainable in two dimensions if the azimuthal dependence of the modes is assumed ***a priori*** [7]. In the work presented here, the authors reduce the finite element analysis of the anisotropic dielectric resonator to two dimensions by assuming an exponential ϕ-dependence, and limiting the permittivity tensor to possess longitudinal and transverse components only. While ruling out most anisotropic dielectric configurations, this approach makes possible the first two-dimensional finite element treatment for sapphire "whispering gallery" resonators.

0-7803-1945-1/94 $3.00 © 1994 IEEE

II. FINITE ELEMENT FORMULATION.

The problem is governed by the Helmholtz equation, which in terms of the magnetic field intensity $\boldsymbol{H}$ and with the use of a penalty function, it is given by [8]

$$\nabla\times[k]^{-1}\nabla\times\boldsymbol{H} - \alpha\nabla(\nabla\cdot\boldsymbol{H}) - k_o^2\boldsymbol{H} = 0$$

where $[k]$ is the tensor dielectric constant, α is an empirical coefficient of the penalty term $\nabla(\nabla\cdot\boldsymbol{H})$, and k_o is the free-space propagation constant. At the interface between a perfect conductor and a lossless dielectric with a unit normal vector $\mathbf{a}_n$ the boundary conditions are:

$$a_n \times [k]^{-1} (\nabla\times\boldsymbol{H}) = 0$$

$$a_n\cdot\boldsymbol{H} = 0$$

The first condition is naturally satisfied while the second needs to be enforced.

Inside the volume of the cylindrical resonator the magnetic field vector may be described as

$$\{ \boldsymbol{H}(r,\phi,z) \} = \{ \boldsymbol{H}(r,z) \} e^{jn\phi}$$

This finite element formulation considers the use of general ring elements to solve for the magnetic field vectors. These elements are defined in the r-z plane and have m nodes. Within each finite element, $\boldsymbol{H}(r,z)$ is approximated in terms of the standard shape function matrix $[N]$ as

$$\{ \boldsymbol{H}(r,\phi,z) \} = [N(r,z)]^T \{ H \}_e \; e^{jn\phi}$$

Upon substitution of the above equation into the functional of the first equation, and application of the Rayleigh-Ritz criterion on the resulting functional, an eigenvalue equation is obtained as:

$$[S + U]\{ \boldsymbol{H} \} - \lambda [T]\{ \boldsymbol{H} \} = 0$$

where the matrices **S**, **U** and **T** are the global matrices resulting from the superposition of all the finite element matrices [9], and λ is the eigenvalue k_o^2.

III. CYRES SOFTWARE

The CYRES software is a three-part finite element package. The first stage consists of a preprocessor for generating finite element meshes automatically from input information in file or given interactively. The processor uses the finite element mesh data developed in the first stage to solve the eigenvalue equation and obtains a set of eigenvalues or resonant frequencies with their corresponding eigenvectors or electromagnetic modes. The output from the processor is used in the third stage or the postprocessor. The postprocessor uses the mesh generated in the first stage along with the output of the processor to display on the computer screen the electromagnetic modes. Three planes of view are available to visualize the modes of the resonator, these are: the r-z, r-φ, and z-φ planes. Figure 1 depicts a general flow chart of the major components of the CYRES Software.

Preprocessor

The preprocessor generates a finite element mesh from information supplied by the user in the form of a formatted file or as part of an interactive session. The essential information consists of the coordinates of the nodes and the connectivity of four- or eight-node superelements. The program then generates the nodal coordinates and the element connectivities of the mesh to a desired resolution. Error checking is accomplished through a graphical display of the mesh on the computer screen. The preprocessor is capable of generating quadrilateral elements with four nodes and quadrilateral elements with eight node and curved boundaries. This capability allows for the discretization of curved surfaces. The program allows for the graphical definition of the locations of sections for the planes of view to be used in the postprocessor. This allows the user to define from where in the mesh the r-φ and z-φ planes will be displayed. The code is written in **C** language and at present the graphical interface version is available for PC compatible only.

Processor

The processor is written in FORTRAN 77 and has been tested on UNIX workstations, as well as the CRAY Y-MP2E/116 at JPL/Caltech. The processor uses the data file generated by the preprocessor to compute the element matrices, assemble the global matrices, solve the eigenvalue equation to give a set of resonant frequencies with their associated eigenvectors. Upon successful execution, the processor generates three output files, one containing an echo of the input data with the resulting

resonant frequencies and Q factors of each mode. The second file contains the eigenvalues as well as the eigenvectors for each quasi transverse magnetic mode. And the third file contains the eigenvalues and eigenvectors for the curl of H for each mode. Figure 2 illustrates a flow chart of the processor stage of CYRES.

The azimuthal mode number for the run is entered by the user as a range, thus allowing for multiple solutions for several values of the azimuthal mode number in a single run. For automated runs, a simple shell script can be written and the program will not require additional input from the user.

Postprocessor

The postprocessor is written in C language and it runs on the PC compatible with either an EGA or VGA display. The postprocessor uses the mesh data file generated by the preprocessor as well as the magnetic field file and the file containing the curl of H to display a three-dimensional picture of the mode inside the resonator in the three planes of view. The field vector (electric or magnetic) is displayed in the form of color contours representing the field magnitude at a particular point and arrow vectors to display the space orientation of the field. The curl of H file is used to generate the electric field, which is obtained from Maxwell's equation for a source free media, namely:

$$\boldsymbol{E} = \frac{1}{j\omega} [\epsilon]^{-1} \nabla \times \boldsymbol{H}$$

The software is keyboard driven, and several help menus are available. The user is free to select the azimuthal angle and elevation plane, as well as to zoom into a particular plane. The data files have a convenient standard nomenclature which makes them easy to identify and allow the program to display the correct azimuthal number and family mode. The naming convention for the files is as follows:

{mesh name} {mesh density} **n** *{azimuthal order} . {ext}*

For example, the file named *cray4n04.vec* would mean that the mesh used to solve the eigenvalue equation was named *cray*. The density of the mesh was *4* (4 local elements for every superelement) the azimuthal order for this solution is *04*, and the extension *vec* signals that this data file contains the magnetic field information. For ease of use, the user needs only to supply the name of the mesh and the density, the program generates the correct file names automatically.

The output of the postprocessor can be captured using any screen-capturing software and saved as graphical files which can later be integrated into a multimedia presentation for each particular resonator. The automatic generation of these graphical files in different graphical formats is currently under development. The value of a multimedia presentation in analyzing a resonator is really not appreciated until one actually sees one. With a multimedia tool, the user can animate sequential families of modes, and observe the effects the increase in the azimuthal mode number has on the electromagnetic modes. Also, animations along azimuth are also possible as well as animations ranging from the bottom to the top of the can, while passing through the sapphire element. Since the software is capable of displaying any point in the cylindrical plane, an infinite number of images can be generated and used to study the behavior of the modes from every conceivable point in the coordinate system.

The postprocessor can display the electromagnetic fields of any cylindrical resonator with an axisymmetric geometry.

A set of black and white displays of the actual output of this program are included with this paper in subsequent pages. This serves to show the value a graphical display has in the study of these resonators.

IV. VERIFICATION OF SOFTWARE RESULTS

The 2D method to solve for the resonant frequencies and electromagnetic modes was tested and validated by solving the classical empty shielded cylindrical cavity and the coaxial cavity, for which there are closed form solutions, and also by comparing the numerical results of the shielded dielectric-loaded cylindrical resonator to actual measurements made at JPL. These comparisons are fully documented in Reference [9].

The test runs for the empty shielded cylindrical cavity included solutions obtained with variable mesh densities ranging from 16 to 100 four node rectangular ring elements. A penalty factor $\alpha = 1$ was assumed to eliminate spurious modes, and the results were compared to the analytical solutions, the error for the resonant frequency in the 100 element mesh converged to about 0.12% for all modes. For the coaxial resonator, a somewhat higher accuracy was found, the mesh used contained 220 eight-node rectangular ring elements, the error in the resonant frequency for all families identified was below 7 parts per million (ppm), and for the fundamental family, the error was less than 1 ppm.

For the analysis of the shielded dielectric loaded cylindrical resonator, the numerical results obtained by CYRES were compared to measurements taken on the resonator at the Jet Propulsion Laboratory's Time and Frequency Standards Group [5,9]. For mesh densities of 15, 54 and 216 eight-node rectangular ring elements, the error in the resonant frequency converged to less the

0.60% for the fundamental family and to less than 1% for all other families identified. Figure 3 illustrates the correlation between the measured and predicted frequencies for different families identified. A more detailed description of the results and errors can be found in [9].

V. VISUALIZATION OF FIELD VECTORS

The visualization of the field vectors is accomplished by displaying color contour plots in three different plane views as mentioned before. Figure 4 shows the three different planes displayed for the magnetic and electric vectors. Figures 4a and 4b are views of the r-z planes at an angle of $\varphi = 0°$. Similar views can be easily obtained at any angle. The horizontal markers (at left and right) in Figures 4a and 4b correspond to the location of the r-φ plane depicted in Figures 4c and 4d. The vertical markers (at top and bottom) define the location of the z-φ plane depicted in Figures 4e and 4f. Similarly, the markers in Figures 4c through 4f identify the r-z plane viewed in Figures 4a and 4b. All these locations of planes can be changed in the postprocessor for a complete visualization of the field vectors. It should also be noted that because of the possibility of modes with high azimuthal order, the contour plots of the z-φ planes only show one cycle of the azimuthal dependency. The number of cycles in the r-φ plane can be adjusted by the user.

VI. ACKNOWLEDGEMENTS

The authors wish to acknowledge the technical comments and discussions provided by Alfred R. Paiz and Lute Maleki from the Jet Propulsion Laboratory during the course of this research.

VII. SUMMARY AND CONCLUSIONS

A two-dimensional finite element program called CYRES to compute the electromagnetic resonant frequencies and corresponding modes of cylindrical cavity resonators loaded with anisotropic dielectric was presented. The CYRES software consists of PC-based pre and postprocessors, as well as the FORTRAN-language finite element processor implemented in the UNIX operating systems. The preprocessor allows for the automatic graphical generation of the finite element mesh. The finite element processor computes the resonant frequencies along with the electromagnetic modes. The postprocessor displays the magnetic and electric fields at three different sectional planes of the cylindrical resonator. The displays are in the form of color contours representing magnitudes of the field vector with arrows indicating the three-dimensional orientation of the vector. The visualization is available on a longitudinal plane (r-z), a transverse (r-φ) and a circumferential plane (z-φ). The numerical results obtained with the CYRES software have been proven to be in excellent agreement with closed form solutions and with experimental resonant frequency measurements.

The visualization of the electromagnetic modes inside the cylindrical resonator is of invaluable help to both the designer and the student of the field. By providing a visual aid, the behavior of the fields becomes clear. The use of multimedia tools in conjunction with the software developed allows for a full study of the resonator. Many aspects, such as the visualization of the degeneration of a family of modes, is only visible and fully comprehendible when presented in an animation format. Through the use of a multimedia tool and a graphical presentation format, the analysis, design and subsequent validation of the design can be done with one software package. The numerical results the software provides are in excellent agreement with closed form solutions to the hollow cavity and the coaxial resonator, as well as with the measurements made on the shielded sapphire resonator. With confidence on the numerical model, the designer can test several resonators without the cost of actually building them. The distribution of the fields as presented by the software may, for example, pinpoint acceptable locations for coupling ports. The animation of the degeneration of the families of modes, may call the designers attention to the possibility of mode crossover, or other undesirable effects. The uses the software provides are perhaps only limited by the user's imagination and creativity.

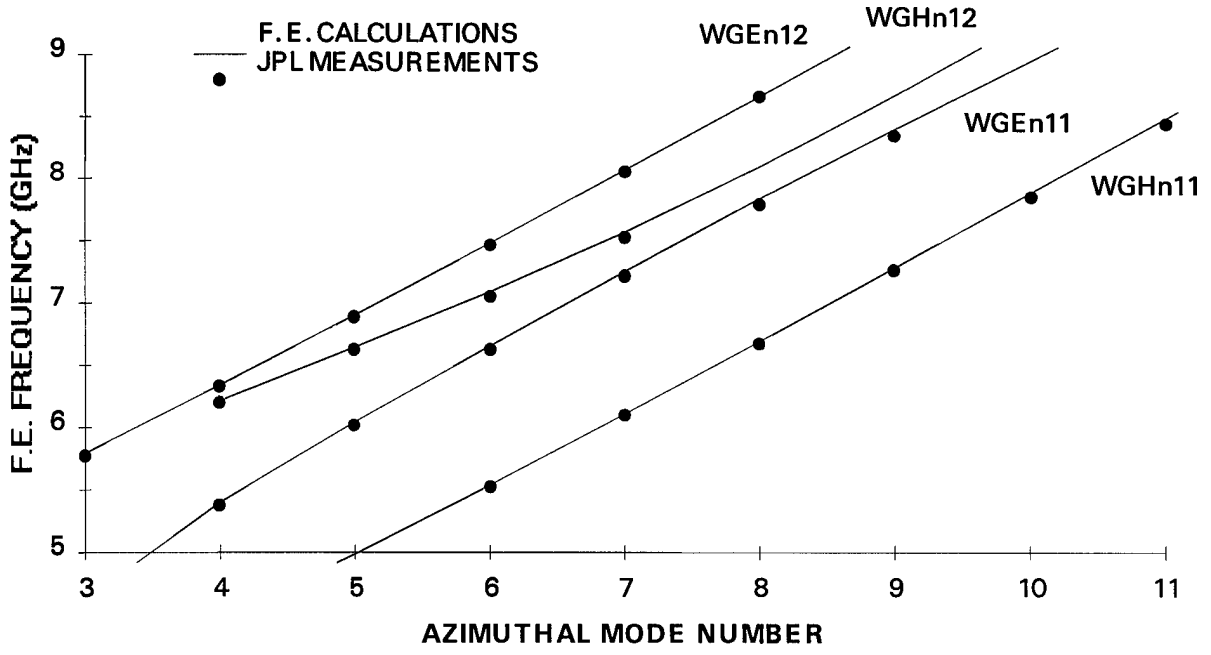


Figure 3: Comparisons of the resonant frequencies obtained from a finite element analysis with the measurements made on a cylindrical sapphire resonator.

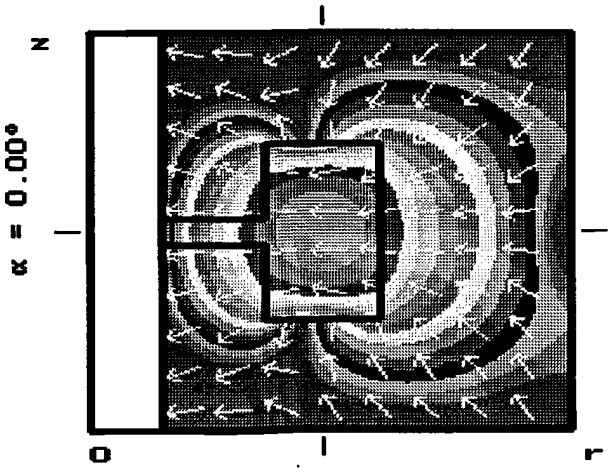


Figure 4.a: Magnetic (H) WGH 411 mode r-z plane.

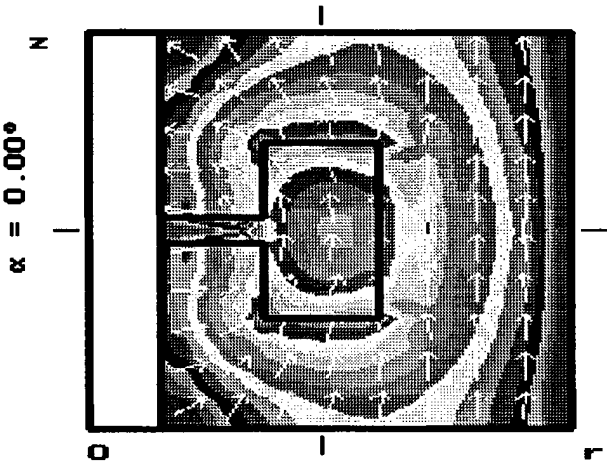
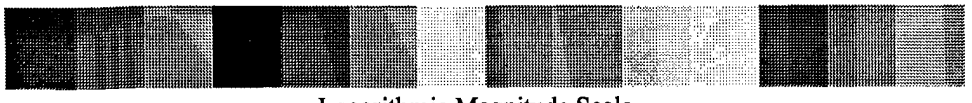


Figure 4.b: Electric (E) WGH 411 mode r-z plane.



Logarithmic Magnitude Scale

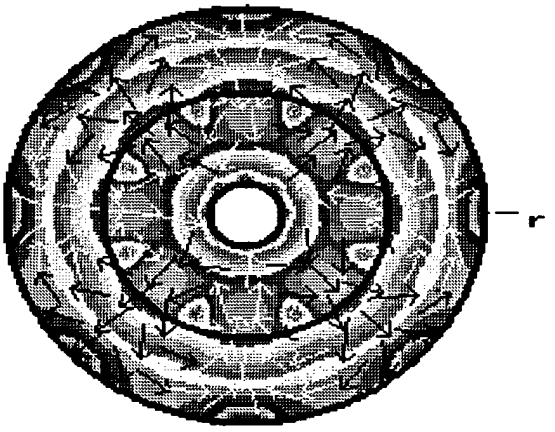


Figure 4.c: Magnetic (H) WGH 411 mode r-ϕ plane.

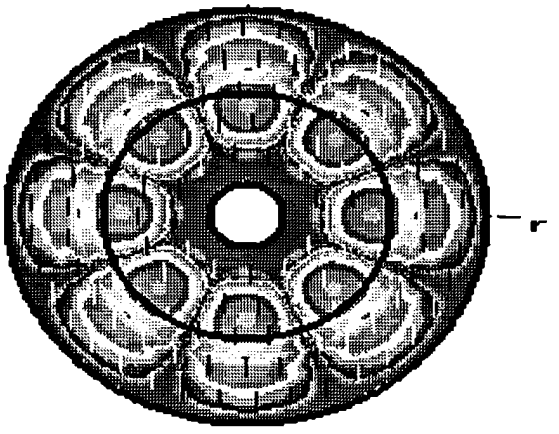


Figure 4.d: Electric (E) WGH 411 mode r-ϕ plane.

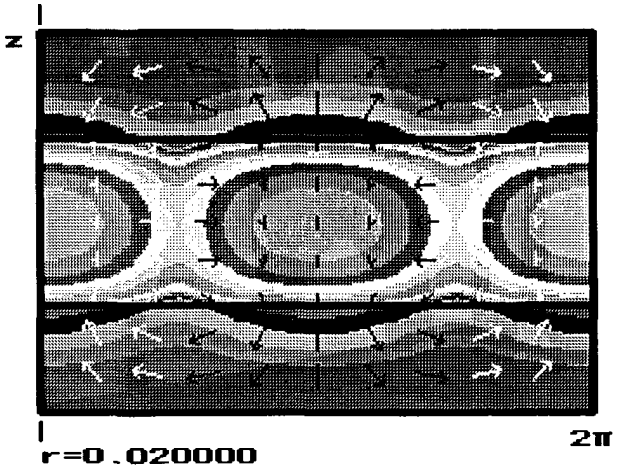


Figure 4.e: Magnetic (H) WGH 411 mode z-ϕ plane.

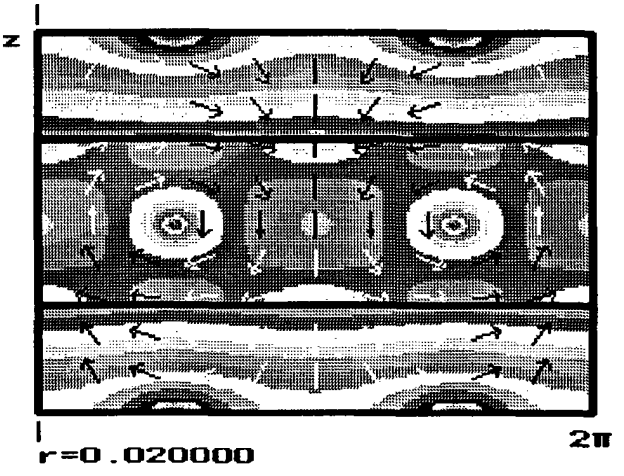


Figure 4.f: Electric (E) WGH 411 mode z-ϕ plane.



Logarithmic Magnitude Scale

VII. REFERENCES

[1]. Kajfez, D. and P. Guillon, Eds., Dielectric Resonators, Artech House, Boston, 1986.

[2]. Tobar, M.E. and A.G. Mann, Resonant Frequencies of Higher-Order Modes in Cylindrical Anisotropic Dielectric Resonators, IEEE MTT-S International Microwave Symposium Digest, pp. 495-498, 1991.

[3]. Konrad, A., Vector Variational Formulation of Electromagnetic Fields in Anisotropic Media, IEEE Trans. Microwave Theory and Techniques, Vol. MTT-24, No. 9, pp. 553-559, 1976.

[4]. Wang, J.S. and N. Ida, Eigenvalue Analysis in Anisotropically Loaded Electromagnetic Cavities Using "Edge" Finite Elements, IEEE Trans. on Magnetics, Vol. 28,, No. 2, pp. 1438-1441, 1992.

[5]. Dick, G.J. and J. Saunders, Measurements and Analysis of a Microwave Oscillator Stabilized by a Sapphire Dielectric Resonator for Ultra-Low Noise, IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 37, No. 5, pp.339-346, 1990.

[6]. Jiao, X., P. Guillon, and L.S. Bermudez, Resonant Frequencies of Whispering-Gallery Dielectric Resonator Modes, IEEE Proc., Vol. 134, Pt. H. No. 6, pp 497-501, 1987.

[7]. Davies, J.B., F.A. Fernandez, and G.Y. Phileppou, Finite Element Analysis of All Modes in Cavities with Circular Symmetry, IEEE Trans. Microwave Theory & Techniques, Vol. MTT-30, No. 11, pp. 1975-1980, 1982.

[8]. Koshiba, M., K. Hayata, and M. Suzuki, Improved Finite Element Formulation in Terms of the Magnetic Field Vector for Dielectric Waveguides, IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-33, No. 3, pp. 227-233, 1985.

[9]. R.A.Osegueda, et. al., Azimuthally-Dependent Finite Element Solution to the Cylindrical Resonator, Proceedings of the 10th Annual Applied Computational Electromagnetics Conference, Vol I. pp 159. March 21-26, 1994

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Sapphire Loaded Microwave Resonators with Enhanced Quality Factor

E. N. Ivanov, V. I. Kalinichev* and D. G. Blair

Department of Physics, The University of Western Australia, Nedlands, WA, 6009
*Radio Engineering Department, Moscow Institute of Energy, Krasnokazarmennaya st., 14

Abstract This paper describes the properties of two-dimensional shielded dielectric ring resonators consisting of a few coaxial dielectric layers surrounded by the coaxial metal shield. It is shown that the Q factor of such resonators excited in whispering-gallery modes increases with azimuthal mode number due to the effective concentration of electromagnetic energy outside the dielectric in the inner region of the resonator as well as in the gap between the dielectric layers. At moderately high azimuthal numbers (m=15...25) the Q-factor can significantly (2 to 3 times) exceed the limit imposed by the intrinsic losses in the dielectric.

I. Introduction

The ultra low loss tangent of sapphire allows the construction of microwave resonators possessing Q factors varying from more than 10^5 at room temperature to $5*10^7$ at 77 K and more than 10^9 at liquid helium temperature [1]. Such resonators have been successfully used in the design of microwave oscillators with the state of art frequency stability and phase noise [2-5]. The unique performance of the sapphire resonators is achieved using electromagnetic whispering gallery modes. These modes can be thought of as electromagnetic waves propagating along the circumference of a dielectric cylinder and experiencing total internal reflection at the surface of the dielectric [4-6].

From the analysis of whispering-gallery modes in the conventional type resonators it follows that the maximum achievable Q-factor is limited by intrinsic losses in dielectric and can not exceed $1/\tan\delta$, where $\tan\delta$ is the loss tangent of the dielectric [7]. This work presents a special class of resonators in which the dielectric has the form of a few coaxial pipes surrounded by a coaxial metal shield. The idea of using the dielectric pipes is to concentrate the electromagnetic field in the internal area of the resonator as well as in the air gap between the dielectric layers where the dielectric losses are absent [8]. The external metal shield eliminates the radiation losses and reduces the overall size of the resonator [7, 9]. A distinctive feature of such a shielded dielectric resonators is that its Q-factor increases with the growth of azimuthal mode number and can significantly exceed the $1/\tan\delta$ level. The predicted Q-factor of the resonator containing two coaxial dielectric layers and operating in the $H_{20,1}$ - mode, is about 2.5 times higher than the dielectric loss limit.

II. Shielded Dielectric Ring Resonators (Basic Equations)

A shielded dielectric ring resonator (SDRR) is a circular dielectric pipe of infinite length surrounded by a coaxial metal shield (Figure 1). We assume that electromagnetic fields are independent on the longitudinal coordinate, and the optical axis of the crystal coincides with the rotation axis of the pipe. The solutions of Maxwell's equations are divided into

0-7803-1945-1/94 $3.00 © 1994 IEEE

two independent families corresponding to E- and H-modes.

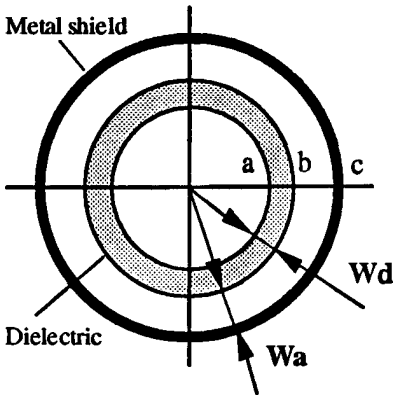


Figure 1. Shielded dielectric ring resonator

For E-modes, the field components E_z, H_ρ, H_φ are non-zero, while for H-modes, the components H_z, E_ρ, E_φ are non-zero. Expressing the field dependence on the azimuthal angle in the form of cos ($m\varphi$), and satisfying the boundary conditions on the surfaces of the dielectric pipe and metal shield, one can obtain the following characteristic equation, which allows the calculation of the resonant frequencies:

$$\Delta_2[A_m J'_m(kb\sqrt{\varepsilon}) + B_m Y'_m(kb\sqrt{\varepsilon})] - \Delta_1[A_m J_m(kb\sqrt{\varepsilon}) + B_m Y_m(kb\sqrt{\varepsilon})] = 0, \qquad (1)$$

where: b is an external radius of dielectric pipe, k is the free-space wave number, m is the azimuthal number, J_m and Y_m are Bessel and Neumann functions of order m, and the primes signify differentiation with respect to the whole argument. For H-modes, the quantities Δ_1, Δ_2, A_m and B_m are given as follows,

$$\Delta_1 = J'_m(kb)Y'_m(kc) - J'_m(kc)Y'_m(kb)$$
$$\Delta_2 = J_m(kb)Y'_m(kc) - J'_m(kc)Y_m(kb) \qquad (2.1)$$
$$A_m = (1/\sqrt{\varepsilon})J_m(ka)Y'_m(ka\sqrt{\varepsilon}) - J'_m(ka)Y_m(ka\sqrt{\varepsilon})$$
$$B_m = J_m(ka\sqrt{\varepsilon})\,J'_m(ka) - (1/\sqrt{\varepsilon})\,J_m(ka)J'_m(ka\sqrt{\varepsilon}),$$

where a is the internal radius of the dielectric pipe and c is the shield radius. For E-modes, the analogous coefficients are given by

$$\Delta_1 = J'_m(kb)Y_m(kc) - J_m(kc)Y'_m(kb)$$
$$\Delta_2 = J_m(kb)Y_m(kc) - J_m(kc)Y_m(kb) \qquad (2.2)$$
$$A_m = \sqrt{\varepsilon}J_m(ka)Y'_m(ka\sqrt{\varepsilon}) - J'_m(ka)Y_m(ka\sqrt{\varepsilon})$$
$$B_m = J_m(ka\sqrt{\varepsilon})\,J'_m(ka) - \sqrt{\varepsilon}J'_m(ka)J'_m(ka\sqrt{\varepsilon}).$$

The relative dielectric permittivity ε in the above equations is assumed to be equal to ε_z for E-modes and ε_r for H-modes, where ε_z and ε_r are the components of sapphire permittivity tensor approximately equal to 11.62 and 9.41, respectively, at room temperature [10].

The first root of the transcendental equation (1) determines the resonant frequency of fundamental mode with one field variation along the radius and m-variations along the azimuth. The second root corresponds to the mode with two variations along radius and m-variations along the azimuth, and so forth.

The Q-factor of the SDRR can be expressed as a combination of partial quality factors Q_m and Q_d :

$$\frac{1}{Q} = \frac{1}{Q_m} + \frac{1}{Q_d} \qquad (3)$$

one of them Q_m is due to the power losses in the metal shield and another one Q_d is due to the dielectric loss only. The Q-factor due to the power loss in the metal shield is given by:

$$Q_m = \frac{2}{\Delta_0}\frac{\int_S H_z^2 ds}{\int_L H_z^2 d\tau} \qquad \text{(H-modes)} \qquad (4.1)$$

$$Q_m = \frac{2}{\Delta_0}\frac{\varepsilon_0}{\mu_0}\frac{\varepsilon\int_S E_z^2 ds}{\int_L H_\varphi^2 d\tau} \qquad \text{(E-modes)} \qquad (4.2)$$

where Δ_0 is the skin-layer depth. One of the integrals in (4) is along the circumference of the shield L, and another one is over the cross-section area of the resonator S. The dielectric loss limited Q-factor is given by,

$$Q_d = \frac{1}{\tan\delta}\frac{\mu_0}{\varepsilon_0}\frac{\int_S H_z^2 ds}{\varepsilon\int_{S2}\left\{E_\rho^2 + E_\varphi^2\right\}ds} \qquad \text{(H-modes)}$$

$$Q_d = \frac{1}{\tan\delta}\left\{1 + \frac{\int_{S1} E_z^2 ds + \int_{S3} E_z^2 ds}{\varepsilon\int_{S2} E_z^2 ds}\right\} \qquad \text{(E-modes)}$$

where the surface integrals are taken either over the internal circle of radius a (S1), or dielectric layer with radii a and b (S2) or external annular layer confined by radii b and c (S3).

III. Properties of Shielded Dielectric Ring Resonators

The results of numerical simulations performed for the H-modes with azimuthal number m = 10 and radial numbers n=1, 2, 3 are presented in Figures 2 and 3. Figure 2 shows the dependencies of mode resonant frequencies on the thickness of the dielectric layer W_d. The shield radius c was chosen to be 56 mm which corresponds to the resonance frequency of the fundamental mode $H_{10,1}$ equal to 10 GHz in the empty resonator (W_d=0).

The results shown in Figure 2 are typical for the system of coupled harmonical oscillators and one can expect the effects of energy exchange between the modes with close resonant frequencies.

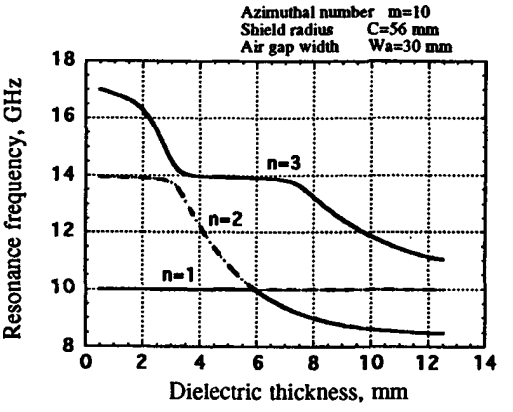


Figure 2. Resonant frequency of the first three radial modes as a function of the dielectric ring thickness.

These effects can be clearly seen from the behaviour of the mode Q-factors shown in Figure 3.

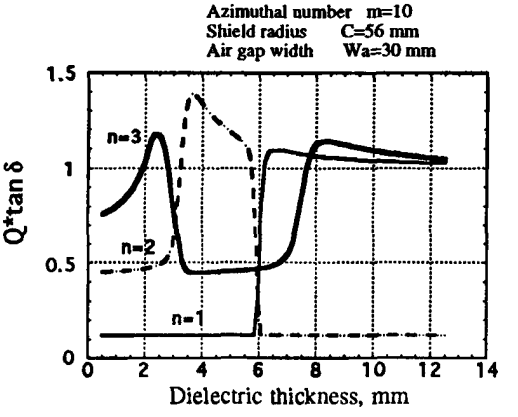


Figure 3. Product Q*tanδ as a function of the dielectric ring thickness for the first three radial modes.

Figure 3 also shows that for each mode there is an interval of W_d where the mode Q-factor is greater than 1/tanδ. In particular, at W_d =3.6 mm the maximum value of the product Q*tanδ at second radial mode is equal to 1.4. The above results have been obtained for a room temperature SDRR with a copper shield. Here the skin depth depends on frequency as: $\Delta_0(mm) = 66/\sqrt{f(Hz)}$ and sapphire loss tangent as $\tan\delta = 5*10^{-16} * f(Hz)$.

By varying the dielectric thickness W_d, the air gap

width W_a, and shield radius c one can maximise the Q-factor of the SDRR for any given mode in the resonator. Figure 4 shows the results of this type of optimisation for the 10 GHz room temperature SDRR excited in fundamental H- and E-modes (n=1).

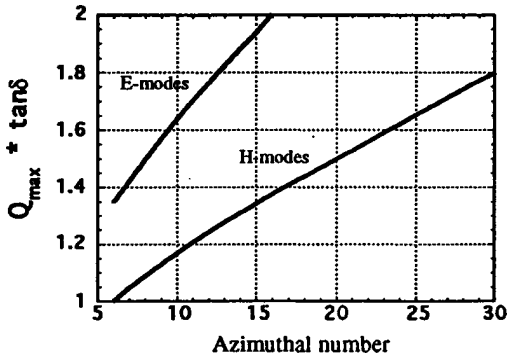


Figure 4. Maximum value of the product Q*tanδ as a function of azimuthal number for two mode families (computer simulations)

For both mode families, the maximum Q-factor increases linearly with azimuth number and at high azimuthal numbers it surpasses the dielectric loss limit. In particular, at m=16 the product Q_{max}*tanδ is equal to 1.4 for H-modes, and 2 for E-modes. However, the higher Q-factors for E-modes are achieved only in larger diameter dielectric rings with reduced wall thickness. This makes such resonator on E-mode less practical compared to a H-mode resonator.

IV. Shielded Multi-Layer Dielectric Structures

The simplest realisation of the shielded multi-layer dielectric resonator (SMDR) is a double dielectric ring structure shown in Figure 5. By varying the gap between the dielectric rings, W_g, and their thickness, W_d, one can find the optimal configuration allowing the maximum of the product Q*tanδ.

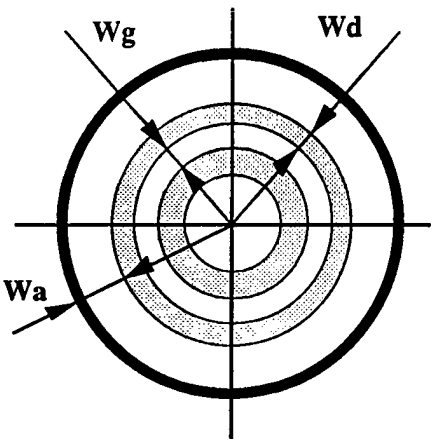


Figure 5. Shielded double dielectric ring resonator

The effect of the gap spacing variations W_g on the quality factor for the room temperature SMDR excited in the fundamental H-mode with m = 16 is shown in Figure 6.

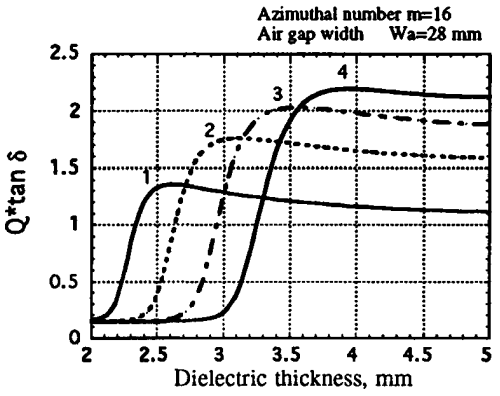


Figure 6. The product Q*tanδ as a function of dielectric ring thickness for different values of the gap between dielectric layers: Wg=0 (curve 1), Wg=0.5 mm (curve 2), Wg=1 mm (curve 3), Wg=1.5 mm (curve 4).

For the distance between the metal shield and external dielectric ring, W_a = 28 mm, the maximum value of Q*tanδ increases from 1.38 to 2.2 when W_g is increased from 0 to 1.5 mm respectively.

A practical advantage of having a large value of the

gap spacing between the dielectric layers is associated with the possibility of increasing the thickness of dielectric rings without sacrificing the Q-factor (see Figure 6). This allows the resonator to be more rigid and less susceptible to vibration.

Figure 7 shows the radial dependence of the H_z component for the $H_{16,1}$ -mode in a SMDR with optimal configuration. It is clear that a significant part of the electromagnetic energy can be stored in the air gap between the dielectric layers.

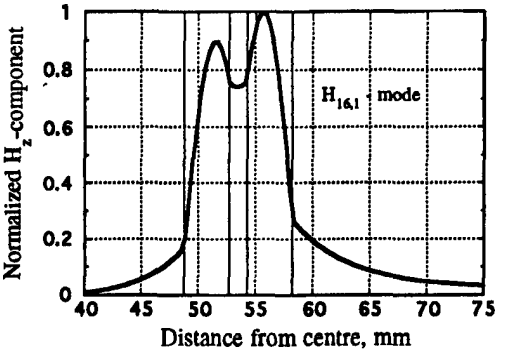


Figure 7. Radial dependence of the normalised z-component of magnetic field for $H_{16,1}$-mode in shielded double ring resonator

In this particular case 15% of the total energy is stored in the gap between the two rings. For H-modes with azimuthal numbers ranging from m = 10 to 25 this value varies from 8% to 21% respectively.

Figure 8 allows the comparison of three different types of room temperature sapphire loaded resonators. Curves 1 and 2 correspond to the single- and the double-ring shielded resonators respectively. Curve 3 is obtained for the double-ring structure surrounded by a corrugated metal shield [11]. Such a shield consists of a number of narrow metal fins placed in the longitudinal direction along the circumference of the resonator. For whispering-gallery modes with large azimuthal number ($m \geq 10$), the losses in the corrugated shield is significantly smaller than in the usual smooth shield provided that the height of the fins and the distance between them are chosen to be equal to $\lambda o/4$, where λo is the free space wavelength [11]. As shown in Figure 8 the combination of a double-ring structure with the corrugated shield achieves a value for $Qmax*\tan\delta$ of the order of 3 at azimuth number between m = 21 and m = 22.

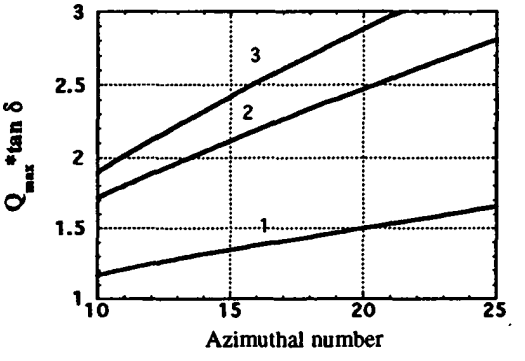


Figure 8. The maximum values of the product $Q*\tan\delta$ versus azimuthal number for three different resonators: shielded dielectric ring resonator (1), shielded double ring resonator (2), double dielectric ring resonator with corrugated shield (3).

V. Conclusion

The numerical analysis of a two dimensional model of shielded dielectric ring resonators shows that Q-factors up to 3 times higher than the intrinsic dielectric loss limit can be achieved in such resonators excited in whispering gallery modes. As far as the frequency stability of the SDRR due to the shield deformation is concerned, it is a bit worse than in the conventional sapphire loaded resonators but much better than in the hollow metal cavities. At room temperature, for instance, the variations of resonant frequency caused by the shield deformation are approximately 5000 times smaller in the SDRR than in the hollow cavity

provided that both resonators are excited in the same mode ($H_{16,1}$), and SDRR is optimised for maximum Q-factor.

Some prototypes of the shielded dielectric ring resonator are currently being tested in the Gravitational Radiation Laboratory at the University of Western Australia. The dielectric rings are made of alumina (polycrystalline Al_2O_3) which will allow us a relatively cheap way of testing the whole concept of the dielectric loaded resonators with enhanced quality factor.

Acknowledgments

The authors would like to thank Dr. M. E. Tobar and Mr. P. J. Turner of the University of Western Australia for helpful advice and assistance in preparing of this paper.

References

1. D. G. Blair and S. K. Jones, "High Q sapphire loaded superconducting cavities and application to ultra stable clocks", IEEE Trans. Magn., Magn-21, pp. 142-143, March 1985.

2. 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", in Proc. 43rd Annual Symposium of Frequency Control, pp 89-93, 1989.

3. M. E. Tobar and D. G. Blair,"Phase noise analysis of the sapphire loaded superconducting niobium cavity oscillator", IEEE MTT, V.42, N2, 1994.

4. J. R. Wait,"Electromagnetic whispering gallery modes in a dielectric rod", Radio Science, V. 2, N9, pp. 1005-1017, 1967.

5. X. Jiao, P. Guillon and L. A. Bermudez,"Resonant frequencies of sapphire dielectric resonator modes", Proc. IEE, V. 134, pp.497-501, 1987

6. J. Krupka, D. Cros, M. Aubourg and P. Guillon,"Study of whispering gallery modes in anisotropic single-crystal dielectric resonators", IEEE MTT, V. 42, N1, pp. 56-61, 1994.

7. Vzyatyshev V F, Dobromyslov V. S., Kalinichev V I and Kuimov V I, "Shielded Dielectric Resonators with Azimuthal Modes", *Radiophysica,* 30, No. 1, pp 79-87, 1987.

8. Kazancev J N, Manenkov A B, Kcharlashkin O A, *Radiophysica*, 17, 10, pp. 1529, 1974.

9. E. N. Ivanov, V. I. Kalinichev and D. G. Blair,"Approximate approach to the design of shielded dielectric disk resonators with whispering-gallery modes", IEEE MTT, V. 41, N4, pp. 632-638, 1993.

10. R. Shelby and J. Fontanella,"The low temperature electrical properties of some anisotropic crystals", J. Phys. Chem. Solids, V. 41, pp. 69-74, 1980.

11. G. Y. Slepijan and A. S. El'insky, "Oscillations and Waves in Electrodynamic Systems with Losses", Izd-vo Moscow State University, 1983.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

FUNDAMENTAL LIMITS ON THE FREQUENCY INSTABILITIES OF QUARTZ CRYSTAL OSCILLATORS

John R. Vig* and Fred L. Walls**
*U. S. Army Research Laboratory, Ft. Monmouth, NJ 07703
**National Institute of Standards and Technology, Boulder, CO 80303

Abstract

The frequency instabilities of precision bulk acoustic wave (BAW) quartz crystal oscillators, and the contributions from the instabilities of resonators, sustaining circuits, and ovens, are reviewed. The fundamental limits on the achievable frequency stabilities, and the degree to which the fundamental limits have been approached to date are examined. Included are the instabilities as a function of time, temperature, acceleration, ionizing radiation, electromagnetic fields, humidity, atmospheric pressure, power supply and load impedance.

Most of the fundamental limits are zero or negligibly small, a few are finite. The authors speculate about the progress which may be achievable in the future with respect to approaching the fundamental limits. Suggestions are provided about the paths that may lead to significant stability improvements.

Introduction

Timekeeping accuracy has improved about 10 orders of magnitude during the past 600 years [1]. The last 10^6-fold improvement occurred during the first sixty years after the introduction of quartz crystal oscillators (about 1920). During the last two decades, however, progress in certain areas, e.g., in long term aging [2,3], has not shown significant improvement. Does this mean that we have reached some fundamental limit and, therefore, further improvements are no longer possible?

The task assigned to the authors of this (invited) paper was to examine what the fundamental limits are, and the degree to which further progress in stability improvements will be limited by the fundamental limits. The discussions will concentrate on oven controlled crystal oscillators (OCXOs), as OCXOs are the highest stability crystal oscillators.

U.S. Government Work Not Subject to U.S. Copyright

Oscillator Instabilities - General Discussion

Figure 1 shows a block diagram of the essential components of a precision crystal oscillator. Each of the three main parts of the oscillator, i.e., the crystal resonator, the sustaining circuit, and the oven, contribute to instabilities. The general expression for the oscillator's instabilities is [4]:

$$\frac{\Delta\nu}{\nu_{osc}} \approx \frac{\Delta\nu}{\nu_{res}} + \left[1 + \left(\frac{2fQ_L}{\nu_o}\right)^2\right]^{-\frac{1}{2}} \frac{d\varphi(f)}{2Q_L}, \qquad (1)$$

where Q_L = loaded Q-factor of the resonator, and $d\phi(f)$ is a small change in loop phase at offset frequency f away from carrier frequency ν_o. Phase changes and phase noise within the loop can originate in either the resonator or the sustaining electronics. Maximizing Q_L helps reduce the effects of noise and environmental effects in the sustaining electronics. In a properly designed oscillator the short-term frequency stability or phase modulation (PM) noise is controlled by the resonator for Fourier frequencies smaller than the half bandwidth of the resonator. For frequencies larger than the half bandwidth of the resonator, the phase noise is determined by the sustaining circuit and the oscillation power (i.e., the amount of power

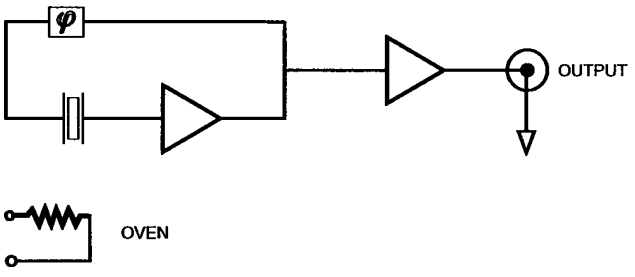


Fig. 1 Simplified block diagram of a quartz oscillator. Loop gain = 1 and loop phase = $2n\pi$, with n = 1, 1, 2, 3.. [4].

delivered from the oscillation loop). More detailed discussions are found in [4-7].

The main influences on the stability of an oscillator can be categorized as follows [4-7]:

- Time: short-term (phase noise), intermediate-term (e.g., due to oven fluctuations), and aging (i.e., long-term stability)

- Temperature: static frequency vs. temperature characteristics, dynamic frequency vs. temperature effects (warmup, thermal shock), and thermal history dependence ("retrace")

- Acceleration: gravity (2g tipover), vibration, acoustic noise, and shock

- Ionizing radiation: photons (X-rays, γ-rays), and particles (neutrons, protons, electrons, α-particles); and pulsed and steady state effects

- Other: sustaining electronics power supply voltage, load impedance, atmospheric pressure (altitude), humidity, and electromagnetic fields

The fundamental and some practical limits to frequency stability due to each of these influences are discussed below [4-7].

Short-Term Instabilities

The causes of short-term instabilities of fundamental nature are:

- Johnson noise in the motional resistance of the resonator and within the oscillator circuitry [8].

- Phonon scattering noise in the resonator [9]

Causes of frequency instability which are technical rather than fundamental in nature are [4]:

- Changes in external load

- Oven temperature fluctuations

- Thermal response of resonator - static and dynamic,

- Random vibration [10],

- Fluctuations in the number of adsorbed molecules [11],

- Stress relief and fluctuations at interfaces (between the quartz plate and its electrodes, and its mounting and bonding structure) [4,7].

Thermal noise (Johnson noise) due to the resonator's motional resistance creates white FM noise [8]:

$$\mathcal{L}(f) = \frac{kT}{2Q_L^2 P f^2}, \qquad (2)$$

and

$$\sigma_y(\tau) = \frac{1}{Q_L}\sqrt{\frac{kT}{2P\tau}}, \qquad (3)$$

where f is the frequency offset from the carrier frequency, k is Boltzmann's constant, T is the temperature in K, Q_L is the loaded Q of the resonator in the circuit, P is the total power dissipated in the resonator plus load, and τ is the measurement time. For $Q_L = 10^6$, $P = 10^{-5}$ W, and T = 330K, $\sigma_y(\tau) = 1.4 \times 10^{-14}\tau^{-1/2}$.

Phonon Scattering in the Resonator

Quartz bulk acoustic wave (BAW) resonators work by transferring electrical power into acoustic power and back into electrical power. The transfer function is characterized by many acoustic modes, some of which have high Q-factors. Scattering of the phonons within the crystal lattice due to defects and scattering at the crystal-electrode or in the case of BVA resonators [12], crystal-vacuum interface, dissipates energy. This lowers the Q-factor, and there is theoretical and some preliminary experimental evidence that it also increases the 1/f frequency modulation (FM) noise in the resonator. Data from many sources show (see for example [5, 9, 13]) that the best flicker level obtained from a given resonator design generally scales as $1/Q^4$, however, there other processes that cause the flicker FM level to vary between resonators of the same design and same Q-factor [5, 9, 13]. Data from [14] shows that the phase noise or short-term frequency stability can exhibit flicker FM from about 0.02 s to beyond 10^6 s., indicating that this is a fundamental process within the resonator and not an accidental coincidence of several types of noise processes to mimic flicker FM.

At present this process sets the limit for short- and medium-term frequency stability in all quartz resonators. There is beginning to be some evidence that appropriate changes in the electrodes in AT and SC-cut resonators may reduce the flicker FM noise [9,15,16]. If the flicker PM level is reduced, then Johnson noise could become the limiting effect in the short-term stability.

Noise in Sustaining Circuit

Flicker PM noise in the sustaining circuit causes flicker FM contribution to the oscillator output frequency given by:

$$\mathcal{L}_{osc}(f) = \mathcal{L}_{osc}(1Hz)\frac{\nu^2}{4f^3Q_L^2}, \qquad (4)$$

and

$$\sigma_y(\tau) = \frac{1}{Q_L}\sqrt{\ln 2\,\mathcal{L}_{ckt}(1Hz)}, \qquad (5)$$

where f is the frequency offset from the carrier frequency ν, Q_L is the loaded Q of the resonator in the circuit, $\mathcal{L}_{osc}(1Hz)$ is the flicker PM noise at f = 1Hz, and τ is any measurement time in the flicker floor range. For $Q_L = 10^6$ and $\mathcal{L}_{osc}(1Hz)$ = -140 dBc/Hz, $\sigma_y(\tau) = 8.3 \times 10^{-14}$.

To achieve $\sigma_y(\tau)$ lower than this value will require improvements in Q_L-factors or amplifier PM noise performance. Since $Q\nu_o$ ~ a constant for BAW resonators of a given cut, these effects strongly favor oscillators at the lowest practical frequency. The highest resonator Q-factors reported are approximately 5×10^6 for 5th overtone AT-cut 2.5 MHz resonators manufactured some 30 years ago.

BT-cut resonators can have higher Q-factors by approximately a factor of two over AT- or SC-cut resonators. The use of crystal cuts which are not stress compensated, however, lead to serious problems with dynamic temperature fluctuations, amplitude changes, and stress as discussed below. We are not aware, however, of any fundamental work which shows a maximum theoretical value for the Q-factor. All of the evidence is empirical (based on physically realized resonators). Changes in resonator design, material purity, isotopic purity, crystallographic perfection, and/or processing might lead to higher Q-factors.

The lowest flicker PM reported in amplifiers is approximately $\mathcal{L}$(1 Hz) = -152 dBc/Hz [16]. Flicker PM noise depends on the amplifier design as well as on the active element. As transistor fabrication techniques have improved, so has the flicker PM noise. We expect further reductions in amplifier 1/f PM noise of 5 to 10 dB.

Loop Phase, Mode Selection, Matching, and Filter Circuits

Many oscillators contain tuned circuits to suppress unwanted modes, as matching circuits, and as filters. The effects of small changes in the tuned circuit's inductance and capacitance is given by [4,5]

$$\frac{\Delta\nu}{\nu_{osc}} \approx \frac{d\phi(f_f)}{2Q_L} \approx \left(\frac{1}{1+\frac{2f_f}{BW}}\right)\left(\frac{Q_c}{Q}\right)\left(\frac{dC_c}{C_c} + \frac{dL_c}{L_c}\right) \qquad (6)$$

where BW is the bandwidth of the filter, f_f is the frequency offset of the center frequency of the filter from the carrier frequency, Q_L is the loaded Q of the resonator and Q_c, L_c and C_c are the tuned circuit's Q, inductance and capacitance, respectively. Many environmental parameters cause a change in the phase around the oscillator loop. The most important are temperature, humidity, pressure, acceleration and vibration, magnetic field, voltage, load and radiation. These environmental sensitivities often lead to increases in the level of wide-band phase noise in the short-term, random-walk frequency modulation in the medium-term, and drift in the long-term. The loop phase shift depends on the circuit design and the loaded Q-factor of the oscillator. Figure 2 illustrates the effect of humidity on an oscillator with a nominal $\sigma_y(\tau)$ of 3×10^{-13}.

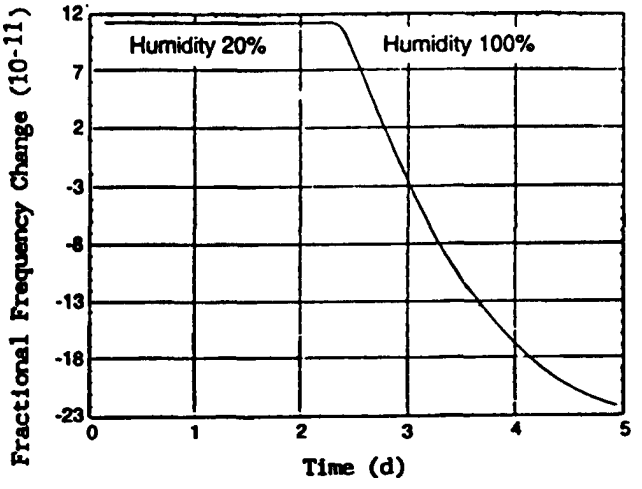


Fig. 2 Fractional frequency change of quartz oscillator due to a change in humidity [4].

Pressure and humidity effects can be eliminated by hermetic sealing in a rigid enclosure.

Equation 6 indicates that notch filters are much less likely to perturb the phase of the carrier than bandpass filters because, in equation 6, $f_f = 0$ for a bandpass filter and the circuit, therefore, sees the full instability of the filter. However, for a notch filter, f_f can be large compared to the BW, and the effect of the instability of the filter on the oscillator is reduced by f_f/BW. All circuit elements, especially reactances in the signal path should be selected with great care. If bandpass filters are used, Q_C should be as small as feasible. Inductors are often noisy and are susceptible to magnetic field disturbances.

Changes in the External Load

If the external load changes, there is a change in the amplitude or phase of the signal reflected back into the oscillator [17]. The portion of that signal which reaches the oscillating loop changes the oscillation phase, and hence the frequency by

$$\frac{\Delta\nu}{\nu_{osc}} \approx \frac{d\phi(f_f)}{2Q} \approx \left(\frac{1}{2Q}\right)\left(\frac{\Gamma-1}{\Gamma+1}\right)(\sin\theta)\sqrt{\text{isolation}} \quad (7)$$

where Γ is the VSWR of the load, and θ is the phase angle of the reflected wave; e.g., if $Q \sim 10^6$, and isolation ~40 dB (i.e., $\sim 10^{-4}$), then the worst case (100% reflection) pulling is $\sim 5 \times 10^{-9}$. A VSWR of 2 reduces the maximum pulling by only a factor of 3. The problem of load pulling becomes worse at higher frequencies, because both the Q and the isolation decrease.

RF Excitation Level

The frequency of the resonator is also a function of amplitude of the signal level (i.e., the drive level) as shown in Figs. 3 and 4 [4, 18-20]. The sensitivity to this effect, usually called the amplitude-frequency effect, is a function of a large number of resonator fabrication details. The frequency change varies as the square of the drive current [19,21]. Figure 4 shows the dependence on blank curvature with most other parameters held constant [20].

Typical sensitivities to this parameter range from approximately $10^{-9}/\mu W$ for 5th overtone AT- or BT-cut resonators to parts in $10^{-11}/\mu W$ for 3rd overtone, SC-cut resonators at 5 MHz. The primary environment drivers for this effect are temperature, humidity, or radiation changing the excitation level,

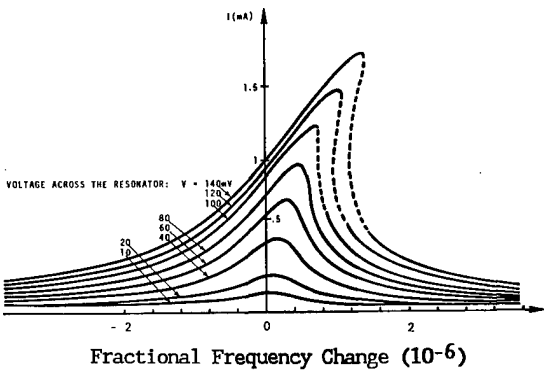


Fig. 3 Non-linear resonance as a function of RF drive for a 5 MHz AT-cut resonator. From [22].

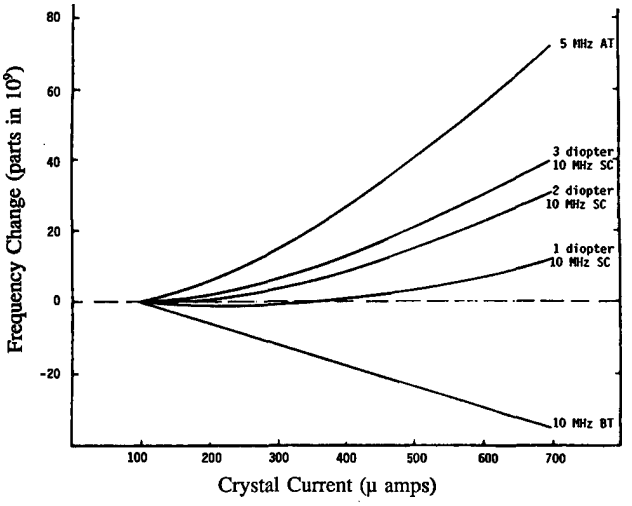


Fig.4 Fractional frequency change versus RF drive level for AT-, BT-, and several SC-cut resonators. From [20].

e.g., through interaction with the automatic gain control (AGC) and the gain of the sustaining stage. Changes in circuit components can also affect the amplitude of oscillation by changing the AGC circuitry.

Stress

Changes in stress on the resonator plate can change the resonance frequency [22-27]. Stress is transmitted to the resonator through the mounting structure or from the plating of the electrodes. It can originate from temperature-driven dimensional variations in the vacuum enclosure, changes in the pressure surrounding the vacuum enclosure, changes in the magnetic field causing a change in the mounting force due to the use of magnetic components, or from changes in the body forces due to acceleration and vibration as described below. Stresses also exist in the quartz due to defects

introduced in the growing process. The stresses on the resonator can change with time, temperature, resonator drive level, and radiation exposure. The stress effects due to the electrodes have been estimated [23-25].

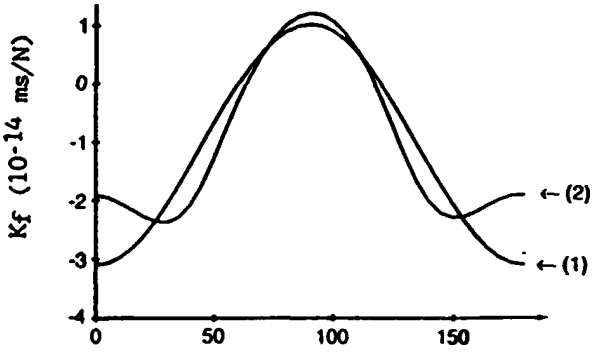


Fig. 5 Force-frequency coefficient for diametrically opposed forces in the plane of an AT-cut resonator as a function of the angle between the force and the azimuthal angle [26].

Figure 5 shows the change in frequency of a traditional AT-cut plate due to diametrically opposed forces in the plane of the resonator as a function of the angle between the applied force and the x-axis [26]. Bending stresses can also produce significant frequency shifts [27], as can stresses due to the bonding of the mounting clips to the resonator plate. The influences due to the mounting stresses can be minimized in traditional resonators by proper choice of the mounting angles and procedures. Mounting stresses in BVA resonators are minimized by both the choice of mounting angle and by the use of a supporting ring machined from the same monolithic piece of quartz as the resonator [12,28]. The effects of electrode stresses are also absent from the BVA electrodeless resonators.

Electromagnetic Fields

The frequency of certain resonator cuts are directly affected by the application of even small electric fields through changes in dimension and effective mass and through interaction with the nonlinear coefficients. The application of electric fields also tends to cause ions within the crystal to move which changes the frequency [29-33]. The result is that the change in frequency generally has a fast component due to piezoelectricity and the interaction with the crystal constants, and one or more slower components associated with the movement of ions, as shown in Figure 6. The shorter time constants depend exponentially on temperature.

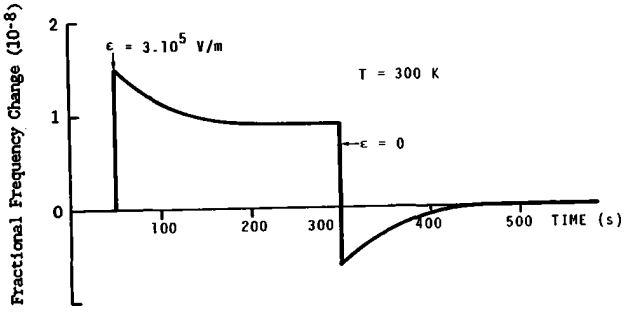


Fig. 6 Fractional frequency change as a function of applied voltage for an AT-cut resonator. The slow variation after the change in voltage is due to the movement of ions in the quartz. From [38].

The aging of resonators can be significantly changed by applying a dc voltage to the electrodes, even at "normal" temperatures [34]. The electric field effect has also been used to vibration compensate SC-cut resonators [35] and to create an ultra-linear phase modulator [29]. The sensitivity to this effect is highly dependent on resonator cut, material, and electrode configuration. Coefficients range from approximately 10^{-11} to 10^{-8}/volt applied across the resonator and scale approximately as the reciprocal of the plate thickness or $1/v_o$. For a resonator to exhibit a sensitivity, the dc electric field must have a component along the x axis of the crystal. AT-cuts with regular electrodes therefore exhibit very little sensitivity to electric fields and doubly rotated cuts, such as the SC-cut, generally show significant sensitivity to dc electric fields. Large electric fields and elevated temperatures are sometimes used to "sweep" ions out of the quartz bar prior to resonator fabrication [4,30-32]. This is typically used on resonators intended for radiation environments (and also, to reduce the etch-channel density of the quartz). Perturbations due to unintentional electric fields are virtually eliminated by placing a dc resistance of ~10^5 Ω across the resonator.

Quartz resonators' inherent magnetic field sensitivity is smaller than 10^{-11}/T [36-38]. Many resonators are, however, constructed with magnetic holders. As the magnetic field changes, the force on the various components of the resonator changes. This causes a frequency shift through the force-frequency coefficient discussed above and the circuit phase shifts discussed below. Generally, the magnetic field effect is unchanged under inversion and maximized for rotation through 90°. This effect is reduced to below 10^{-10}/T by using non-magnetic holders and shielding all inductors within the oscillator [37].

Power Supply Voltage

Instabilities in the oscillator can often be traced to instabilities of the power supply. The frequency changes occur because changes in various voltages change the capacitance of active and some passive devices. The gain of the sustaining or buffer amplifiers can change causing a phase shift. If the phase shift is in the sustaining stage, there is a frequency change given by Eq. 1, whereas if the phase shift is in the buffer amplifier, there is no first order frequency change if the isolation is high enough and a matched load is used. The gain changes can change the amplitude of drive thereby causing a frequency change.

The remedy is to measure the coefficient for frequency and phase changes with voltage in various parts of the circuit and design the voltage regulation to provide regulation that does not compromise the performance. We are not aware of any fundamental limit to oscillator frequency stability due to this effect.

Aging

The main causes of aging [2] are: mass transfer due to contamination, stress relief in the resonator's mounting and bonding structure and in its electrodes, sustaining circuitry aging, and oven control circuitry aging. Another possible aging mechanism is the outgassing of the quartz, which only recently has been investigated [39].

Nearly all precision oscillators use resonators which are enclosed in either metal or glass enclosures. These metals and glasses are known to outgas orders of magnitude more than some other materials, such as the high-alumina ceramics and sapphire [40-43].

Stainless steel has been the material of choice for ultrahigh vacuum (UHV) systems, however, aluminum has been shown to allow better vacuums. A capability of achieving 10^{-11} Pa (10^{-13} torr) has been demonstrated with a vacuum chamber the interior of which was coated with high purity aluminum, and which used aluminum seals [45,46]. The ultimate pressure in such an "extremely high vacuum (XHV)" system is determined by the desorption of hydrogen from the chamber walls [46]. An aluminum alloy vacuum chamber which had been extruded and tempered in an oxygen and argon atmosphere (in order to form a dense oxide layer on the inner surfaces) has also been shown to allow vacuums in the 10^{-11} Pa (10^{-13} torr) range [46].

Aluminum's low outgassing rates and high thermal conductivity may make it a good candidate material for future resonator packages. It may provide a lower cost alternative to sapphire and ceramic packages.

Near atomically clean surfaces can be produced immediately before resonator sealing by a combination of UV/ozone cleaning [47] and high-temperature baking in ultrahigh vacuum. However, if the enclosure and the quartz outgas after the enclosure is sealed, aging will result. It is clear that the current practices for making high-stability resonators leave room for improvement with respect to minimizing the mass transfer due to contamination resulting from the outgassing of the enclosure and of the materials inside the enclosure.

Aging due to stress relief can be minimized by "electrodeless" designs [12,28,48], and by a combination of an electrodeless design and levitation. (The possibility of using levitation is discussed below.)

The effects of sustaining circuitry component aging can be minimized by using: 1) maximum loaded Q-factor, 2) properly selected components, 3) resonator designs which provide a small motional capacitance, e.g., a fifth overtone SC-cut, and, 4) a sufficiently large load capacitance.

The effects of oven control circuitry aging can be minimized by using: 1) properly selected components and design, 2) an SC-cut resonator the turnover points of which coincide with the inflection point, 3) a temperature sensitive mode (e.g., the B-mode [49] or a combination of the fundamental and third overtone [50]), at least for setting the oven temperature to the turnover temperature, if not for oven control, and 4) an oven temperature set point which coincides with the turnover temperatures of the resonator. As is shown below, when conditions 2 and 4 are satisfied, small changes in the oven temperature result in negligible frequency shifts.

Although further progress in reducing the known aging mechanisms present serious practical problems, none of them appear to present a fundamental limit. With sufficient effort, significant improvements ought to be possible in the future. For example, in the "ideal" resonator described below, all the known aging mechanisms would be virtually absent.

Frequency vs. Temperature Stability

The frequency vs. temperature (f vs. T) stability of an OCXO depends on: the static and dynamic f vs. T characteristics of the resonator in the sustaining circuit, the difference between the oven set point and the point where the static f vs. T characteristic has zero slope, the oven's temperature excursions from the set point, and the rate of change of temperature during the oven's temperature excursions [4,51].

Dynamic f vs. T Stability

The dynamic f vs. T stability can dominate the static f vs. T stability when the resonator is not thermal transient compensated (i.e., when it is not an SC-cut). For maximum stability, a thermal transient compensated resonator must be used [52]. Although many so-called SC-cut resonators exhibit a finite thermal transient effect, in the vicinity of the "true" thermal transient compensated cut, small changes in the angles of cut can produce either positive or negative thermal transient effects. The thermal transient effect is exactly zero at one set of angles of cut in the vicinity of $\theta = 34°$ and $\phi = 22°$. The exact values of the "true" SC-cut angles vary with resonator configuration. The effect can also be influenced by factors such as transient effects in the oven control and sustaining circuit components [52].

Although making an oscillator which exhibits zero thermal transient effect presents some non-trivial practical problems, the fundamental limit to the thermal transient effect is zero.

Static f vs. T Stability

For a thermal transient compensated oscillator, the f vs. T stability is determined by the static f vs. T characteristic, the oven stability, and the difference between the oven set point temperature and the temperature where the f vs. T characteristic has zero slope, i.e., the oven offset. Table I shows how the f vs. T instabilities of an oscillator vary with oven characteristics. The static f vs. T characteristic assumed in Table I is for an SC-cut resonator the turnover point of which is 10°C from the inflection point [4,6].

As can be seen from Table I, for small oven offsets and high-stability ovens, the f vs. T instability can be less than the oscillator's $\sigma_y(\tau)$. The assumed f vs. T characteristic for Table I is not the best that can be achieved. Figure 7 shows the f vs. T for the optimum resonator, i.e., for one where the turnover

Table I - Frequency offsets due to thermal cycling as a function of the offset of the oven set point from the resonator turnover point.

for $T_I - T_{TP}$ = 10°C		Oven Cycling Range (mK)			
		10	1	0.1	0.01
Oven Offset (mK)	100	4×10^{-12}	4×10^{-13}	4×10^{-14}	4×10^{-15}
	10	6×10^{-13}	4×10^{-14}	4×10^{-15}	4×10^{-16}
	1	2×10^{-13}	6×10^{-15}	4×10^{-16}	4×10^{-17}
	0.1	2×10^{-13}	2×10^{-15}	6×10^{-17}	4×10^{-18}
	0	2×10^{-13}	2×10^{-15}	2×10^{-17}	2×10^{-19}

temperatures coincide with the inflection temperature [6]. Such a resonator is not easy to make, however, with careful cutting, angle correction, recontouring [53], design and fabrication, it is possible to make such a resonator. The numbers in Table I would be significantly smaller for such a resonator.

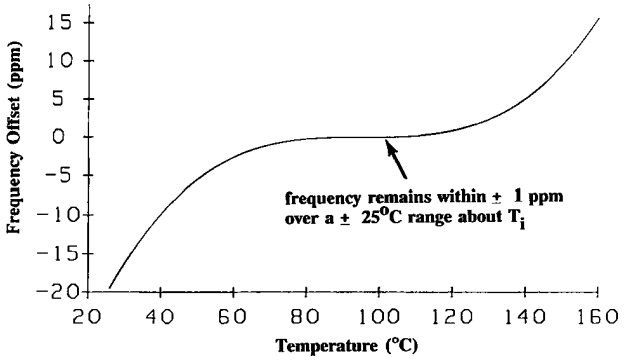


Fig. 7 Frequency versus temperature of an SC-cut resonator with zero f vs. T slope at the inflection temperature [6].

How accurately one can set the oven temperature to the optimum set point depends primarily on the hysteresis of the resonator and of the thermometer. Secondary factors are the noise of the resonator, thermometer and amplifier. If, for example, the normalized beat frequency between three times the fundamental mode frequency minus the third overtone frequency is used as the thermometer [50], then the slope of this frequency is about 10^{-4} /K, and the hysteresis of the best such SC-cut resonators (over a much wider temperature than is necessary for setting the oven to turnover) is less than 10^{-8}. Therefore, in principle, the temperature can be set to about 10^{-4} K of the optimum set point. As the hysteresis over the very narrow temperature range needed for oven temperature setting can be expected to be much smaller than 10^{-8}, and, as will

be discussed later, the hysteresis of existing resonators is not known to be limited by any fundamental limits, the fundamental limit to setting the oven temperature may ultimately be limited only by the $\sigma_y(\tau)$ of the resonator. If so, then the oven temperature can be set much closer to the optimum temperature than 10^{-4} K, if necessary.

Oven Stability

Most ovenized crystal oscillators use a thermistor to sense and ultimately control the temperature of the oven shell. All thermistors exhibit drift of the apparent temperature with time and also change with large temperature cycling. Typical drift is from 0.1 to 0.001 K/y [54].

Figure 8 shows a simplified electrical analog for a thermal enclosure around a crystal resonator. T_H, the temperature of the heater is replaced by V_H, the temperature of the resonator, T_R by V_R, and the outside temperature T_o by V_o. The heat capacity of the resonator is represented by C_R, the thermal impedance between the heater and resonator by R_{HR}, and the thermal impedance between the resonator and the outside by R_{RO} [55].

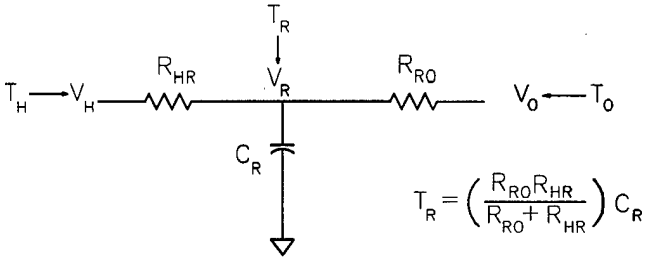


Fig. 8 Simplified electrical analog of an oven. From [55].

If we assume that the thermal sensor and heater are tightly coupled, the fundamental thermodynamic limit for temperature stability of the heater is [56]

$$\Delta T_H = \sqrt{\frac{kT_H^2}{C_H}} \qquad (8)$$

where k is Boltzmann's constant and C_H is the heat capacity of the sensor/heater system. For an aluminum heater shell and sensor weight of approximately 7 gm and T_H = 80°C, Eq. (8) indicates a minimum temperature fluctuation of 0.5 nK. At equilibrium, the rate of change of heater temperature due to the temperature fluctuations is given by

$$\frac{dT_H}{dt} = \frac{T_H}{\tau_H}\sqrt{\frac{k}{C_H}} \qquad (9)$$

where $\tau_H = R_{oH}C_H$ is the thermal time constant of the sensor/heater system to the outside. For τ_H longer than 0.01 s the temperature fluctuations due fundamental thermodynamic processes are so small that they can be ignored (unlike in infrared imaging technology, where the much smaller heat capacities of pixels makes this effect a fundamental limit.)

For simplicity, let us assume a thermistor with a resistance of 50 kΩ at 25°C and a slope at 80°C of roughly 0.8 kΩ/C at 80°C. If the current passing through the thermistor is 0.5 mA then the power dissipated in the thermistor is 1.8 mW. If the time constant for the oven cooling is 100 s and the heat capacity is 6 J/K, which corresponds to roughly 7 gm of aluminum, then the thermal conductance is 6 J/K/100 s = 0.06 W/K. This results in a self heating error of 1.8 mW/0.06 W/s = 30 mK. The error signal from the thermal sensing circuit is roughly

$$dV/dT = 0.5\ mA \times 1.8\ k\Omega/K = 0.9\ V/K. \qquad (10)$$

A change in the sensing current of 5×10^{-6} will change the self heating by about 300 nK. Reducing the sensing current will reduce the self heating and the inherent sensitivity to small changes in current but will also reduce the error signal.

The noise in the thermal bridge, due to shot noise in the current, Johnson noise in the thermistor, and typical noise in amplifiers, can be roughly estimated to be 100 nK at a few seconds for a dc bridge, and about 30 nK for an ac bridge. The long-term stability in both cases should be dominated by the thermistor aging of approximately 0.1 to 0.001 K/y [54].

The rate of change of the resonator temperature is approximately given by

$$\frac{dT_R}{dt} = \left[\frac{dT_H}{dt} + \frac{dT_0}{dt}\frac{R_{HR}}{R_{RO}}\right]\left(\frac{1}{1+\tau_R}\right) \qquad (11)$$

where

$$\tau_R = C_R\frac{R_{RH}R_{RO}}{R_{RH}+R_{RO}}$$

The first term in Eq. 11 is due to fluctuations in the oven temperature and the second term is due to changes in the outside temperature. For times long

compared to τ_o the temperature of the resonator follows

$$T_R \approx T_H - \frac{(T_H - T_O)R_{HR}}{R_{RO}} \qquad (12)$$

The temperature error of the resonator is due to the thermal coupling to the outside and the finite thermal impedance to the sensor/heater system. The thermal isolation of the resonator G_{iso} is limited to

$$G_{iso} = \frac{R_{RO}}{R_{HR}} \qquad (13)$$

even if the average temperature of the sensor and heater are perfectly stable. A typical value of G_{iso} is of order 1000. A temperature change in the outside temperature of 1°C results in a transient change in the resonator of approximately $(1/1000)(1/1+\tau_R)$. If τ_R is 100 s then the thermal transient is approximately 10 μK/s with a total offset of 1 mK.

The temperature error of the resonator can be greatly reduced if the set point for the sensor is changed slightly to compensate for the changes in the outside temperature. This can be accomplished by applying a signal proportional to the difference between the current outside temperature and the nominal outside temperature, to change the temperature of the heater by an amount

$$dT_H = (T_H - T_O)G_{ff} \qquad (14)$$

This results in a resonator temperature of

$$T_R = T_H - (T_H - T_O)\left(\frac{R_{HR}}{R_{RO}} - G_{ff}\right) \qquad (15)$$

By adjusting the feed-forward gain G_{ff}, it is possible to compensate for the thermal coupling to the outside and thereby improve the thermal isolation by a factor of 10 to 100 over the simple traditional approach. Practical limits to achievable thermal isolation are due to temperature gradients. Thermal gains of 10^5 in a single oven have been achieved using this technique. The transient response of this feed forward compensated oven arrangement is also greatly improved over conventional ovens [55].

Retrace

An OCXO's frequency offset upon turn-on and stabilization after a temperature excursion due to an off-period is called "retrace". The magnitude of retrace is a function of the duration of the off period and the oscillator's temperature-time profile during that period. A larger temperature excursion from the oven set point and a longer duration usually result in a larger frequency offset upon turn-on and stabilization.

Retrace is closely related to the hysteresis exhibited by temperature compensated crystal oscillators [57]. The mechanisms responsible for retrace are believed to be the same as those responsible for aging and hysteresis. In addition to the aging mechanisms discussed above, stress changes resulting from the temperature excursion, from temperature gradients, and from thermal expansion coefficient mismatches can worsen the effect.

Although minimizing retrace presents serious practical problems, none of the known retrace mechanisms appear to present a fundamental limit. For example, for the "ideal" resonator described below, virtually all known retrace mechanisms would be absent.

Acceleration Sensitivity

Acceleration sensitivity [10] was not well understood until recently. Thanks to the theoretical work of Tiersten [58] and his students, and others [59], the origin of acceleration sensitivity, and how it can be minimized, is now understood.

Leaving out tensor notation for simplicity, the acceleration induced normalized frequency shift $\Delta v/v$ can be represented as

$$\frac{\Delta v}{v} = \int_V cg^2 dV \qquad (16)$$

where c represents the change in the effective elastic constants due to the biasing deformation field resulting from the acceleration, g is the gradient of the normalized mechanical displacement vector associated with the mode of vibration (i.e., of the "mode shape"), and V is the undeformed volume of the resonator. Theoretically, the integral can be zero for bulk acoustic wave quartz resonators, e.g., for SC-cuts, when the mounting is completely symmetrical with respect to the mode shape.

So, although many practical limits must be overcome before zero acceleration sensitivity resonators can be produced, the fundamental limit to acceleration sensitivity is zero. Because of defects in the quartz and finite fabrication tolerances, it may be necessary to adjust the mode shape in order to achieve near zero acceleration sensitivity [59].

Shock Resistance

The frequency excursion during a shock is due to the resonator's acceleration sensitivity. The magnitude of the excursion is a function of resonator design, and of the shock induced strains in the resonator. (Resonances in the mounting structure amplify the strains.)

A shock-induced permanent frequency offset can be due to: shock induced stress changes, e.g., due to exceeding an elastic limit in the mounting structure, the removal or deposition of contamination from or onto the resonator surfaces, and changes in the oscillator circuitry, e.g., a load reactance change can result from displacement of a wire or other component. If a resonator with zero acceleration sensitivity to low level accelerations is shocked so that its mounting structure deforms inelastically, then the symmetry conditions required for zero acceleration sensitivity may be destroyed and the shock induced frequency shift can become non-zero.

Survival under shock is primarily a function of resonator surface imperfections. Chemical-polishing-produced scratch-free resonators have survived shocks of up to 36×10^3 g in air gun tests, and have survived the shocks due to being fired from a 155 mm howitzer (16×10^3 g, 12 ms duration) [60,61]. The feasibility of achieving shock resistance in excess of 10^5 g has been shown [62].

The integral in Eq. 16 for the acceleration induced frequency shift is valid at all acceleration levels, provided that the resonator plate and its mounting are linearly elastic [63]. The quartz plate remains sufficiently linear until near its breaking point, and, in principle, the mounting structure can also be designed to behave linearly (e.g., by using quartz for the mounting). Therefore, although currently available resonators do generally exhibit a frequency shift upon experiencing a shock, the fundamental limit to shock induced frequency shift is zero, and the fundamental limit to shock resistance is greater than 10^5 g.

Ionizing radiation

Ionizing radiation can be divided into two categories: photons (X-rays and γ-rays), and particles (mostly neutrons, but also protons, electrons, α-particles, and other elementary particles) [6, 7, 64-66].

The response of an oscillator to a pulse of ionizing (photon) radiation consists of two parts. Initially, there is a transient frequency change that is due primarily to impurity motion and the thermal-transient effect caused by the sudden deposition of energy into the crystal unit. This effect is a manifestation of the dynamic f vs. T effect discussed earlier. The transient effect is absent in SC-cut resonators made of high purity quartz.

In the second part of the response, after steady state is reached, there is a permanent frequency offset that is a function of the radiation dose and the nature of the crystal unit. The frequency change versus dose is nonlinear, the change per rad being much larger at low doses than at large doses. At doses above 1 kilorad (SiO_2), the rate of frequency change with dose is quartz-impurity-defect dependent.

The impurity defect of major concern in quartz is the substitutional Al^{3+} defect with its associated interstitial charge compensator, which can be an H^+, Li^+, or Na^+ ion, or a hole. This defect substitutes for a Si^{4+} in the quartz lattice. Radiation can result in a change in the position of weakly bound compensators, which modifies the elastic constants of quartz and thereby leads to a frequency change. The movement of ions also results in a decrease in the crystal's Q, i.e., in an increase in the crystal's equivalent series resistance, especially upon exposure to a pulse of ionizing radiation. If the oscillator's gain margin is insufficient, the increased resistance can stop the oscillation for periods lasting many seconds.

A high level pulse of ionizing radiation produces photo currents in the circuit which results in a momentary cessation of oscillation, independent of the type of quartz used in the resonator. In oscillators using properly designed oscillator circuitry and resonators made of swept quartz, the oscillator recovers within about 15 μs after exposure.

Sweeping is a high-temperature, electric-field-driven, solid-state purification process in which the weakly bound alkali compensators are diffused out of the lattice and replaced by more tightly bound H^+ ions and holes [31,32, 65]. In the typical sweeping process, conductive electrodes are applied to the Z surfaces of a quartz bar, the bar is heated to about 500°C, and a voltage is applied so as to produce an electric field of about 1 kV/cm along the Z direction. After the current through the bar decays (due to the diffusion of impurities) to some constant value, the bar is cooled slowly, the voltage is removed, and then the electrodes are removed. Crystal units made from swept quartz exhibit neither the radiation-induced Q degradation nor the large radiation-induced frequency shifts. Swept quartz (or low aluminum content quartz) should be used in oscillators which are expected to be exposed to

ionizing radiation. For example, at a 1 megarad dose, the frequency change can be as large as 10 ppm when the crystal unit is made from natural quartz, which is, usually, relatively impure; it is typically 1 to a few ppm when the crystal is made from cultured quartz, and it can be as small as 0.02 ppm when the crystal is made from swept cultured quartz.

At low doses (e.g., at a few rads) the frequency change per rad can be as high as 10^{-9} per rad [66]. The low-dose effect is not well understood. It is not impurity-dependent, and it saturates at about 300 rads. At very high doses (i.e., at >> 1 Mrad), the impurity-dependent frequency shifts also saturate because, since the number of defects in the crystal are finite, the effects of the radiation interacting with the defects are also finite.

When a fast neutron hurtles into a crystal lattice and collides with an atom, it is scattered like a billiard ball. A single such neutron can produce numerous vacancies, interstitial, and broken interatomic bonds. The effect of this "displacement damage" on oscillator frequency is dependent primarily upon the neutron fluence. The frequency of oscillation increases nearly linearly with neutron fluence. The rates for AT-cut and SC-cut resonators are: 8×10^{-21} neutrons per square centimeter (n/cm^2) at a fluence range of 10^{10} to 10^{12} n/cm^2, 5×10^{-21}/n/cm^2 at 10^{12} to 10^{13} n/cm^2, and 0.7×10^{-21}/n/cm^2 at 10^{17} to 10^{18} n/cm^2.

As the lattice damage produced by neutrons is unavoidable (except for using a great deal of shielding), the frequency changes produced by neutrons appear to be unavoidable, at first. However, because the frequency change per neutron per cm^2 is constant for a given resonator, if one can sense the neutron fluence, then in principle, one can compensate for the frequency changes [6, 67].

Would an SC-cut crystal made of quartz that is free of impurities exhibit a frequency change due to radiation? As the reported frequency offsets produced by photons are attributed to changes at impurities, the answer is probably that there would be no permanent frequency offset produced by photons. Would there be a temporary frequency change? Probably yes, because a pulse of radiation liberates electrons which can produce temporary local electric fields in the quartz, and on the surfaces of the resonator. The magnitude of this frequency change is unknown, but is probably small because the integral of the charge distribution between the electrodes remains zero (except for charges that may be emitted from the surfaces under an intense pulse).

Background radiation due to radioactive trace elements in the soil and in building materials, α-particle emitters in the resonator package [68], and cosmic rays will produce drift that is difficult to distinguish from aging. This drift is not aging according to the definition of the term "aging". It is a radiation effect. The definition of aging [69] is that it is "the systematic change in frequency with time due to internal changes in the oscillator." Added to the definition is: "Note - It is the frequency change with time when factors external to the oscillator (environment, power supply, etc.) are kept constant." Drift is defined as "the systematic change in frequency with time of an oscillator." Drift is due to aging plus changes in the environment and other factors external to the oscillator.

Alpha-particle induced errors in high-density integrated circuits have been of concern in the design and manufacturing of such devices [68]. These particles originate from radioactive trace impurities, such as uranium and thorium, which are present in many materials, including packaging materials. Although α-particles originating from outside a resonator package would not be able to reach the resonator (because α-particles typically penetrate solids to depths of only ~10 to 30 μm), α-particles can originate from materials inside the package, e.g., from gold-plated Kovar, sealing glasses and adhesives. Some of these α-particles can have energies as high as 8.8 MeV, however, the flux is low, typically < 1/cm^2/hour. As the sensitivity of AT-cut resonators to energetic α-particles is linear, with a slope of ~ -2×10^{-17}/cm^2/particle [70], the effect of α-emitters on long term stability appears to be < 1×10^{-13}/y. The effect on noise (or frequency jumps) also appears to be negligible, unless there is a significant transient effect whenever a high-energy alpha particle penetrates the resonator surface.

The effect of cosmic rays depends on the altitude [36], the effect being more significant at higher altitudes, and on the mix of photon and particle energies in the cosmic rays. The amount of background radiation depends on location. The average annual radiation dose from natural sources in the USA has been reported to be on the order of 0.1 rad [71]. The frequency shift due to irradiation with 0.1 rad per year is unknown, because the annealing of radiation effects over such a long period is unknown, however, if one conjectures that the sensitivity to a dose of 0.1 rad is independent of

the dose rate and that ~0.1 rad per year reaches the resonator, then the reported sensitivity to low doses of about 10^{-10} to 10^{-9} per year [66] would produce apparent aging rates of ~10^{-13} per day, which is about the order of magnitude of the best aging reported to date [2]. Therefore, background radiation (on the surface of the earth) is a significant practical limit to the achievable long term stability of crystal oscillators. Presumably, there may be locations, e.g., in a deep mine, where the background radiation effects are negligible, so the limit is a practical one, not a fundamental one.

Fundamental Limit to Oscillator Frequency

The highest frequency (bulk-acoustic-wave - BAW) quartz resonator reported to date is a 1.655 GHz AT-cut resonator made with a chemical polishing etching technique [72]. Such a resonator is 1 μm thick. The performance of this resonator was probably limited by the lack of sufficient parallelism of the quartz plate, and by the electrodes, which at such a resonator thickness can no longer be a negligible portion of the total thickness.

How thin can a resonator be? If an electrodeless design is used, and resonators can be polished and etched uniformly enough to maintain the required flatness and parallelness, then theoretically, one can achieve frequencies far above 1.6 GHz; e.g., at 0.1 μm thickness, the frequency of an AT- or SC-cut resonator would be about 16 GHz. At such a thickness, since the molecular spacing is about 0.5 nm, a resonator is about 200 molecular layers thick. At 100 molecular layers, the frequency would be 32 GHz, at 50 molecular layers, the frequency would be 64 GHz, etc.

How many molecular layers are needed to make a resonator? From a theoretical point of view [73, 74], for fundamental mode thickness shear resonators, the results in terms of frequency and mode shape are similar for a fifteen molecular layer crystal whether it is a discreet or continuum model. By applying semiconductor and silicon micromachining [75] fabrication techniques, where nanometer feature sizes are now being fabricated, and using levitation in order to eliminate mounting and bonding stresses, it may be possible to make ultrathin, ultra-miniature quartz resonators in the future. Rather than using etching techniques, it may be possible to grow a few molecular layers of single crystal quartz (or langasite, which does not undergo a phase transition below its melting point). Such extremely thin resonators would be extremely sensitive to mass loading changes, e.g., the frequency vs. thickness sensitivity would be 1% per molecular layers at 32 GHz. On the other hand, if outgassing of the resonator material is a significant instability mechanism, then being ultrathin would be an advantage because the thinner the resonator, the easier it is to outgas it (as the impurity atoms have a shorter distance to travel before reaching the surfaces). If the electrode wafers of the electrodeless structure are made of e.g., quartz or silicon, then they could be made to be good getters by baking them at high temperature, or by depositing a getter material onto their inside surfaces.

Summary of the Fundamental Limits

Table II below summarizes the fundamental limits of the various instabilities. The best instability values reported to date are included for comparison

Some Thoughts on the Paths One Might Take to Approach the Fundamental Limits

A large number of possibilities remain to be explored in order to approach the fundamental limits. In the authors' opinion, the following are some of the more important ingredients in future attempts to make significant improvements in the stabilities of oscillators:

Dislocation-free, high-purity, high-Q (and twinning-resistant) material - quartz or other.

Defect-free surfaces, e.g., via chemical polishing plus colloidal silica polishing followed by final chemical polishing [77]

SC-cut, biconvex, low-frequency (2.5 MHz 5th overtone), unconventional (noncircular) geometry

No electrodes in contact with the active area

Mount symmetrically between two identically oriented quartz plates joined by clean-surface to clean-surface atomic bonding

Alternative mounting: no mounting - i.e., levitate the resonator plate

Final clean with UV/ozone [47]

High-temperature bake in UHV immediately before sealing

Sapphire, or high-alumina ceramic, or pure Al or Al alloy or clad Al enclosure

Table II - Summary of the fundamental limits of instabilities. The best instability values reported to date are included for comparison.

Instability	Best Reported	Fundamental Limit
Aging	10^{-13} per day	0
2-sample deviation floor, $\sigma_y(\tau)$	4×10^{-14} @ τ = 10s [76]	$< 1 \times 10^{-14}$
f vs. T, static	10^{-10} (-55°C to +85°C)	0
f vs. T, dynamic	0 for a true SC-cut	0
f vs. T, retrace	10^{-10} after 24 hrs at -55°C	0
f vs. T, hysteresis	10^{-9} (-55°C to +85°C)	0
Acceleration sensitivity, Γ	~10^{-11} per g	0
Radiation - photons, steady state	10^{-14} per rad @ 1 Mrad 10^{-10} per rad @ 1 rad	0? @ 1 Mrad 0?
Radiation - photons, pulse	? (e.g., 10^{-9} @ 10^{11} rad/s)	? - no permanent offset
Radiation - neutrons	10^{-21}/n/cm^2	10^{-21}/n/cm^2 (but can compensate)
Atmospheric pressure	0 if hermetically sealed	0
Humidity	0 if hermetically sealed	0
Power supply	negligible	0
Load impedance	~10^{-15} for VSWR =2	0
Magnetic field	~10^{-10}/T [37]	0
Electric field	~0	0

Seal in UHV with clean-surface to clean-surface atomic bonding

Compensate oven for thermal gain > 10^4; use dual-mode for temperature setting and control

R ~ 10^5 Ω across resonator

Non-magnetic mounting & electrode materials; shield transformers and inductors

Low flicker noise sustaining circuit

Use notch instead of bandpass filter for mode selection

Measure voltage & power sensitivities & regulate accordingly

Total isolation in loop and output amplifier > 120 dB

Hermetically seal the oscillator enclosure

Use intelligent compensation for all systematic effects

The Ideal Resonator?

The ideal quartz resonator may be a high-perfection quartz plate levitated in an atomically clean environment, with no electrodes, mounting structures or anything else in contact with the quartz plate.

Figure 9 shows a method that may be applicable to levitating a quartz plate [78-80]. A resonant circuit is formed consisting of an inductor L that is external to the enclosure, and a parallel plate capacitor inside the enclosure. The capacitor consists of three plates, two of which are edge-to-edge, with a small spacing between them, and a third, which is a conducting surface on the plate being levitated. This third conductor is a distance d from the other two. The capacitance C of this arrangement is a function

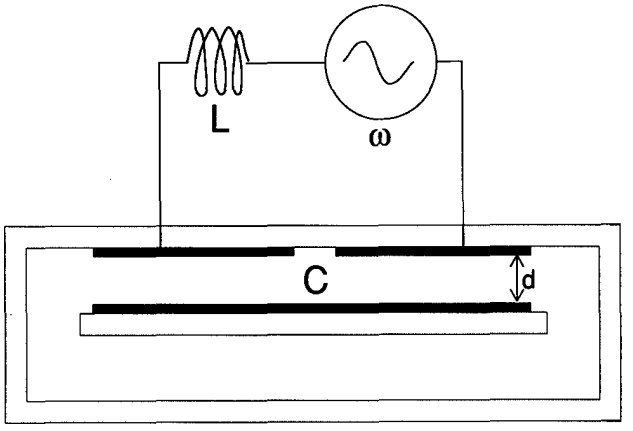


Fig. 9 Diagram illustrating levitation [78-80].

of the gap dimension d. The plate is in stable equilibrium with respect to vertical displacements when the LC-circuit is driven at an angular frequency $\omega = (LC)^{-1/2}$. This method is open-loop, i.e., no feedback sensor is used. The concept has been demonstrated experimentally via the levitation of a 24 mm x 24 mm x 180 μm microscope cover slide.

It has been shown [78] that when the plate is displaced vertically, there is a restoring force, i.e., the plate is in a potential well (in the vertical direction). This confinement is similar to that used for ion traps [81, 82]. It has also been shown [80] for a flat plate, that with four sets of levitation electrodes, it is possible to obtain stability in all six degrees of freedom required for a rigid body. Moreover, for capacitor plate area A, and a driving voltage V, the vertical levitation force F is proportional to AV^2d^{-2}. It is possible to levitate a 10 MHz 3rd overtone SC-cut resonator with V ~ 10 V at a few micrometers gap distance.

In the currently used electrodeless (BVA) resonators [28, 83], the gap between the electrodes and the plate is 5 μm on either side of the plate, and the next generation BVA resonators are expected to use 1 μm gaps [83], so the BVA gap dimensions are highly compatible with the dimensions needed for levitation at reasonable voltages. Moreover, the frequency of an electrodeless resonator depends only on the sum of the two gaps [63,83], i.e., a slight displacement of the plate in the gaps does not affect the frequency (e.g., if two equal 5 μm gaps are changed to a 4 μm gap on one side and a 6 μm gap on the other, the frequency is unaffected). Therefore, slight displacements of a levitated plate, e.g., due to vibration, ought not affect the frequency. One possible way to approximate the ideal resonator is to combine levitation with a BVA-type structure, e.g., as shown conceptually in Fig. 10 [84].

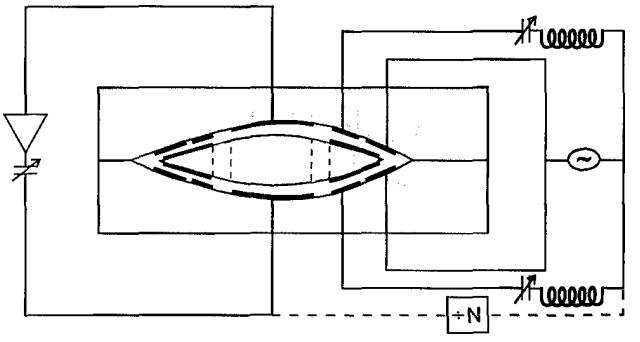


Fig. 10 One approach to combine levitation with the BVA style resonator. From [84].

The area on which the levitation electrodes are deposited are outside the active area of the resonator, separated from the active area by quartz bridges. The levitation frequency can be derived from the oscillator frequency.

An interesting side effect of levitation is that heat losses from the resonator are limited to radiation losses. The amount of energy needed to heat the resonator to the turnover temperature is small, and the thermal time constant for heating the resonator to its turnover temperature is short, e.g., when heated in the thickness direction by IR radiation [85], the time constant is a fraction of a second, even for a 5 MHz 5th overtone SC-cut resonator. The time constant is proportional to the square of the plate thickness, so the time constant for warmup can be in the milliseconds for high frequency resonators, whereas the time constant for cooling is much longer, due to the small thermal conductance to the outside.

Conclusions

We have explored many parameters that affect the stability of quartz crystal oscillators and find only two significant fundamental limits to stability, Johnson noise of the resonator, and phonon scattering within the resonator. Current technology appears to be sufficient to significantly reduce the effects of all other parameters on frequency stability below the best presently available in quartz oscillators. We expect advances in quartz (and possibly other) materials, resonator design, fabrication methods, and electronics to make possible crystal oscillators of greatly improved stability in the future.

Acknowledgments

The authors gratefully acknowledge the many fruitful discussions on this topic with their colleagues, especially, Arthur Ballato, Raymond J. Besson, Michael M. Driscoll, Raymond L. Filler, Jean-Jacques Gagnepain, Joel J. Martin, Jerry R. Norton, Thomas E. Parker, Charles Stone, Harry F. Tiersten, and Yook-Kong Yong.

References

[1] J. A. Barnes, "Basic concepts of precise time and frequency," in Time and Frequency: Theory and Fundamentals, NBS Monograph 140, B. E. Blair, editor, pp. 1-14, 1974.

[2] J. R. Vig and T. R. Meeker, "The aging of bulk acoustic wave resonators, filters and oscillators," in Proc. 45th Annu. Symp. Freq. Contr., 1991, pp. 77-101.

[3] R. L. Filler and J. R. Vig, "Long-term aging of oscillators," in IEEE Trans. Ultrason., Ferroelec., and Freq. Contr., vol. 40, 1993, pp. 387-394.

[4] F. L. Walls & J.-J. Gagnepain, "Environmental sensitivities of quartz oscillators," in IEEE Trans. on Ultrason., Ferroelec., and Freq. Contr., vol. 39, 1992, pp. 241-249.

[5] T. E. Parker, "Characteristics and sources of phase noise in stable oscillators," in Proc. of the 41st Annu. Freq. Contr. Symp., 54, 1966, pp. 329-330.

[6] J. R. Vig, "Quartz crystal resonators & oscillators - for frequency control and timing applications - a tutorial," R&D Technical Report SLCET- TR-88-1 (REV 6.1), August 1993, AD-A268205.

[7] J. R. Vig, "Introduction to quartz frequency standards," in Tutorials from the Twenty-third Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 1-49, December 1991, AD-A254745; also available as R & D Technical Report SLCET-TR-92-1 (Rev. 1), October 1992, AD-A256373.

[8] L. Cutler and C. Searl, "Some aspects of the theory and measurement of frequency fluctuations in frequency standards," in Proc. IEEE, vol. 54, 1966, pp. 136-154.

[9] F. L. Walls, P.H. Handel, R. Besson, and J.-J. Gagnepain, "A new model of 1/f noise in BAW quartz resonators," in Proc. 1992 IEEE Freq. Contr. Symp., 1992, pp. 327-333.

[10] J. R. Vig, C. Audoin, M. M. Driscoll, E. P. EerNisse, R. L. Filler, R. M. Garvey, W. J. Riley, R. C. Smythe, and R. D. Weglein, "The effects of acceleration on precision frequency sources," R&D Technical Report SLCET-TR-91-3 (Rev. 1), July 1992, AD-A255465.

[11] Y. K. Yong and J. R. Vig, "Resonator surface contamination - a cause of frequency fluctuations?," in Proc. of the 42nd Ann. Symp. on Freq. Contr., 1988, pp. 397-403.

[12] R. Besson, "A new piezoelectric resonator design," in Proc. 30th Ann. Symp. on Freq. Contr., 1976, pp. 78-83.

[13] J.-J. Gagnepain, J. Ubersfeld, G. Goujon, and P. Handel, "Relationship between 1/f noise and Q factor in quartz resonators at room and low temperatures, first theoretical interpretations," in Proc. 35th Annu. Symp. on Freq. Contr., 1981, pp. 476-483.

[14] F. L. Walls, "The influence of pressure and humidity on the medium and long-term frequency stability of quartz oscillators," in Proc. 42nd Ann. Symp. Freq. Contr., 1988, pp. 279-283, IEEE Cat. No. 88CH2588-2.

[15] M.M. Driscoll and W.P. Hansen, "Measured VS volume model-predictor flicker-of-frequency instability in VHF quartz crystal resonators," in Proc. 1993 IEEE Freq. Contr. Symp., 1993, pp. 186-192.

[16] C. Nelson, "High isolation distribution amplifier for 5 MHz and 10 MHz," in Proc. 1994 IEEE Freq. Contr. Symp., 1994.

[17] L.M. Nelson and F.L. Walls, "Environmental effects in mixers and frequency distribution systems," in Proc. of 1992 IEEE Freq. Contr. Symp., Hershey, PA, May 27-29, 1992, pp. 831-837.

[18] J.-J. Gagnepain, J.C. Poncot, C. Pegeot, "Amplitude-frequency behavior of doubly rotated quartz resonators," in Proc. 31st Annu. Symp. Freq. Contr., 1977, pp. 17-22.

[19] R. L. Filler, "The amplitude-frequency effect in SC-cut resonators" in Proc. 39th Annu. Symp. Freq. Contr., 1985, pp. 311-316.

[20] J. A. Kusters, "The SC-cut crystal--An overview," in Proc. IEEE Ultrason. Symp., 1981, pp. 402-409.

[21] H. F. Tiersten and D. S. Stevens, "The Evaluation of the Coefficient of Nonlinear Resonance for SC-cut Quartz Resonators," *Proc. 39th Annu. Symp. Freq. Contr.*, pp. 325-332, 1985.

[22] J.-J. Gagnepain, "Nonlinear properties of quartz crystal and quartz resonators: A review," in Proc. 35th Annu. Symp. Freq. Contr. 1981, pp. 14-30.

[23] J.A. Kusters, "Transient thermal compensation for quartz resonators," IEEE Trans. Sonics Ultrasonics, vol. SU-23, 1976, pp. 273-276.

[24] J.A. Kusters and J. G. Leach, "Further experimental data on stress and thermal gradient compensated crystals," in Proc. IEEE 1977, pp. 282-284.

[25] E.P. EerNisse, "Quartz resonator frequency shifts arising from electrode stress," in 29th Proc. Annu. Symp. Freq. Contr., 1975, pp. 1-4.

[26] S. Ballandras, "Sensitivity of BAW devices to radial in-plane stress distribution, comparison between analytic and finite element results," in Proc. 5th Eur. Freq. and Time Forum, Besancon, 1991.

[27] E.D. Fletcher and A.J. Douglas, "A Comparison of the Effects of Bending Moments on the Vibrations of AT and SC (or TTC) Cuts of Quartz," in Proc. 33rd Annu. Symp. Freq. Contr., 1979, pp. 346-350.

[28] R. Besson and U.R. Peier, "Further advances in B.V.A. quartz resonators," in Proc. 34th Annu, Symp. Freq. Contr., 1980, pp. 175-182.

[29] J. Lowe and F.L. Walls, "Ultralinear small angle phase modulator," in Proc. 45th Annu Symp. Freq. Contr., 1991, pp. 645-648.

[30] R. Brendel and J.-J. Gagnepain, "Electroelastic effects and impurity relaxation in quartz resonators," in Proc. 36th Annu. Symp. Freq. Contr., 1982, pp. 97-107.

[31] J. J. Martin, "Electrodiffusion (sweeping) of ions in quartz," IEEE Trans. Ultrason. Ferroelec., Freq. Contr., vol. 35, 1988, pp. 288-296.

[32] J.G. Gualtieri, "Sweeping quartz crystals," in Proc. 1989 IEEE Ultrason. Symp. 1989, pp. 381-391.

[33] C.K. Hruska and M. Kucera, "The dependence of the polarizing effect on the frequency of quartz resonators," J. Canadian Ceramic Society, vol. 55, 1986, pp. 39-41.

[34] R.L. Filler, J.A. Kosinski, V.J. Rosati, and J.R. Vig, "Aging Studies on Quartz Crystal Resonators and Oscillators," in Proc. 38th Annu. Symp. Freq. Contr., 1984, pp. 225-231.

[35] V.J. Rosati and R.L. Filler, "Reduction in the effects of vibration on SC-cut quartz oscillators," in Proc. 35th Annu. Symp. Freq. Contr., 1981, pp. 117-121.

[36] A. Ballato, T.J. Lukaszek & G.J. Iafrate, "Subtle effects in high-stability quartz resonators," in Proc. 34th Annu. Symp. Freq. Contr., 1980, pp. 431-444.

[37] R. Brendel, B. Cretin, and, A. Habti, "Origin and measurement of quartz resonator magnetic sensitivity," in Proc. 1994 IEEE Int. Freq. Contr. Symp., 1994.

[38] J.-J. Gagnepain, "Sensitivity of quartz oscillators to the environment: Characterization methods and pitfalls," in IEEE Trans. Ultrason. Ferroelec. Freq. Contr., vol. 37, 1990, pp. 347-354.

[39] D. Pierce, S. Laffey, R. Lareau, R. A. Murray, and J. R. Vig, "The outgassing of quartz," in Proc. 1994 IEEE Int. Freq. Contr. Symp., 1994.

[40] J.R. Vig and E. Hafner, "Packaging precision quartz crystal resonators," TR ECOM-4134, July 1973, NTIS Accession No. AD-763215.

[41] P.D. Wilcox, G. S. Snow, E. Hafner, and J. R. Vig, "A new ceramic flatpack for quartz resonators," in Proc. 29th Annu. Symp. on Freq. Contr., 1975, pp. 202-210.

[42] R. Glang, et al., "High vacuum technology," in Handbook of Thin Film Technology, L. I. Maissel and R. Glang, editors, McGraw Hill, 1970, pp. 2-39 to 2-59.

[43] W.G. Perkins, "Permeation and outgassing of vacuum materials," in J. Vac. Sci. Technol., vol. 10, 1973, pp. 543-556.

[44] P.H. Singer, "Today's changing vacuum requirements," Semiconductor International, Sept. 1990, pp. 58-64.

[45] Y. Hirohata, et al., "Hydrogen desorption behavior of aluminum materials used for extremely high vacuum chamber," in J. Vac. Sci & Technol. A, vol. 11, 1993, pp. 2637-2641.

[46] H. Ishimaru, "Ultimate pressure of the order of 10^{-13} Torr in an aluminum-alloy vacuum chamber," J. Vac. Sci & Technol. A - Vacuum Surfaces and Films, vol. 7, pp. 2439-2442, 1989.

[47] J.R. Vig, "Ultraviolet-ozone cleaning of semiconductor surfaces," in Handbook of Semiconductor Wafer Cleaning Technology, W. Kern, Editor, Noyes Publications, 1993, pp. 233-273.

[48] 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.

[49] J.A. Kusters, M. C. Fischer, and J. G. Leach, "Dual mode operation of temperature and stress compensated crystals," in Proc. 32nd Annu. Symp. Freq. Contr., 1978, pp. 389-397.

[50] R.L. Filler and J. Vig, "Resonators for the microcomputer-compensated crystal oscillator," in Proc. of the 43rd Annu. Symp. on Freq. Contr., pp. 8-15, 1989, pp. 8-15.

[51] A. Ballato and J. R. Vig, "Static and dynamic frequency- temperature behavior of singly and doubly rotated oven controlled quartz resonators," Proc. of the 32nd Annu. Symp. Freq. Contr., 1978, pp. 180-188.

[52] E.P. EerNisse and J. A. Kusters, "Orientation dependence of 'True' SC-cuts," in Proc. of the 44th Annu. Symp. Freq. Contr., 1990, pp. 185-192.

[53] J.R. Vig, W. Washington, and R. L. Filler, "Adjusting the frequency vs. temperature characteristics of SC-cut resonators by contouring," Proc. of the 35th Annu. Symp. Freq. Contr., 1981, pp. 104-109.

[54] J.F. Schooley, Thermometry, Boca Raton, FL: CRC Press, 1986.

[55] F.L. Walls, "Analysis of high performance compensated thermal enclosures," in Proc. 41st Annu. Symp. Freq. Contr., 1987, pp. 439-443.

[56] H.C. Wright, Infrared Techniques, pp. 61-63, Clarendon Press, Oxford, 1973 from A. Einstein, "Zur Allgemeinen Molecularen Theorie der Warme," Ann. Physik, vol. 14, 1904, pp. 354-362.

[57] J.A. Kusters and J. R. Vig, "Hysteresis in quartz resonators - a review," IEEE Trans. on Ultrasonics, Ferroeletrics, and Freq. Contr., vol. 38, May 1991, pp. 281-290.

[58] B.J. Lwo & H. F. Tiersten, "Calculation of the optimal clip dimensioning to minimize the influence of fabrication imperfections on the acceleration sensitivity of SC-cut quartz resonators with stiffened rectangular support systems," in Proc. 1994 IEEE Int. Freq. Contr. Symp., 1994.

[59] L.D. Clayton, and E. EerNisse, "The effects of metal deposition upon acceleration induced frequency shifts for quartz crystal units," in Proc. 1993 IEEE Annu. Freq. Contr. Symp., 1993, pp. 560-573.

[60] J.R. Vig, J. W. LeBus, and R. L. Filler, "Chemically polished quartz," in Proc. 31st Annu. Symp. Freq. Contr., 1977, pp. 131-143.

[61] R.L. Filler, L. J. Keres, T. M. Snowden, and J. R. Vig, "Ceramic flatpack enclosed AT and SC-cut resonators," in Proc. 1980 IEEE Ultrason. Symp., 1980, pp. 819-824.

[62] R.C. Smythe, J. Greenstein, and J. Lipeles, " A very low profile, shock-resistant, surface- mountable quartz crystal unit," in Proc. 45th Annu. Symp. Freq. Contr., 1991, pp. 117-123.

[63] Private communication with Harry F. Tiersten, Department of Mechanical Engineering, Aeronautical Engineering & Mechanics; Rensselaer Polytechnic Institute; Troy, New York 12180-3590, USA, June 1994.

[64] J.C. King, and D. R. Koehler, "Radiation effects on resonators," in Precision Frequency Control, vol. 2 (E. A. Gerber and A. Ballato, eds.), Academic Press, New York, 1985, pp. 147-159.

[65] J.J. Martin, "Radiation-induced frequency offsets and acoustic loss in AT-cut quartz crystals," in J. Appl. Phys., vol. 68, 1990, pp. 5095-5104.

[66] T.M. Flanagan & R. E. Leadon, "Evaluation of mechanisms for low-dose frequency shifts in crystal oscillators," in Proc. 40th Annu. Symp. Freq. Contr., 1986, pp. 127-133.

[67] J. R. Vig and A. Ballato, "Method of Making Radiation Hardened Quartz Crystal Oscillators," U.S. patent application SN 719,922, patent to be issued in 1994.

[68] Microelectronics Packaging Handbook, ed. by R. R. Tummala and E. J. Rymaszewski, pp. 252, 389, 541-544, Van Nostrand Reinhold, New York, 1989.

[69] Consultative Committee on International Radio (CCIR); Recommendation No. 686, "Glossary," CCIR 17th Plenary Assembly, Vol. VII, "Standard Frequency and Time Signals (Study Group 7)," 1990

[70] T. Aoki, K., et al., "Frequency Change of Quartz Resonators Irradiated by Alpha Particles," Jap. J. of Appl. Phys. vol. 15, pp. 2307-2310, Dec. 1976.

[71] Nuclear Energy, pamphlet no. EDM-1016, U.S. Department of Energy, May 1976.

[72] J. R. Hunt and R. C. Smythe, "Chemically Milled VHF and UHF AT-cut Resonators," Proc. 39th Ann. Freq. Contr. Symp., pp. 292-300, 1985.

[73] R. D. Mindlin, "Lattice Theory of Shear Modes of Vibration and Torsional Equilibrium of Simple Cubic Crystal Plates and Bars," Int'l J. of Solids & Structures, vol. 6, pp. 725-738, 1970.

[74] Private communication, Y.K. Yong, Rutgers Univ., Dept. of Civil and Env. Eng., P.O. Box 909, New Brunswick, NJ 08854, April 1994.

[75] R. T. Howe, "Applications of Silicon Micromachining to Resonator Fabrication," Proc. 1994 IEEE Annu. Freq. Contr. Symp., 1994.

[76] J.R. Norton, "Performance of ultrastable quartz oscillators using BVA resonators," in Proc. 8th Eur. Freq. and Time Forum, 1994.
[77] S. Laffey, M. Hendrickson and J. R. Vig, "Polishing and Etching Langasite and Quartz Crystals," in Proc. 1994 IEEE Int. Freq. Contr. Symp., 1994.
[78] S. Kumar, D. Cho and W. N. Carr, "Experimental study of electric suspension for microbearings," in J. of Microelectromechanical Systems, vol. 1, 1992, pp. 23-30.
[79] U.S. Patent No. 5,015,906, "Electrostatic levitation control system for micromechanical devices," D. Cho, S. Kumar and W. Carr.
[80] S. Kumar and D. Cho, "Electrostatically levitated microactuators," in J. Micromech. Microeng., vol. 2, 1992, pp. 96-103.
[81] H.G. Dehmelt, "Radio-frequency spectroscopy of stored ions. I. storage," Adv. Mol. Phy. 3, 1967, pp. 53-72.
[82] H.G. Dehmelt, "Radio-frequency spectroscopy of stored ions. II. spectrology," Adv. Mol. Phy. 5, 1969, pp. 109-154.
[83] Private communication with Raymond J. Besson, Ecole Nationale Superieure de Mecanique et des Microtechniques; La Bouloie, Route de Gray 25030 Besancon Cedex, France, June 1994.
[84] J. R. Vig, U.S. Army CECOM invention disclosure Docket No. 5058, "Methods of supporting a resonator."
[85] J. R Vig, "Fast warm-up oven controlled piezoelectric oscillator," U.S. Pat. No. 4,259,606, Mar. 31, 1981.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

1/f FREQUENCY FLUCTUATIONS AND NONLINEARITY OF QUARTZ RESONATORS AND QUARTZ CRYSTAL OSCILLATORS

*V.N.Kuleshov and **G.D.Janushevsky

*Moscow Power Engineering Institute, Moscow, Russia;
**VEGA-M Scientific and Production Corporation, Moscow, Russia

Abstract: Levels of flicker fluctuations of VHF quartz crystal resonators parameters have been obtained from experimental data.For the same resonators nonlinear model characteristics have been measured.It turned out that resonators with unusual nonlinearity characteristics had very high levels of parameters fluctuations.

The data reported in this paper are the result both in-bridge and in-oscillator measurements for 5th overtone AT-cut quartz crystal resonators operating at 140 MHz.

The in-oscillator phase noise measurement results for 92 resonators are presented as functions of their unloaded Q-factors and series resistance R_s.The results show that on the average phase noise goes down when Q and $1/R_s$ go up.But a scattering of the phase noise levels for resonators with the same Q and R_s is more than average lowering.An approach to these results explanation and quartz resonator noise model development is suggested.

Introduction

Oscillators employing bulk acoustic wave quartz crystal resonators operating in the frequency range 40 - 160 MHz have been widely used as sources of oscillations with low phase noise for Fourier frequencies higher than 20 -30 Hz [1].To decrease the phase noise due to active devices one has to use the resonators under rather high drive levels.Therefore nonlinear effects in the resonators have to be taken into consideration.They give rise to amplitude -frequency conversion which increases the phase noise.If it were the only reason of drive level limitation and sources of 1/f noise in transistors were the only cause of phase fluctuations the phase noise levels in VHF quartz crystal oscillators had to be lower than they are really [2].Experimental study of VHF oscillators with quartz resonators of the same type have shown that their phase noise levels had variations up to 20 - 30 dB with the lowest levels close to the ones predicted by theory based on calculations of transistors noise sources influence.

The aim of this work was to understand the causes of the excess phase noise and to look for opportunities to lower it.

It seemed evident that the cause of the excess phase noise could be 1/f frequency fluctuations in quartz resonators.They were measured and investigated since 1974 [3,4], but up to 1985 the main attention was payed to HF resonators with very high unloaded Q and very low drive level.First theoretical models were also based on the data for such resonators [5-7].

At the beginning of our work it was necessary to measure 1/f frequency fluctuations in VHF resonators used as passive devices.It was planned also to find out the level of resonator model series resistance Rs fluctuations.

It was important to know a dependence of series resonant frequency f_S on the amplitude of the current through a dynamic branch of the resonator circuit model.It could be useful to find out if there was some interconnection between frequency fluctuations level and nonlinearity of the resonator.

An interpretation of the measurement results was given basing on the nonlinear circuit model of a quartz crystal resonator suggested by J.Gagnepain and R.Besson [8].We assumed that linear parts of its elements L_S,C_S,R_S were fluctuating and calculated these fluctuations basing on the results of our measurement.This approach gave an opportunity to calculate an amplitude and phase fluctuations in quartz crystal oscillator with nonlinear resonator [9].

Since 1985 some papers containing data about 1/f frequency fluctuations in VHF resonators have been published [10 - 16].But there is no yet a noise model of quartz resonator that could explain all basic experiments [12 - 15].We suppose that measurements of R_S fluctuations could be useful to create such model.

Measurement Systems

The fluctuations measurement system block diagram is shown in Fig.1.It consists of the reference oscillator (RO),limiting amplifier (LA) with low amplitude and phase noise,power divider (D), two test fixtures,two low flicker noise RF amplifiers (RFA),phase shifter (PS),phase detector (PD),two amplitude detectors (AD) ,two low noise video amplifiers (VA) one of them subtracting the AD outputs and spectrum analyzer.

One of the test fixtures contains a resonator under test, capacitor C_C compensating lead inductances of the resonator, variable inductance L_C that could compensate a resonator static capacitance C_0, resistors

0-7803-1945-1/94 $3.00 © 1994 IEEE

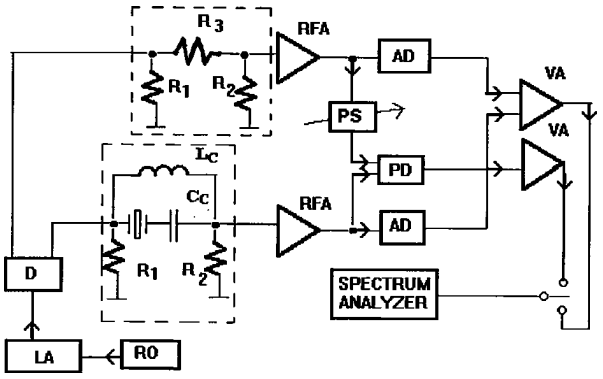


Figure 1. Block diagram for measurement of crystal resonator parameters fluctuations

R_1 and R_2 that were chosen to achieve a necessary drive level,the best sensitivity and accuracy of the measurement.The other test fixture contains a resistor R_3 equal to R_S.It is inserted instead of the resonator, capacitor C_C and inductance L_C. The reference oscillator could be tuned to adjust its frequency in the bandwidth of each quartz resonator.

A quartz resonator for the reference oscillator was selected from the whole set of the resonators under test. It had the lowest 1/f frequency fluctuations.

Inserting of amplitude detectors in the measurement system is the main feature that differs our system from the ones presented in [3,4,14].It gave an opportunity to get additional information about 1/f fluctuations in quartz resonators. Subtracting of amplitude detectors output signals helped to reduce an influence of amplitude fluctuations of driving oscillations.

Nonlinearity of each resonator was investigated by means of dynamic measurements of its amplitude-frequency and phase - frequency responces under different drive levels.The measurement system consisted of frequency synthesizer, limiting amplifier,adjustable attenuator,test fixture similar to the one shown in Fig.1,voltmeter and phase meter connected with X-Y recorders.A set of the responces has been obtained for each resonator under different drive levels.From this set a dependence of zero phase shift frequency f_S on the resonator current amplitude I_q at this frequency has been found.The function $f_S(I_q)$ was considered as characteristic of the resonator nonlinearity.

Quartz resonator model and basic ideas of the experiments

A circuit model of a quartz resonator used for planning of our experiments and data interpretation is shown in Fig.2. Basing on the paper [8] we considered L_{sn}, C_{sn}, R_{sn} as nonlinear elements that

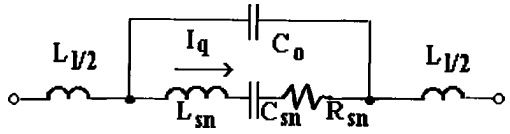


Figure 2. Circuit model of a crystal resonator

depended on the charge $q_q(t)$ and current $i_q(t)=dq_q/dt$ flowing through dynamic branch

$$L_{sn} = L_s(1 + \gamma_L i_q^2),\quad C_{sn} = C_s/(1 + \gamma_C \omega_s^2 q_q^2), \quad (1)$$

$$R_{sn} = R_s(1 + \beta i_q^2), \quad (2)$$

$\gamma_L = -\alpha(1 + Q)$, $\gamma_C = \alpha(1 - Q)$, α,β - constants that depend on the material and geometry of the resonator . Using (1) and (2) one can obtain a series resonant frequency f_S as a function of the current amplitude I_q

$$(f_S - f_{S0})/f_{S0} = \alpha I_q^2/(1 - 2\alpha Q I_q^2), \quad (3)$$

where $f_{S0} = 1/2\pi\sqrt{L_S C_S}$. If $2\alpha Q I_q^2 \ll 1$, then

$$(f_S - f_{S0})/f_{S0} \approx \alpha I_q^2. \quad (4)$$

As the dependence $f_S(I_q)$ gives rise to the amplitude - frequency conversion of fluctuations it was important to extract it from dynamic measurements with the resonator being not in thermal equilibrium.If there were some defects in the resonator that had influence both on its nonlinearity and its fluctuations we expected to notice it.

We assumed also that the elements L_S,C_S,R_S of the model Fig.2 had fractional 1/f fluctuations $w_L(t)$, $w_C(t)$, $w_R(t)$ and so

$$L_S = L_{S0}(1 + w_L(t)),\quad C_S = C_{S0}(1 + w_C(t)), \quad (5)$$

$$R_S = R_{S0}(1 + w_R(t)). \quad (6)$$

It was not possible to distinguish between influence of $w_L(t)$ and $w_C(t)$ in our experiments. We could measure only an influence of

$$w_X(t) = w_L(t) + w_C(t) \quad (7)$$

and $w_R(t)$.

For the test circuit shown in Fig.1 transfer functions from $w_X(t)$ and $w_R(t)$ to fractional amplitude fluctuations m(t) and phase fluctuations $\psi(t)$ of the output voltage were found.They were obtained under assumption that formula (4) could be used and are given in Appendix.

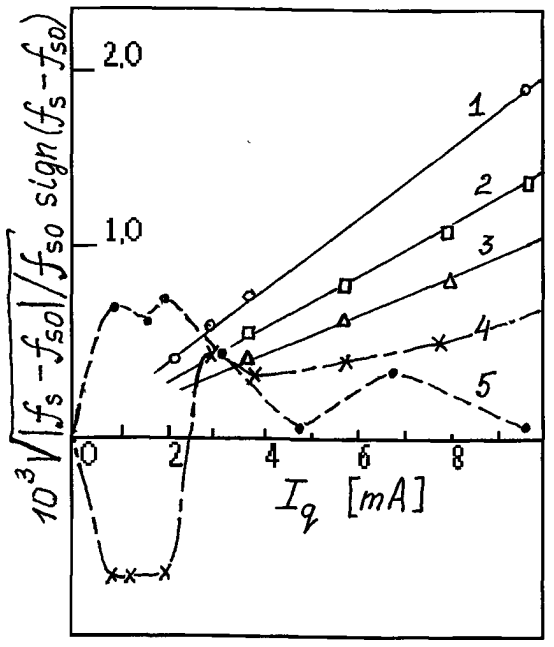


Figure 3. Measurement results of series resonant frequency shift as a function of resonant current amplitude

For the case when a reference oscillator frequency ω is equal to $\omega_S = 2\pi f_S$, $\omega_S L_C = 1/\omega_S C_0$, nonlinearity influence is negligible one can get from the Appendix simple formulae for power spectral densities $S_m(f)$, $S_\psi(f)$, $S_{WX}(f)$ and $S_{WR}(f)$ of the random functions m(t), ψ(t), $w_X(t)$ and $w_R(t)$

$$S_m(f) = S_{WR}(f)/(1 + a)^2, \quad (8)$$

$$S_\psi(f) = Q^2 S_{WX}(f)/(1 + a)^2, \quad (9)$$

where $a = (R_1 + R_2)/R_S$.

Fractional series resonant frequency fluctuations $y(t) = \Delta f_S(t)/f_S$ are connected with $w_X(t)$ by formula $y(t) = 0{,}5 w_X(t)$.So one can find from (8) and (9)

$$S_{WR}(f) = (1 + a)^2 S_m(f), \quad (10)$$

$$S_y(f) = (1 + a)^2 S_\psi(f)/4Q^2. \quad (11)$$

Using (10) and (11) one can find basic noise characteristics of quartz resonator from results of measurement.

Measurement results

The resonators under test had f_S close to 140 MHz (5th overtone,AT-cut crystal).

Results of measurement of fractional frequency shift $(f_S - f_{S0})/f_{S0}$ as a function of I_q are shown in Fig.3.Most of the resonators had these functions close to the function (4) (curves 1,2,3 in Fig.3). Resonators with such nonlinearity are called further regular ones.But there were some resonators that had very significant distinction of their nonlinearity characteristics from the function (4) (curves 4,5 in Fig.3).These resonators are called irregular ones.

Fig.4 and 5 show typical results of $S_m(f)$ and $S_\psi(f)$ measurement for the investigated set of resonators.Scattering of these power spectral densities for different resonators of the same type reaches 25 dB.(Power dissipated in each crystal was close to 0.5 mW).Nonlinearity influence estimated for regular resonators was negligible.

It turned out that all irregular resonators had high level both $S_\psi(f)$ and $S_m(f)$. Scattering of $S_\psi(f)$ and $S_m(f)$ in the set of regular resonators was about 16 dB.

Basing on these rezults and using (10) and (11) one could find power spectral densities $S_{WR}(f)$ and Sy(f). For example resonator N2 had $Q = 85 \times 10^3$ and $R_S = 36$ Ohm.In the test circuit we had $R_1 + R_2 = 18$ Ohm. So a=0.5 and from curves 2 in Fig.4 and 5 one can obtain

$$S_{WR}(f) = 1.1 \times 10^{-12}/f, \quad (12)$$

$$S_y(f) = 2.1 \times 10^{-22}/f. \quad (13)$$

The power spectral densities for the best resonators of the set were 5 - 6 dB lower.For the worst ones they were about 10 dB higher.

Some measurements have been done in oscillator similar to the one presented in the paper [1].It has been ajusted so that oscillations frequency was equal to f_S and resonator power dissipation was close to 0.5 mW.The results of in-oscillator measurement of phase noise power spectral density $S_\phi(f)$ were compared with results of calculations based on in-bridge measurement $S_{\phi c}(f)$.The value of $S_{\phi c}(f)$ was found using an apporximate formula

$$S_{\phi c}(f) = [(1 + a) f_S/2Qf]^2 S_\psi(f). \quad (14)$$

The results of the comparison are shown in Fig.6 for the same resonators as were characterized by Fig.2-5.For regular resonators the values of $S_{\phi c}(f)$ and $S_\phi(f)$ were close.For irregular ones they differed significantly.As a rule $S_\phi(f)$ was higher than $S_{\phi c}(f)$.

In this oscillator the values of $S_\phi(f)$ with f = 1 kHz were measured for all 92 resonators of the set. Fig.7 and 8 show the results in the planes (Q, S_ϕ) and $(1/R_S, S_\phi)$.They show both average biases of the measured values considered as functions of Q and $1/R_S$ and scattering.

Discussion

The main objective of our work was to contribute in understanding of 1/f fluctuation sources in quartz crystal resonators.For 20 years since the paper [3] was published several models were proposed.The discussion of the experimental data and models given in papers [12 - 15] has

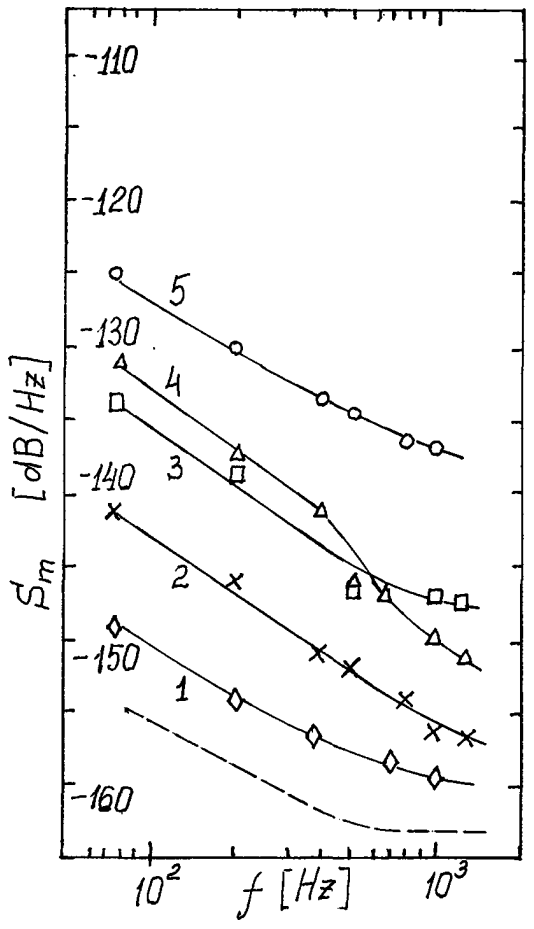


Figure 4. Measurement results of fractional amplitude fluctuations for 140 MHz AT-cut quartz resonators (- - - - sensitivity curve)

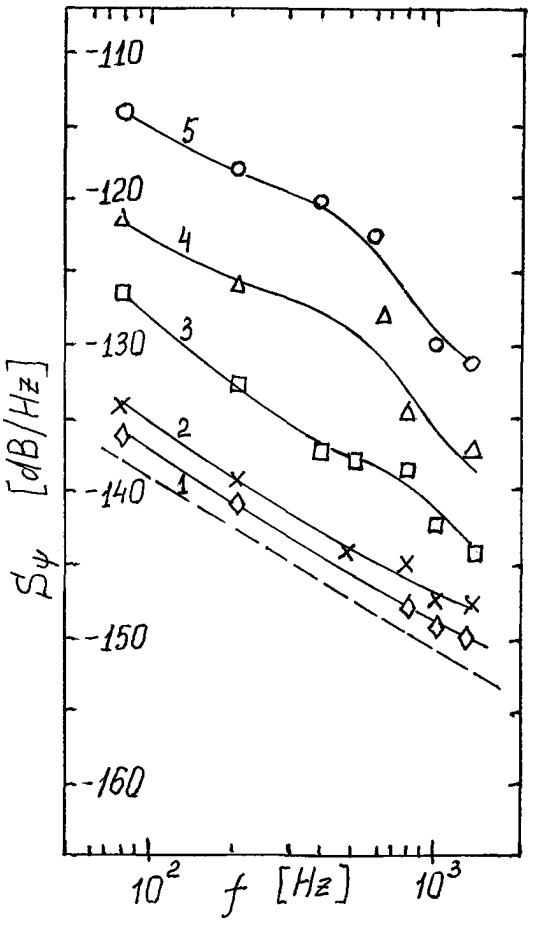


Figure 5. Measurement results of phase shift fluctuations for 140 MHz AT-cut quartz resonators (- - - - sensitivity curve)

shown that there are several sources of the fluctuations in real resonators [14,15].It seems reasonable to suppose that 1/f fluctuations are caused by localized defects of the crystal distributed in its active volume [13,15,16].They give rise the elasticity,density and acoustic energy dissipation fluctuations.So the circuit model parameters have to fluctuate. Using this approach one can explain results of the experiments presented in this paper.

When the defects concentration becomes high one can expect to notice their influence on nonlinearity of the resonator.In this case 1/f parameters fluctuations have to be high.That is what we see in Fig.3,4,5.

The measurements (Fig.4,5) have shown that on the average the resonators having higher $S_\psi(f)$ have also higher $S_m(f)$.But when we compare two resonators a difference between levels of $S_\psi(f)$ is not equal to a difference of $S_m(f)$.It could be explained by dependence of each defect contribution in resonant frequency and energy dissipation fluctuations on its position in active volume.

Direct measurements of $S_{WR}(f)$ and $S_y(f)$ for the same resonator give an opportunity to compare the calculations of $S_y(f)$ based on the paper [5] with

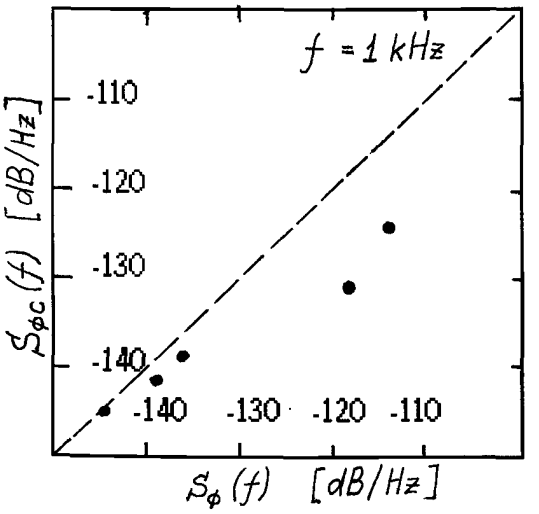


Figure 6. Comparison of power spectral density measured in oscillator $S\phi(f)$ with the one calculated on the base of in-bridge measurement $S_{\phi c}(f)$

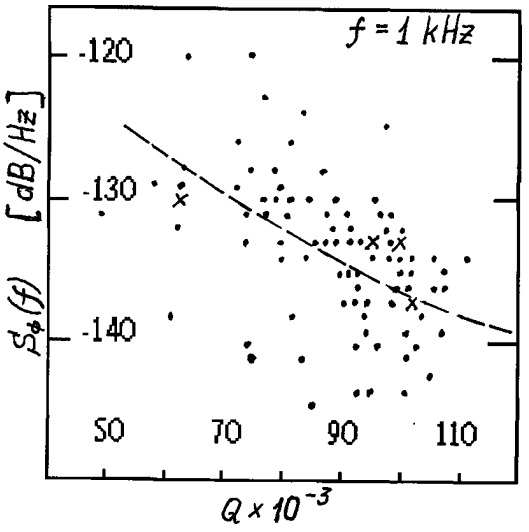


Figure 7. Results of in-oscillator phase noise measurements for 140 MHzAT-cut resonators as functions of their Q. (Crosses are double points)

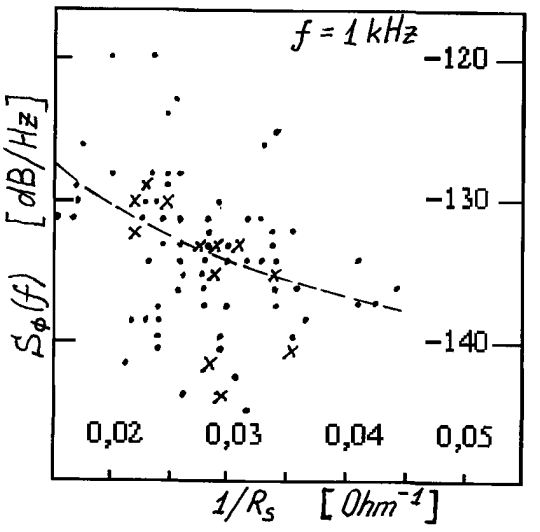


Figure 8. Results of in-oscillator phase noise measurements for 140 MHz AT-cut resonators as functions of their R_S. (Crosses are double points)

experiment.Using formula (12) from [5] one can obtain $S_y(f) = S_{wR}(f)/4Q^4$. For the resonator N2 this formula and measurement result (12) give $S_y(f) = 5{,}8\text{x}10^{-33}/f$. It is not consistent with experimental result (13).So we have to assume an independent influence of fluctuation sources both on the resonant frequency and on energy dissipation.

The results of measurement presented in Fig.7,8 show that on the average the values of $S_\phi(f)$ change approximately as $1/Q^4$ and R_S^2 (dotted lines in Fig.7 and 8).But scattering of the results was more than average lowering and the lowest value of $S_\phi(f)$ was obtained with resonator that had not the highest Q or lowest R_S.The scattering of measured values was the same both in-bridge and in-oscillator measurements.

Comparing these data with results presented by M.Driscoll and W.Hanson (Fig.3 in the paper [14]) one can see that an average level of $S_\phi(f)$ in [14] is lower but the scattering of results is almost the same.

It seems important to develop a model of 1/f fluctuations in quartz crystal resonator that could explain their scattering for resonators with the same values of Q and R_S.

Conclusion

Measurements of resonant frequency and equivalent resistance fluctuations of VHF AT-cut quartz crystal resonators have been carried out with resonators inserted in passive circuit.Results of frequency fluctuation measurement were used to estimate a phase noise of oscillators and compare them with direct measurement.In-oscillator phase noise measurements have been done for 92 resonators of the same type and interconnection between the noise level and resonator parameters (Q and R_S) was discussed.

Special attention was payed to scattering of fluctuations levels for resonators with the same values of Q and R_S that is more than for example average lowering of this levels with Q growing.

For resonators of the same set dynamic measurements of series resonant frequencies f_S as functions of resonant current amplitude Iq have been done.It was found that there were some resonators that had these functions sufficiently differing from usual ones and these resonators had very high levels of 1/f parameters fluctuations.

It is supposed that results of the experiments could be explained basing on the model with localized sources of resonant frequency and energy dissipation fluctuations distributed in active acoustic volume of resonator.

Acknowledgements

The authors would like to acknowledge T.Boldyreva for valuable discussion and help in this paper preparation.

Appendix: Transfer Functions

Transfer functions from $w_X(t), w_R(t)$ to the values $\psi(t), m(t)$ measured in the test circuit shown in Fig.1 are designated $K(w_X,\psi;j\Omega)$, $K(w_R,m;j\Omega)$, $K(w_X,m;j\Omega)$ and $K(w_R,\psi;j\Omega)$,where $\Omega = 2\pi f$.They were obtained under next assumptions: $\omega = \omega_S$, where ω is a frequency of the reference oscillator, $\omega_S C_0 = 1/\omega_S L_c$, $\omega_S L_l = 1/\omega_S C_c$.

Introducing designations

$$T = 2Q/\omega_S, \qquad \text{(A1)}$$

$$N(j\Omega) = 1 + a + j\Omega T + (2.25\beta - 1.5\alpha Q j\Omega T) I_q^2, \qquad \text{(A2)}$$

$$D(j\Omega) = N^2(j\Omega) + (1.5\alpha Q I_q^2)^2, \quad \text{(A3)}$$

one can present the transfer functions in the form

$$K(w_X,\psi;j\Omega) = - QN(j\Omega)/D(j\Omega), \quad \text{(A4)}$$

$$K(w_R,m;j\Omega) = - N(j\Omega)/D(j\Omega), \quad \text{(A5)}$$

$$K(w_X,m;j\Omega) = 0, \quad \text{(A6)}$$

$$K(w_R,\psi;j\Omega) = - 4.5\alpha Q Iq^2/D(j\Omega)\,. \quad \text{(A7)}$$

Transfer functions for quartz resonator linear model can be obtained from (A1 - A7) if $\alpha = \beta = 0$. For low Fourier frequencies ($(\Omega T)^2 \ll (1+a)^2$) one can put in (A2 - A7) $\Omega = 0$ and get the formulae used to obtain (8) and (9).

References

[1] M.M.Driscoll,"Two Stage Self-Limiting Series Mode Type Crystal Oscillator Exhibiting Improved Short-Term Frequency Stability", in Proc.26th Annual Symp. on Freq. Control,1972, pp.29-42.

[2] В.Н.Кулешов, Г.Д.Янушевский,"Влияние нелинейности кварцевого резонатора на флуктуации в высокочастотном кварцевом автогенераторе", в Сб. Стабилизация частоты . Тезисы докл., ВИМИ, 1980,с. 112-114.

[3] A.E.Wainright, F.L.Walls, and W.D.Mc Caa, "Direct Measurements of the Inherent Frequency Stability of Quartz Crystal Resonators",in Proc.28th Annual Symp. on Freq. Control,1974, pp.177-180.

[4] J.J.Gagnepain,"Fundamental Noise Studies of Quartz Crystal Resonators", in Proc.30th Annual Symp. on Freq. Control,1976,pp.84-91.

[5] P.H.Handel,"Nature of 1/f Frequency Fluctuations in Quartz Crystal Resonators", Solid State Electronics, vol.22, pp.875-876, 1979.

[6] J.J. Gagnepain, J. Ubersfeld, G. Goujon, and P.H.Handel, "Relation Between 1/f Noise and Q Factor in Quartz Resonators at Room and Low Temperature, First Theoretical Interpretation", in Proc.35th Annual Symp.on Freq. Control,1981, pp.476-483.

[7] J.J.Gagnepain,M.Olivier,and F.L.Walls,"Excess Noise in Quartz Crystal Resonators", in Proc. 37th Annual Symp. on Frequency Control,1983, pp.135-137.

[8] J.J. Gagnepain and R.Besson, "Nonlinear Effects in Piezoelectric Quartz Crystals",J.Physical Acoustics. Principls and Methods, vol.11, pp.245-288, 1975.

[9] В.Н. Кулешов, Г.Д.Яиушевский, "Флуктуационные характеристикн кварцевых резонаторов", в Сб. Стабилизация частоты и прецизнонная радиотехника,2.Тезнсы докл.,1983,б.71-76.

[10] M. M. Driscoll and B. W. Kramer, " Spectral Degradation in VHF Crystal-Controlled Oscillators Due to Short-Term Instability in the Quartz Resonator",in Proc. 1985 IEEE Ultras. Symp., Oct.1985, pp.340-345.

[11] T.E.Parker, "Characteristics and Sources of Phase Noise in Stable Oscillators",in Proc. 41st Annual Symp. on Freq. Control, 1987, pp.99-100.

[12] F.L.Walls, P.H. Handel, R. Besson, and J.J.Gagnepain, "A New Model of 1/f Noise in BAW Quartz Resonators", in Proc. 1992 IEEE Freq. Control Symp., pp.327-333.

[13] T.E.Parker and G.K.Montress, "Spectral Purity of Acoustic Resonator Oscillators",in Proc. 1992 IEEE Freq.Control Symp.,pp.340-348.

[14] M.M.Driscoll and W.R.Hanson, "Measured vs. Volume Model-Predicted Flicker-of-Frequency Instability in the Quartz Crystal Resonators",in Proc. 1993 IEEE Freq. Control Symp., pp.186-192.

[15] T.E.Parker and D.Andres, " Dependence of SAW Resonator 1/f Noise on Device Size",in Proc. 1993 IEEE Freq. Control Symp., pp.178-185.

[16] I.A.Chaban," Flicker Noise in Quartz Resonators", Sov.Phys.Acoust., vol.36, no.6, pp.614-617, November 1990.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

1/f NOISE IN SURFACE ACOUSTIC WAVE (SAW) RESONATORS

Thomas E. Parker† and Donald Andres

Raytheon Company
Research Division
131 Spring Street
Lexington, MA 02173 USA

Abstract

Flicker noise measurements on 40 additional two-port, ion etched groove SAW resonators (giving a total of 56 devices) have confirmed that resonators with large active acoustic areas exhibit lower *1/f* noise levels than smaller devices. The observed noise levels vary inversely with the size of the SAW resonator. It has also been observed that there are no large differences in the *1/f* noise levels of devices fabricated on either different grades of quartz or on swept and unswept material. Previous estimates of the active acoustic area (or volume) of a SAW resonator were based on the dimensions of rectangular acoustic energy profiles that had the same peak energy density and the same total energy as the actual resonator. More precise calculations of device acoustic volume are presented herein using perturbation theory to correctly handle the non-uniform acoustic energy distributions in a SAW resonator.

Introduction

Previously reported measurements of *1/f* (or flicker) frequency noise in groups of eight [1], [2] and sixteen [3] two-port, 450 MHz ion etched groove SAW resonators have indicated that there is an inverse size dependence to the observed noise level. Specifically, the spectral density of frequency fluctuations, $S_y(f)$, decreases as the active acoustic area of the resonator increases. These measurements were made on resonators that varied in size by approximately a factor of four. The observed size dependence is consistent with frequency noise that originates in the quartz from localized, independent velocity fluctuations whose dimensions are small compared to the in-plane dimensions of the SAW resonators [1]-[3]. This size dependence has now been further substantiated by measurements on 40 additional SAW resonators. These new resonators extend the range of device sizes to more than a factor of ten. The flicker noise levels of devices fabricated on four different grades or types of quartz have also been measured and no statistically significant difference was observed.

It has also become clear that calculating the effective active area in SAW devices with non-uniform energy distributions requires special care. Techniques for calculating the effective averaging areas of typical SAW resonators are also presented in this paper.

† Current address:
National Institute of Standards and Technology
Time and Frequency Division
325 Broadway
Boulder, CO 80303 USA

Devices and Measurements

The data on *1/f* noise levels of the 40 additional 450 MHz SAW resonators comes primarily from residual noise measurements on the devices. The fabrication parameters and measurement techniques are described in detail in Refs. 2 and 3. Figure 1 illustrates the basic design of the 450 MHz two-port SAW resonators used for these measurements. All devices contained 49 fingers in each unapodized transducer and 1000 ion etched grooves

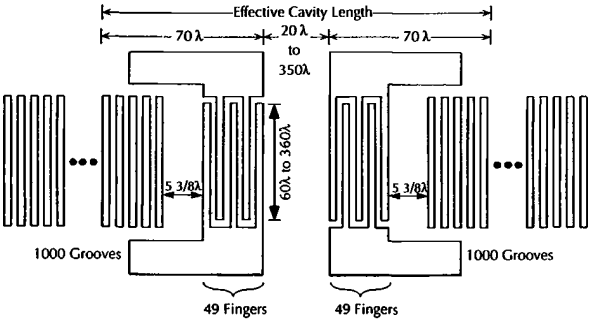


Figure 1. Illustration of the different two-port SAW resonator designs.

0-7803-1945-1/94 $3.00 © 1994 IEEE

Table 1

SUMMARY OF 450 MHZ RESONATOR DESIGN PARAMETERS AND PERFORMANCE DATA
(56 Devices Total)

F = 449.68 MHz ± 120 kHz λ = 7.01 microns

Design Ident.	No. of Dev.	Finger Overlap	Effec. Cavity Length	Active Acoustic Area (λ^2)	I.L. (dB)	Q_L	Q_u
A	12	60λ	160λ	9.6×10^3	8.0	13,200	21,800
B	12	90λ	190λ	1.71×10^4	6.4	11,900	21,800
C	4	180λ	190λ	3.42×10^4	5.3	6,450	14,500
D	4	90λ	390λ	3.51×10^4	10.6	16,550	22,850
E	12	180λ	390λ	7.02×10^4	7.7	11,575	20,675
F	12	360λ	490λ	1.76×10^5	5.7	9,800	20,900

in each grating reflector. Only the acoustic aperture and the separation between the transducers were changed among the various designs. Table 1 summarizes some of the design and performance parameters of the six different designs. The parameters listed (from left to right) are the design identifier, number of devices measured, finger overlap (acoustic aperture), effective cavity length, active acoustic area, insertion loss, loaded Q and unloaded Q. The designs identified as B through E were used for the measurements in Refs. 1, 2 and 3. Designs A and F are new and provide noise data over a wider range of device sizes. The active acoustic area as listed in Table 1 is the product of the finger overlap and the effective cavity length, and covers a range of almost a factor of twenty from the smallest to the largest devices. Data for the 16 devices from Ref. 3 is included in Table 1 along with the data for the 40 new devices, bringing the total to 56.

All of the A, C, D, and F device designs were fabricated on Premium Q cultured quartz. However, six of the twelve resonators for each of the B and E designs were fabricated on other types or grades of quartz. Specifically, two devices for each of the B and E designs were fabricated on swept Premium Q, Super Premium Q, and swept Super Premium Q material. All material was Pure Z and all of the substrates were cut and polished by the same vendor. Therefore, of the twelve devices listed in Table 1 for each of the B and E designs, six of the devices were on Premium Q material, two were on swept Premium Q, two were on Super Premium Q, and two were on swept Super Premium Q.

Calculation of Effective Averaging Area

As discussed in Refs. 1, 2 and 3, the observed size or area dependence of the flicker noise can be explained by the presence of a large number of independent, localized velocity fluctuations in the quartz. The acoustic energy of the surface wave is coherent over the entire active acoustic area, but the velocity fluctuations are not. Therefore, the spectral density of the fractional frequency fluctuations of the resonator is reduced with respect to the spectral density of the fractional velocity fluctuations of a single noise source by the number of independent noise sources contained within the active area of the resonator. In effect, the influence of the velocity fluctuations on the resonator frequency is averaged to a lower level over the entire size of the resonator due to the incoherent nature of the noise sources. Thus an area (or volume dependence) is obtained.

If the acoustic energy distribution in a SAW resonator were rectangular (a constant energy density in some locations and zero everywhere else), the calculation of the noise level reduction would be trivial. It would be simply $1/N_\varepsilon$, where N_ε is the number of independent noise sources contained within the active acoustic area. This is illustrated in Eq. 1 below (from Ref. 3)

$$S_y(f) = \frac{\varepsilon^3}{Vol} S_v(f) = \frac{1}{N_\varepsilon} S_v(f), \qquad (1)$$

where $S_y(f)$ is the spectral density of frequency fluctuations of the resonator, ε is the average separation between independent noise sources, *Vol* is the volume of

the active acoustic region, N_ε is the number of independent noise sources in the active acoustic region of the resonator, and $S_v(f)$ is the average power spectral density of fractional velocity fluctuations in a single noise element. However, since the energy distribution in a real resonator is not rectangular, a more sophisticated method of calculation is necessary. One approach is to use perturbation theory.

Following the method of Auld [4], the basic concept is that the fractional frequency fluctuation of the resonator, $\Delta F/F$, caused by a single local velocity fluctuation is weighted by a factor equal to the ratio of the local acoustic energy density to the total acoustic energy contained within the resonator. This is illustrated in Eq. 2

$$\frac{\Delta F}{F} = \frac{\frac{\Delta V}{V} E_d \Delta Vol}{E_T}, \tag{2}$$

where $\Delta V/V$ is the average local fractional velocity fluctuation, E_d is the local acoustic energy density, ΔVol is the volume of the local velocity fluctuation, and E_T is the total acoustic energy contained in the resonator. E_T is given by Eq. 3.

$$E_T = \sum_{Vol} E_d \Delta Vol \tag{3}$$

The frequency fluctuations of the resonator caused by all of the velocity fluctuations are obtained by a summation over the entire volume of the resonator. If the velocity fluctuations were coherent over the resonator volume, and $\Delta V/V$ is small, then this would yield

$$\left(\frac{\Delta F}{F}\right)_T = \frac{\frac{\Delta V}{V}\sum_{Vol} E_d \Delta Vol}{\sum_{Vol} E_d \Delta Vol} = \frac{\Delta V}{V}, \tag{4}$$

which is exactly as expected. The fractional frequency fluctuations are equal to the fractional velocity fluctuations.

If the velocity fluctuations are not coherent over the entire acoustic volume, the square of the resonator's fractional frequency fluctuation caused by all of the localized velocity fluctuations is calculated by summing the squares of the individual weighted velocity fluctuations, as shown in Eq. 5. It is again assumed that the magnitude of $\Delta V/V$ is the same for all of the localized velocity fluctuations, and that it also is much less than 1.

$$\left(\frac{\Delta F}{F}\right)_T^2 = \left(\frac{\Delta V}{V}\right)^2 \frac{\sum_{Vol}(E_d \Delta Vol)^2}{\left[\sum_{Vol} E_d \Delta Vol\right]^2} \tag{5}$$

In terms of spectral densities Eq. 5 can be written as

$$S_y(f) = S_v(f) \frac{\sum_{Vol}(E_d \Delta Vol)^2}{\left[\sum_{Vol} E_d \Delta Vol\right]^2}. \tag{6}$$

Note that $S_v(f)$ is the average power spectral density of a single velocity fluctuation.

As in Refs. 1, 2, and 3, ε is defined as the average separation between independent noise sources. Therefore, we can set $\Delta Vol = \varepsilon^3$ and also use N_ε as the total number of noise sources in the resonator. Defining $E_\varepsilon = E_d \Delta Vol$, Eq. 6 can be written as

$$S_y(f) = S_v(f) \frac{\sum_{N_\varepsilon}(E_\varepsilon)^2}{\left[\sum_{N_\varepsilon} E_\varepsilon\right]^2}. \tag{7}$$

Note that E_ε is not a constant, but varies as a function of position in the resonator.

It is useful to define a new parameter, $AF(N_\varepsilon)$, called the averaging factor which is shown in Eq. 8.

$$AF(N_\varepsilon) = \frac{\sum_{N_\varepsilon}(E_\varepsilon)^2}{\left[\sum_{N_\varepsilon} E_\varepsilon\right]^2} \tag{8}$$

Equation 7 can then be written as

$$S_y(f) = AF(N_\varepsilon) S_v(f). \tag{9}$$

$AF(N_\varepsilon)$ is a multiplicative factor which represents the amount by which the influence of the magnitude of the average spectral density of fractional velocity fluctuations, $S_y(f)$, is reduced by averaging over the entire resonator to give the spectral density of frequency fluctuations for the resonator. Another way of interpreting $AF(N_\varepsilon)$ is that it represents the amount $S_y(f)$ is reduced relative to what it would be if the velocity fluctuations were coherent rather than incoherent (see Eq. 4). Note that the summation in the numerator of Eqs. 7 and 8 cannot be replaced by an integral since the finite size of the individual noise volumes, ε^3, is critical to calculating the averaging factor.

Consider the hypothetical example of a resonator with a rectangular energy distribution (E_ε has a constant value inside the resonator, and is zero outside). Equation 8 then becomes

$$AF(N_\varepsilon) = \frac{E_\varepsilon^2 \sum_{N_\varepsilon} (1)^2}{E_\varepsilon^2 \left[\sum_{N_\varepsilon} 1\right]^2} = \frac{N_\varepsilon}{N_\varepsilon^2} = \frac{1}{N_\varepsilon}. \qquad (10)$$

This is the expected result from Refs. 1, 2, and 3. Since $(1/N_\varepsilon) = (\varepsilon^3/Vol)$, the correct volume, or size, dependence is obtained and the result is the same as Eq. 1.

A rectangular energy distribution does not accurately represent a real SAW resonator, so an attempt must be made to calculate $AF(N_\varepsilon)$ for distributions that more nearly reflect the actual situation in the SAW resonators under investigation. The distribution of the acoustic energy along the direction of the transducer fingers (the transverse direction) is not known exactly for extended cavity resonators, but it should be close to a cos^2 function [5]. This is illustrated in Fig. 2 which shows only the right half of the acoustic profile since the energy distribution is symmetric about the center of the resonator. The X axis is shown in units of ε. (Figure 2 is shown for illustrative purposes and does not represent any of the designs discussed in this paper.) The dashed line represents an equivalent, or effective rectangular energy distribution that has the same peak energy density and the same total energy as the cos^2 distribution. The width of the rectangular distribution is exactly half that of the cos^2 distribution. The value of the averaging factor for the rectangular profile in Fig. 2, $AF_r(6)$, is $1/N_\varepsilon = 1/6 = 0.167$ since each unit on the X axis

PBN-94-1414

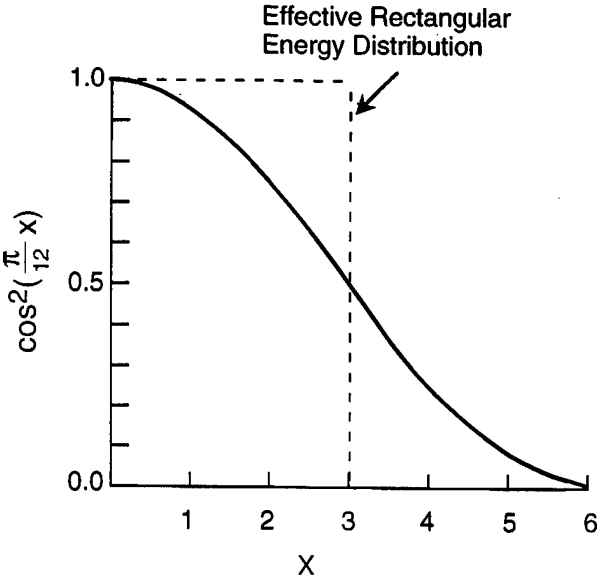


Figure 2. Example of a cos^2 acoustic energy distribution.

represents one independent noise source. ($N_\varepsilon = 6$ rather than 3 since Fig. 2 shows only one-half of the acoustic aperture.) The averaging factor for the $E_\varepsilon = cos^2[(\pi/12)X]$ distribution, $AF_c(12)$, can be calculated numerically by using Eq. 8 and is found to be 0.125. The amount of averaging (or noise reduction) for the cos^2 distribution is more than for the effective rectangular distribution, but is less than that of a rectangular distribution with the full width of the cos^2 function (which would be 1/12 = 0.083).

Another useful parameter to define is the size factor, $SF(N_\varepsilon)$, which is the ratio of the averaging factor of the effective rectangular distribution to that of the actual non-uniform distribution. For the example in Fig. 2 this is

$$SF_c(N_\varepsilon = 12) = \frac{AF_r(N_\varepsilon = 6)}{AF_c(N_\varepsilon = 12)} = 1.333. \qquad (11)$$

The importance of the size factor is that it can be used to convert effective rectangular dimensions into effective averaging dimensions, which can then be used as if they were rectangular profiles to calculate the correct level of flicker frequency noise. Thus Vol in Eq. 1 can be replaced by Vol_a, which is an effective averaging volume calculated from effective averaging dimensions (see Eq. 12 below). W_e, which is the effective rectangular width of the acoustic energy (and is equal to approximately

Table 2

EFFECTIVE RECTANGULAR ENERGY AND AVERAGING DIMENSIONS FOR THE RESONATOR DESIGNS

Design Ident	W_e (cm)	SF_c (Trans.)	W_a (cm)	C_e (cm)	SF_l (Long.)	C_a (cm)	A_e (cm^2)	A_a (cm^2)
A	2.10×10^{-2}	1.333	2.80×10^{-2}	1.12×10^{-1}	1.30	1.46×10^{-1}	2.35×10^{-3}	4.09×10^{-3}
B	3.15×10^{-2}	1.333	4.20×10^{-2}	1.33×10^{-1}	1.24	1.65×10^{-1}	4.19×10^{-3}	6.93×10^{-3}
C	6.31×10^{-2}	1.333	8.41×10^{-2}	1.33×10^{-1}	1.24	1.65×10^{-1}	8.39×10^{-3}	1.39×10^{-2}
D	3.15×10^{-2}	1.333	4.20×10^{-2}	2.73×10^{-1}	1.10	3.01×10^{-1}	8.60×10^{-3}	1.26×10^{-2}
E	6.31×10^{-2}	1.333	8.41×10^{-2}	2.73×10^{-1}	1.10	3.01×10^{-1}	1.72×10^{-2}	2.53×10^{-2}
F	1.26×10^{-1}	1.333	1.68×10^{-1}	3.43×10^{-1}	1.08	3.71×10^{-1}	4.32×10^{-2}	6.23×10^{-2}

one-half of the finger overlap) can be converted to an effective averaging width, W_a, by multiplying W_e by the size factor for the cos^2 distribution (i.e., $W_a = SF_c \cdot W_e$). W_a can then be used to calculate Vol_a. As long as the acoustic aperture is much larger than ε, the size factor for the cos^2 distribution is very close to 1.333. This is assumed to be the case for all of the resonator designs in Table 1.

The distribution of acoustic energy as the surface wave penetrates into the grating reflector is an exponential decay. However, the effective cavity lengths [6] given in Table 1 are based on effective rectangular dimensions. The apparent point of reflection for the gratings in the resonators used in this investigation is about 40 wavelengths (80 grooves) into each grating (see Fig. 1). This point is the distance into the grating where the acoustic energy has been reduced to a value equal to 1/e times the value at the beginning of the grating. It is also equal to the dimension of an effective rectangular profile with the same total energy in the grating. The same technique used for the cos^2 distribution can also be used to calculate the averaging factor, $AF_{ex}(N_\varepsilon)$, and the size factor, SF_{ex}, for an exponential function. The size factor, SF_{ex}, for an exponential decay is equal to 2.0 as long as the penetration distance of the acoustic energy into the grating is large compared to ε. However, SF_{ex} = 2.0 cannot be directly used to calculate the effective averaging cavity lengths, since all of the resonator designs have regions between the gratings in which the acoustic energy density is relatively constant. Thus the energy distribution along the longitudinal dimension is a combination of a central region with constant energy density bounded on both ends with regions of exponential decay. Equations 8 and 11 have been used to calculate longitudinal size factors, SF_l, for all of the designs in Table 1 and they are listed in Table 2, along with the calculated effective averaging dimensions for each design. W_e and W_a in Table 2 are the effective rectangular and effective averaging widths of the resonators, respectively, and C_e and C_a are the respective effective rectangular and averaging cavity lengths. A_e and A_a are the respective areas of the resonators ($A_e = W_e C_e$ and $A_a = W_a C_a$).

The distribution of the SAW energy density down into the substrate can also be approximated by an exponential decay in which ~ 80% of the acoustic energy is confined within one wavelength of the surface. The effective rectangular distribution has a penetration depth, d_e, of about 0.5 λ, but the effective averaging penetration depth, d_a, is closer to 1.0 λ since the energy distribution nearly follows an exponential decay. (Calculated from depth profiles in Ref. 7.)

$S_y(f)$ can be calculated for realistic SAW resonators by using in Eq. 12 the effective averaging dimensions calculated above and shown in Table 2.

$$S_y(f) = \frac{\varepsilon^3}{Vol_a} S_v(f) = \frac{\varepsilon^3}{W_a C_a d_a} S_v(f)$$

$$= \frac{\varepsilon^3}{A_a d_a} S_v(f) \qquad (12)$$

Equation 12 yields a more accurate estimate of the size dependence of the power spectral density of frequency fluctuations than obtained from the use of effective rectangular dimensions in Eq. 1.

The effective averaging dimension into the depth of the substrate is the same for all devices measured as part

of the size dependence experiments since all of the resonators were at the same frequency (450 MHz). Therefore this dimension does not enter into the size dependence. The exact transverse energy profiles are not known for these designs and it is not absolutely certain that they are the same for each design. Until the profiles are measured it has been assumed that they are close to the $\cos^2$ function discussed in this section. Therefore, the effective averaging dimensions in the transverse direction are simply scaled from W_e (where W_e is one-half of the finger overlap) by a constant factor for all of the designs. However, this is not true for the longitudinal dimension where the various designs have different size factors as indicated in Table 2. As will be discussed in the next section, this affects the relative sizes of the different designs.

Experimental Results

The first phase of the measurement program is illustrated in Fig. 3 which shows the measured values of $S_y(f)$ at f = 1 Hz as a function of device area for the original set of 16 resonators from the middle four designs [3]. Four devices of each design were evaluated. The device areas used for this figure are simply the products of the finger overlaps and effective cavity lengths given in Table 1. (Data using effective averaging dimensions is presented later in this section.) The solid line in Fig. 3 represents a 1/A dependence, and the size dependence of $S_y(f)$ is clearly visible. This figure is repeated from Ref. 3 since the data was obtained from both residual noise measurements on the resonators, as well as from

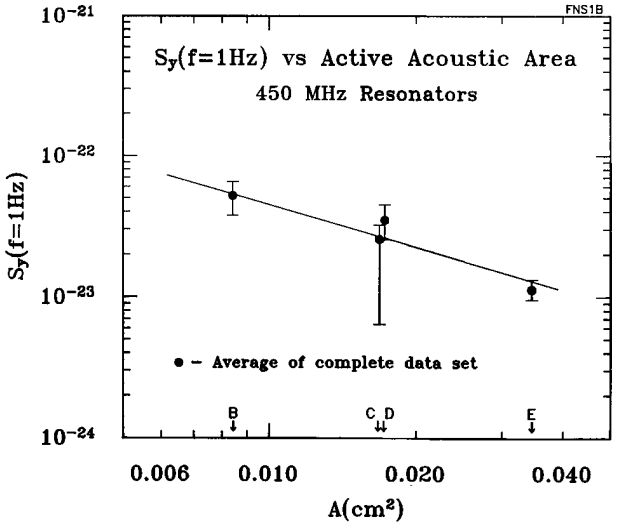


Figure 3. Measured values of $S_y(f)$ as a function of device area for the original set of 16 resonators.

oscillator noise measurements in which each resonator was installed in the same oscillator circuit. The size dependence is present in both the resonator and oscillator measurements, therefore reducing the possibility that it is a measurement artifact. The following data in this section was obtained only from residual noise measurements on the SAW devices.

The next phase of the investigation was to evaluate the $1/f$ noise for several different grades or types of quartz. As discussed earlier, B and E design devices fabricated on swept and unswept Premium Q and Super Premium Q materials were tested. Figure 4 shows the results of the measurements. The distribution of measured values of Log[$A_a S_y(f$=1Hz)] is shown for the different qualities of quartz. The parameter $A_a S_y(f)$ is a size corrected flicker noise value and was used to remove the effect of the size difference between the B and E designs (A_a is in units of cm^2). The median value indicated by the dashed line was determined from the complete set of Premium Q devices from the A through E designs (32 devices). (The F design devices were not used for reasons discussed later.). Data from twelve B

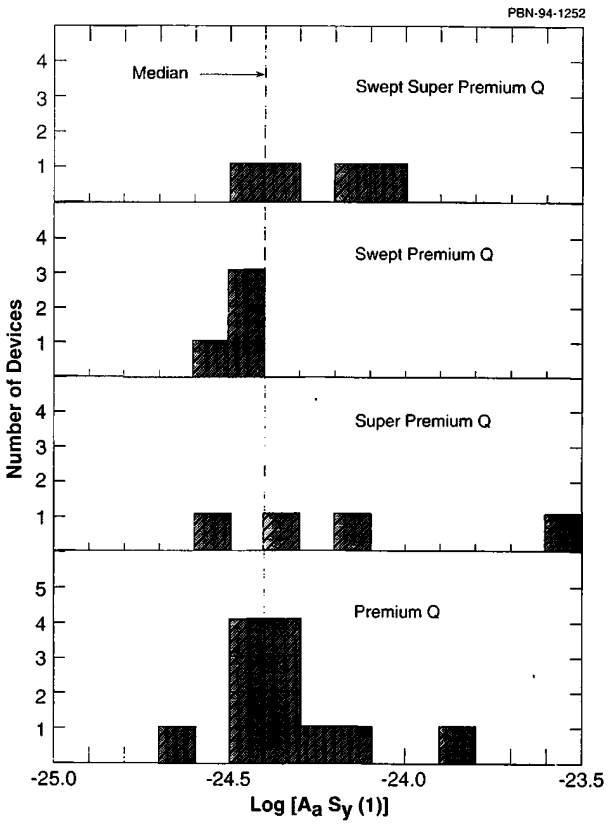


Figure 4. Distribution of measured values of $A_a S_y(f$=1Hz) for four different grades or types of quartz. Only the B and E designs were used.

and E design devices is shown for the unswept Premium Q material, but there are only four devices for each of the other types of quartz. Though the sample size is small, there does not appear to be any statistically significant difference (4 dB or larger) in the noise levels from the various types and grades of quartz. Many more devices would have to be measured to resolve average differences on the order of only 1 or 2 dB. One point worth noting is that the swept Premium Q sample, which gave the lowest average noise level, was over five years old, whereas all of the other material was new. There was no significant statistical difference in the values of $A_a S_y(f)$ between the B and E designs.

Figure 5 shows the measured values of $S_y(f=1\text{Hz})$ versus A_e (the effective rectangular area) for the complete set of 56 devices. (Note that A_e in Fig. 5 is just $0.5A$ in Fig. 3.) Maximum, mean, median and minimum values are shown for each design. The different grades of quartz have been included in the data in Fig. 5 since they did not exhibit any significant difference in noise levels. The solid line represents a $1/A_e$ dependence. The size dependence is still clearly evident even for the larger range of sizes in Fig. 5. However, the slope of the experimental data appears to be slightly shallower than that of the solid line, and the F devices have a higher noise level than expected.

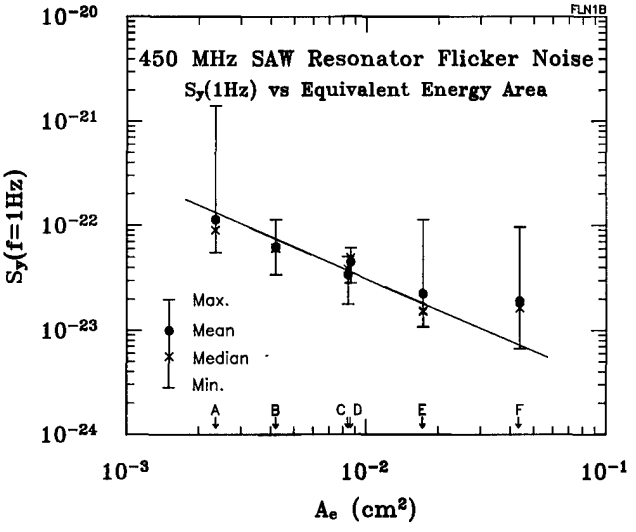


Figure 5. Measured values of $S_y(f)$ versus equivalent energy area for the complete set of 56 devices. The solid line shows a $1/A_e$ dependence.

Figure 6 shows the same noise data as that in Fig. 5, but now plotted as a function of A_a, the effective averaging area (shown in Table 2). This is a more accurate estimate of the effect of device size on the amount of noise averaging. The slight change in relative size improves the fit between the experimental data and the $1/A_a$ line. Note that the D and C devices have reversed their order and are now more consistent with the $1/A_a$ dependence. Only the F devices continue to show a significant deviation from the expected behavior (see discussion below).

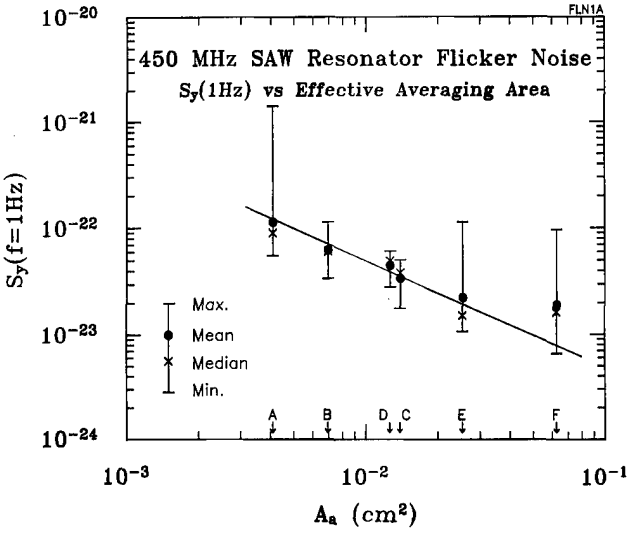


Figure 6. Measured values of $S_y(f)$ versus effective averaging area for the complete set of 56 devices. The solid line shows a $1/A_a$ dependence.

Several steps were taken to further ensure that the size dependence was not an artifact of the measurement or testing procedures. All of the residual noise measurements were conducted with +13 dBm of rf power incident on the SAW resonators, whereas oscillator measurements were made with approximately +15 dBm incident power. Obviously, under these conditions, the smaller devices had higher acoustic power densities during the measurement process than the larger devices. Figure 7 shows measured flicker noise levels on three E design devices for different incident power levels. The data in Fig. 7 clearly shows that there is essentially no power dependence. The data would have to show a power dependence similar to that indicated by the solid line in order to give the appearance of the observed size dependence.

As discussed in Refs. 1, 2, and 3, all of the devices were burned-in to reduce flicker noise levels [8]. The burn-in power was adjusted for each design to produce the same energy density (stress level) in each type of device. The data presented in Figs. 3 through 7 is all

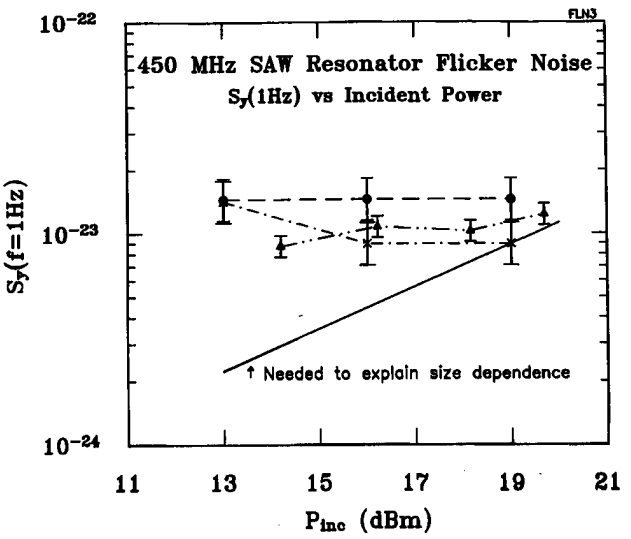


Figure 7. Measured values of $S_y(f)$ as a function of incident test power level for three different resonators.

post burn-in data. However, we must be certain that the burn-in process did not create the apparent size dependence. If the relevant parameter for the burn-in were not energy density, but rather incident rf power level, then the lower noise levels may have been created in the larger devices because they were burned-in at higher power levels. To test for this possibility, the flicker noise levels were also measured on each device prior to burn-in. This data is shown in Fig. 8 on the same scale as the post burn-in data in Fig. 6. Though there is considerably more scatter to the data, the size dependence is definitely present on the A through E designs even before the burn-in process. This is particularly evident in the minimum observed values. Thus it is clear that the size dependence was not created by the burn-in process. A careful comparison of the data in Figs. 6 and 8 also clearly shows the value of the burn-in process.

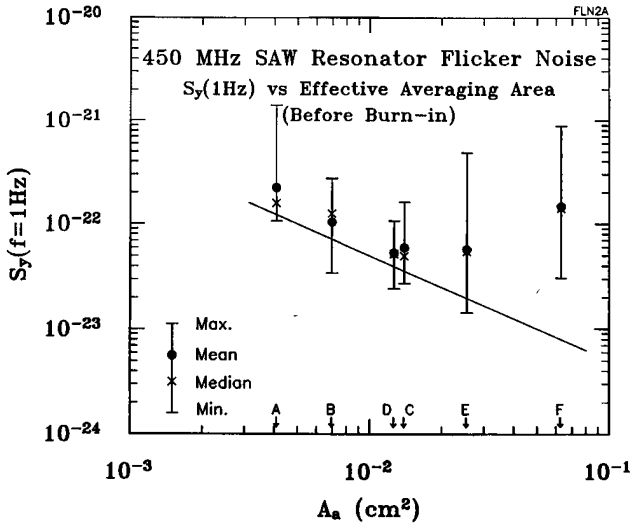


Figure 8. Measured values of $S_y(f)$ versus effective averaging area for the complete set of 56 devices before burn-in. The solid line shows a $1/A_a$ dependence.

The data in Fig. 8 shows very high noise levels for the F design devices, even though these were very large resonators. However, comparing Figs. 6 and 8, it can be seen that the burn-in process reduced the average noise level for this design by almost 10 dB, the largest reduction of any of the designs. The level of the noise process that is reduced by the rf burn-in was originally very high (relatively speaking) in these devices and may not have been reduced to a negligible level by the burn-in. Therefore, it is possible that the post burn-in data for the F design is corrupted by the presence of an additional noise process and consequently does not represent the size dependent process in the quartz. This is the reason the data from the F design devices was not included in the calculation of the median value shown in Fig. 4.

In Ref. 3 there is an extensive discussion of how the size dependent flicker noise in SAW resonators can be modeled when the penetration depth, d_a, of the acoustic energy is less than ε. The dependence on A_a in Eq. 12 is valid for all values of d_a as long as the in-plane resonator dimensions are large compared to ε. However, the dependence on d_a in Eq. 12 is valid only for the case where d_a is greater than ε, which does not appear to be generally true Flicker noise data from residual noise measurements on SAW resonators ranging from 401 to 915 MHz is also presented in Ref. 3, and volume corrected flicker noise levels are plotted versus penetration depth. However, Fig. 8 in Ref. 3 was made using equivalent rectangular dimensions which have been shown herein to be incorrect. Figure 9 in the present paper shows the same data from Ref. 3, but now plotted using effective averaging dimensions. As in Ref. 3, the data is consistent with the penetration depth being less than ε, but since d_a is approximately two times larger than d_e, it now appears that ε must be larger than about 7 to 10 microns, rather than 3 to 6 microns [3]. The data in Fig. 9 is also still consistent with r, the effective range of the velocity fluctuation, being much smaller than ε. The meaning of the dashed and dotted lines in Fig. 9 is discussed extensively in Ref. 3.

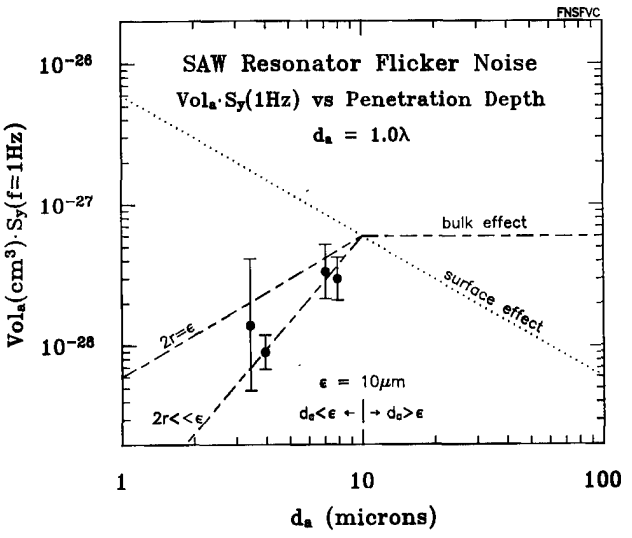


Figure 9. Volume corrected flicker noise levels versus penetration depth for SAW resonators in the 401 to 915 MHz range. (Flicker noise data and theoretical lines from Ref. 3.)

Conclusions

Flicker noise data on fifty-six 450 MHz SAW resonators of varying sizes has been presented which further substantiates the previously reported size dependence of the $1/f$ noise levels. Specifically, it has been established that the $1/f$ noise level is inversely proportional to the active acoustic area of the SAW resonator. Every effort has been made to ensure that the observed size dependence is not an artifact of the burn-in process, or of the phase noise measurement techniques. Data has been presented to support this conclusion. Flicker noise measurements on devices fabricated on swept and unswept Premium Q and Super Premium Q material have not shown any large difference (greater than approximately 3 to 4 dB) in average noise level. Finally, the calculation of effective averaging dimensions is non-trivial, and a technique using perturbation theory has been presented which gives the correct dimensions for calculating relative flicker noise levels. Effective averaging dimensions and data from Ref. 3 have been used to show that ε appears to be larger than approximately 7 to 10 microns. Furthermore, the observed size dependence limits ε to a value less than ~ 300 microns (this is the smallest in-plane dimension).

References

[1] T. E. Parker and D. Andres, "Dependence of SAW Resonator $1/f$ Noise on Device Size," in *Proc. of the IEEE Freq. Control Symp.*, 1993, pp. 178-185.

[2] T. E. Parker, "Dependence of SAW Resonator $1/f$ Noise on Device Size," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 40, no. 6, pp. 831-833, November 1993.

[3] T. E. Parker, D. Andres, J. A. Greer, and G. K. Montress, "$1/f$ Noise in Etched Groove Surface Acoustic Wave (SAW) Resonators," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, (to be published).

[4] B. A. Auld, *Acoustic Fields and Waves in Solids, Vol. II*, New York, John Wiley & Sons, 1973, pp. 322.

[5] H. A. Haus, "Modes in SAW Grating Resonators", *Journal of Applied Physics*, vol. 48, no. 12, pp. 4955-4961, December 1977.

[6] W. R. Shreve and P. S. Cross, "Surface Acoustic Waves and Resonators", in *Precision Frequency Control: Acoustic Resonators and Filters*, vol. 1, E. A. Gerber and A. Ballato, Eds., New York: Academic Press, 1985, pp. 118-145.

[7] *Microwave Acoustics Handbook, Vol 1A, Surface Wave Velocities*, A. J. Slobodnik, Jr., E. D. Conway, and R. T. Delmonico, Eds., Air Force Cambridge Res. Lab., Bedford, MA, AFCRL-TR-73-0597, 1973, pp. 517.

[8] G. K. Montress, T. E. Parker, M. J. Loboda, and J. A. Greer, "Extremely Low-Phase-Noise SAW Resonators and Oscillators: Design and Performance", *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 35, no. 6, pp. 657-667, November 1988.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ANALYSIS OF QUANTUM 1/F EFFECTS IN FREQUENCY STANDARDS

P. H. Handel* and F. L. Walls**

*Dept. of Physics and Astronomy, University of Missouri St. Louis, MO 63121
**Time & Frequency Division, Mail St 576, NIST, Boulder, CO 80303

Abstract

There is a fundamental limitation of the stability of frequency standards, caused by the quantum 1/f effect. This effect represents a quantum fluctuation of the physical currents, scattering cross sections and process rates, caused by the reaction of Bremsstrahlung on the charged particles which have emitted it. It is the only known quantum effect which is inversely proportional to Planck's constant, and which is therefore important in the limit of low, rather than high frequencies. In frequency standards the rate of change of the dipolar moment of the oscillating material system suffers a discontinuity every time an oscillation quantum is absorbed in the loss process. This causes the emission of elf bremsstrahlung and therefore quantum 1/f fluctuations in the loss rate, which translate into $1/fQ^4$ frequency fluctuations.

A newly discovered aspect of quantum physics results in a fundamental stability limit in frequency standards such as quartz resonators, Cs atomic clocks and H maser oscillators. This aspect is the quantum 1/f effect which causes physical cross sections σ and process rates Γ to fluctuate with a 1/f spectral density $2\alpha A\Gamma^2/f$. Physically, whenever the system loses a main oscillator mode phonon (quartz) or photon (Cs or H), the electric polarization current $\mathbf{J}=\dot{\mathbf{P}}$ caused by the oscillation suffers a sudden change which generates a bremsstrahlung amplitude $(\alpha A/f)^{1/2}$ for the emission of photons of arbitrarily low frequency f, which can not be detected. This yields a fluctuation of Γ because the final quantum state of the after the emission becomes a superposition of states with slightly different energies.

According to the general quantum 1/f formula, $\Gamma^{-2}S_\Gamma(f)=2\alpha A/f$ with $\alpha\equiv e^2/\hbar c = 1/137$ and $A\equiv 2(\Delta J/ec)^2/3\pi$. Setting $\mathbf{J}=d\mathbf{P}/dt\equiv\dot{\mathbf{P}}$ where $\mathbf{P}$ is the vector of the dipole moment of the quartz crystal, or of the group of N oscillating atoms of Cs or H, we obtain for the fluctuations in the rate Γ of phonon/photon removal from the main resonator oscillation mode by scattering on a phonon/photon from any other mode of average frequency $\langle\omega\rangle$ of the quartz crystal or of the Cs/H clock blackbody radiation background, (or via a two-phonon-process at a crystal defect or impurity, involving a phonon of average frequency $\langle\omega'\rangle$, or, e.g., by photon absorption in the cavity walls in the case of Cs or H standards) the spectral density

$$S_\Gamma(f) = \Gamma^2 4\alpha(\Delta\dot{\mathbf{P}})^2/3\pi e^2c^2, \qquad (1)$$

where $(\Delta\dot{\mathbf{P}})^2$ is the square of the dipole moment rate change associated with the process causing the removal of a phonon (or photon for Cs and H) from the main oscillator mode. To calculate it, we write the energy W of the interacting mode $\langle\omega\rangle$ in the form

$$W = n\hbar\langle\omega\rangle = 2(Nm/2)(dx/dt)^2$$
$$=(Nm/e^2)(e\,dx/dt)^2=(m/Ne^2)\varepsilon^2(\dot{P})^2; \qquad (2)$$

The factor two includes the potential energy contribution. Here m is the reduced mass of the elementary oscillating dipoles, e their charge, ε a polarization constant, and N their number in the resonator. Applying a variation $\Delta n=1$ we get

$$\Delta n/n = 2|\Delta\dot{\mathbf{P}}|/|\dot{\mathbf{P}}|, \text{ or } \Delta\dot{\mathbf{P}}=\dot{\mathbf{P}}/2n. \qquad (3)$$

Solving Eq. (2) for $\dot{\mathbf{P}}$ and substituting, we obtain

$$|\Delta\dot{\mathbf{P}}| = (N\hbar\langle\omega\rangle/n)^{1/2}(e/2\varepsilon). \qquad (4)$$

Substituting $\Delta\dot{\mathbf{P}}$ into Eq. (6), we get

$$\Gamma^{-2}S_\Gamma(f) = N\alpha\hbar\langle\omega\rangle/3n\pi mc^2f\varepsilon^2 \equiv \Lambda/f\,. \qquad (5)$$

This result is applicable to the fluctuations in the loss rate Γ of the frequency standard.

The corresponding resonance frequency fluctuations of the standard are given by

$$\omega^{-2}S_\omega(f) = (1/4Q^4)(\Lambda/f) = N\alpha\hbar\langle\omega\rangle/12n\pi mc^2f\varepsilon^2Q^4;$$
$$\omega^2 = \omega_0^2 - 2\Gamma^2, \qquad (6)$$

where $Q=\omega/2\Gamma$ is the quality factor of the single-mode resonator considered, and $\langle\omega\rangle$ is the circular frequency of the main oscillation mode ω_0 in the case of cesium clocks, Hydrogen maser clocks and laser clocks. However, for quartz resonators $\langle\omega\rangle$ is not the circular frequency of the main resonator mode, but rather the practically constant frequency of the average interacting phonon (or

0-7803-1945-1/94 $3.00 © 1994 IEEE

photon for Cs and H standards).

For the case of cesium clocks, Hydrogen maser clocks, and for the case of the laser clocks being developed, we can set N/n =1, because the number of excited (oscillating) ions or atoms equals the number of quanta n which they harbor (one quantum per atom), while in the case of a quartz resonator there are an average $n_\omega = kT/\hbar\omega$ quanta (phonons) present in any mode of frequency ω. Therefore, for these (non-quartz) systems, we particularize Eq. (10) by setting N/n=1 and we write

$$\Gamma^{-2} S_\Gamma(f) \equiv \Lambda/f = \alpha\hbar\omega_o/3\pi mc^2 f$$

$$= 10^{-27}\text{erg.s.}\ \omega_o\ /10\ .\ 137\ .\ 0.9\ .\ !0^{-27}\ \text{g.}\ 9\ .\ 10^{20}\text{cm}^2/\text{s}^2$$
$$= 10^{-24}\ \omega_o \qquad (7)$$

Using the approximate values of the resonance frequency in each standard, we obtain the following quantum 1/f flicker floors:

For the Hydrogen Maser: $\omega_o = 10^{10}\ s^{-1}$;
$$S_{\delta\omega/\omega}(f) = 2.5\ .\ 10^{-15}\ /\ Q^4\ f; \qquad (8)$$

For the Cs clock: $\omega_o = 6.3\ .10^{10}\ s^{-1}$;
$$S_{\delta\omega/\omega}(f) = 1.5\ .\ 10^{-14}\ /\ Q^4\ f; \qquad (9)$$

For the Rb clock: $\omega_o = 4\ .10^{10}\ s^{-1}$;
$$S_{\delta\omega/\omega}(f) = 10^{-14}\ /\ Q^4\ f; \qquad (10)$$

For the laser clock: $\omega_o = 10^{14}\ s^{-1}$;
$$S_{\delta\omega/\omega}(f) = 2.5\ .\ 10^{-11}\ /\ Q^4\ f; \qquad (11)$$

These are lower than the observed values; the quantum limit has not been obtained so far.

For the case of quartz resonators[1] we have used the interacting thermal mode of average frequency $<\omega>$ to calculate the quantum 1/f effect. The corresponding $\Delta\dot{\mathbf{P}}$ in the main resonator mode of frequency ω_o has to be also included in principle, but is negligible because of the very large number n of phonons (or photons for Cs present in the main resonator mode.

Eq. (11) can be written with $N \equiv VN/V$ for the case of quartz resonators in the form

$$S(f) = \beta V/fQ^4, \qquad (12)$$

where, with an intermediary value $<\omega> = 10^8/s$, with $n = kT/\hbar<\omega>$, T=300K and $kT = 4 \cdot 10^{-21}$ J

$$\beta = (N/V)\alpha\hbar<\omega>/12n\pi\varepsilon^2 mc^2$$
$$= 10^{22}(1/137)(10^{-27}10^8)^2/12kT\pi 10^{-27}\ 9\ 10^{20} = 1. \qquad (13)$$

The form of Eq. (7) shows that the level of 1/f frequency noise depends not only as Q^{-4} as previously proposed by Handel , but also on the oscillation frequency or the volume of the active region. This theory qualitatively fits the data of Gagnepain who varied the Q-factor with temperature in the same resonator (but not frequency or volume) for the quartz case, and the data of Walls who considered several reonators which differ in volume and frequency, again only for the quartz case. The comparison with the experiment is not yet possible in the case of cesium and hydrogen standards, because they have not reached the quantum 1/f limit. At the present time, these standards are still limited in their stability by well-known factors which do not introduce a 1/f spectral dependence.

The theory also provides the basis for predicting how to improve the 1/f level of resonators, beyond just improving the Q-factor, which has been known for many years, and which has been related to fluctuations in the dissipation. Since the level depends on the active volume, one should use the lowest overtone and smallest diameter consistent with other circuit parameters. This is in good agreement with the experiment from all points of view.

The authors acknowledge the support of the Air Force Office of Scientific Research and of the National Science Foundation. They thank H. Hellwig for advice

REFERENCES

1. F.L. Walls, P.H. Handel, R. Besson and J.J. Gagnepain: "A New Model relating Resonator Volume to 1/f Noise in BAW Quartz Resonators", Proc. 46. Annual Frequency Control Symposium, pp.327-333, 1992.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

IRREVERSIBLE FREQUENCY VARIATIONS IN TIME OF PRECISION QUARTZ CRYSTAL UNITS

A.G.SMAGIN

"Crystal" Co., 105425, Moscow, 9th Parkovaya str. 47-1-29, Russia

Abstract

General ageing reasons, inherent to all piezoelectric resonators, fall into the following groups: subsurface, structural and contact ones. High-Q and high-stability quartz crystal units have irreversible frequency variation with time due to structural reasons only. Long-term frequency instability due to ageing of precision quartz crystal oscillators constitutes the value of $1\text{-}2\text{x}10^{-12}$ per day, which agrees well with theoretical results at temperatures of the order of 310 K.

1.Changes in subsurface crystal layer

The dominating influence on irreversible variation of quartz crystal physical properties exert mechanical treatment processes and, consequently, changes in subsurface crystal layer. During crystal cutting by a diamond saw strong effects of structural deformation arise propagating to a considerable depth: turning of crystal cells occurs and macro- and microcracks appear in it as a consequence. Crystal lapping and crystal polishing, to a lesser extent, are also accompanied by nonelastic processes occurring in the crystal lattice. Then in the process of work or storage of quartz the shear regions of the crystal lattice transit from a metastable state to the main one inherent to the crystal prior its mechanical treatment. Such relaxational process proceeds spontaneously or under heat energy action liberated in quartz, which in a vibrational state (the process accelerates in this case).

At the same time microcrack selfhealing can take place in quartz crystals , if the surface is clean, i.e. at the surfaces of selfhealing structures no particles exist screening intermolecular forces action and no monomolecular films exist. For selfhealing, however, definite conditions are necessary: ion surface cleaning and high vacuum.

Resonator ageing is mainly a consequence of mechanical crystal treatment, therefore ageing processes can be brought to minimum by removing all disturbed crystal layers, i.e. by approximating the surface to a monomolecular structure by an asymptotic, and more precisely, by a limiting lapping and polishing.

Asymptotic methods of crystal surface treatment proposed by us, make it possible to approximate maximally to an undamaged crystal structure, which ensures quality factor Q elevation and ageing lowering by two orders of magnitude, any other conditions being equal. Thus, in accordance with the data of Frequency and Time Service high-stability oscillators of State Frequency Standard of Russia, in which the frequency control has been effected by precision resonators developed by us [1] had an ageing of $1 \div 2\text{x}10^{-11}$ per day (Fig.1) and by the end of 1962 — up to $1\text{x}10^{-12}$. Earlier (before 1959) the same oscillators with the same crystal bars, but treated by using old technology had an ageing of $1\text{x}10^{-9}$ per day.

The ageing processes considered apply to piezoelectric quartz with the surface free from a metal film. If the quartz is metallized, resonator ageing can be also due to changes of thin film physical properties (density, elasticity) and stresses at the crystal-metal film interface. The presence of thin metal films at the surface of quartz causes adsorption of gases from the environment and their absorption during metallization process, metal atoms migration at the piezoelectric element surface, in those moving

Warner [2] has shown that quartz crystal units with metal electrodes directly applied at a piezoelectric element have an ageing of $1\text{x}10^{-10}$ per month. This resonator class has become the main one in resonator developments for systems, especially.

2. Contact and structural changes

In the process of diamond saw and abrasive particles action not only a crystal surface is destructed, but a crystal lattice is distorted to a large depth, as a result of which its potential energy increases. Ageing mechanism consists in the fact, that with the elapse of

0-7803-1945-1/94 $3.00 © 1994 IEEE

time potential energy stored by the crystal lattice decreases. This phenomenon, being relaxational by its nature, is subordinate to Maxwell's equation and changes in accordance with the exponential law. We note, that the above is true with respect to structural changes rather than to surface ones where several different in regularities processes influence the ageing.

Different variations of a random character can influence the main frequency variation at the initial ageing stage caused by phenomena occurring in places of a quartz contact with a crystal holder (phenomenon of "rest" of metal supports, grinding-in effect of crystal holder, particle migration). If the conditions under which the quartz crystal unit operates are invariable, and in the crystal itself and associated devices no jumpwise changes take place the exponential law of ageing should be observed.

In parallel with relaxation mechanism of crystal lattice stresses, diffusive mechanisms exert their influence on irreversible quartz crystal frequency variation even in the vicinity of room temperatures.

Heat motion in crystals represents a totality of vibrations with a consecutive number of elementary movements of equilibrium states over lattice sites in the case of vacant atom site movement or over interstitials in the case of dislocated atoms. If the real vacancy concentration exceeds the equilibrium one, their coagulation takes place with the formation of microcavities (internal vacancy flow). The vacancy number can decrease as a result of establishing thermodynamic equilibrium state by means of excessive vacancies exit at the surface of a crystal (external flows), with this

$$\Delta\omega/\omega \sim V/V \sim N'/N,$$

where in accordance with Frenkel

$$N' = N \exp(-u/kT)$$

N' is the number of vacancies in a crystal;
N is the total number of atoms;
u is the vacancy formation energy;
k is the Boltzmann constant;
T is the absolute temperature.

Application of electrical voltage and, consequently, of a mechanical stress to a crystal causes dislocation movement. During their intersection with each other the moving dislocations leave vacancy chains and those of interstitial atoms after them. For the process of vacancies exit at the crystal surface only vacancy-atom interactions are important. Relation of probabilities interaction vacancy-vacancy p' and vacancy-atom p is proportional to p'/p, i.e. p'/p N'/N=n. As a result of such concentration of excessive vacancies at a temperature of 310 K the process of their exit at the crystal surface is more probable compared with the microcavity formation process.

Calculations of the time of establishing equilibrium vacancy concentration by way of their exit at the crystal surface have been made by Grois, Member of our Laboratory [3]. He has used the value u=0.65 eV from the paper of Shmatov & Grin describing definition of internal friction in metals and has obtained the value $n \approx 10^{-11}$. From my experiments on internal friction in quartz crystals, caused by vacancy mechanism, u=1.14 eV and $n \approx 10^{-12}$ [4]. Improved value of vacancy exit at the surface is $t = 7.67 \times 10^4$ s.

Figure 2 shows ageing results of precision quartz crystal units at 2 MHz, 5th overtone, at piezoelectric elements of which Ag- and Au- films are vacuum evaporated with the thickness of 2000 Å.

Piezoelectric elements have been subjected to asymptotic surface treatment methods and chemical etching and also to a thermal treatment together with the vacuum bulb with continuous pumping out residual gases from the latter.

Based on the analysis of numerous experimental data obtained (Cady, Mason, Mitchell, Dobie, Scheibe, Pritchard, Becker, Zelenka and those of Warner, expecially) and our own, a systematic investigation of phenomena mechanisms connected with irreversible quartz crystal unit frequency variation in time has been undertaken and hierarchial structure levels of ageing phenomena have been proposed.

The general ageing reasons, common with all piezoelectric resonator classes fall into the three groups as follows:

1. Subsurface ones — originating mechanical stresses forming microcracks after crystal treatment; destruction of the surface layer, microcracks healing under definite conditions under action of intermolecular forces;
adsorption by crystal of moisture and gases from the environment; metal film oxidation; lift-off and exit of adsorbed particles from this layer; change in physical properties (density, compliance) of thin films; change in stresses at the crystal metal film interface; particle migration at the crystal surface.

2. Structural ones - changes in orientation of shear regions, originated as a result of mechanical treatment, their transition from metastable to the main state, characteristic for thermodynamic equilibrium (relaxation phenomena);
compensation of surface forces by volume forces;
impurity diffusion in structural channels and impurity

exit at the crystal surface; closing of cavities originated as a result of coagulation of condensated vacancies; exit of excess vacancies at the crystal surface; anharmonic crystal oscillations.

3. Contact ones - change in resonator stable state due to piezoelectric element interaction with the holder and environment; effect of quartz surface and "grinding-in" metal supports of crystal holder; rest phenomena of quartz supporting elements;
change in gap as a result of metal ageing;
migration of quartz crystal particles to the glass bulb walls and backwards; change of vacuum in the bulb, where the resonator is located.

Experimental results have shown that subsurface, contact and certain structural ageing causes can be eliminated or considerably reduced by using asymptotic crystal surface treatment methods.
High-Q and high-stability crystal units have irreversible frequency variation with time due to structural reasons only: (microstress relaxation of the crystal lattice, diffusion and selfdiffusion of atoms and vacancies). Due to fact that the mechanisms of their ageing in crystal structure are known, they subordinate to the Maxwell and Frenkel formulae, the ageing regularities have an exponential character. Quartz crystal controlled oscillators with such resonators have an instability, including ageing of the order of $1x10^{-12}$ per day under stationary conditions and of the order of 10^{-10} in moving systems under the most difficult conditions.

Oscillators of this kind have found a wide application not only for scientific purposes (astronomic observatories, geophysics etc.), but also in a new engineering: telecontrol, radionavigation (primarily a space one) supertelecommication and radiolocation systems with continuous radiation.

References

[1] Smagin A.G. Report of the Academy of Schience of the U.S.S.R., 1962, v. 143, № 2, p.323 (in Russian).
[2] Warner A.W. Bell Syst. Tech. Journal, 1960, v.39, № 5, p.1193.
[3] Grois O.Sh. Radioengineering and Electronics, 1962, v.7, issue 4, p.702 (in Russian).
[4] Smagin A.G. Reports of the Academy of Science of the U.S.S.R., 1978, v. 240, № 6, p.1351 (in Russian).

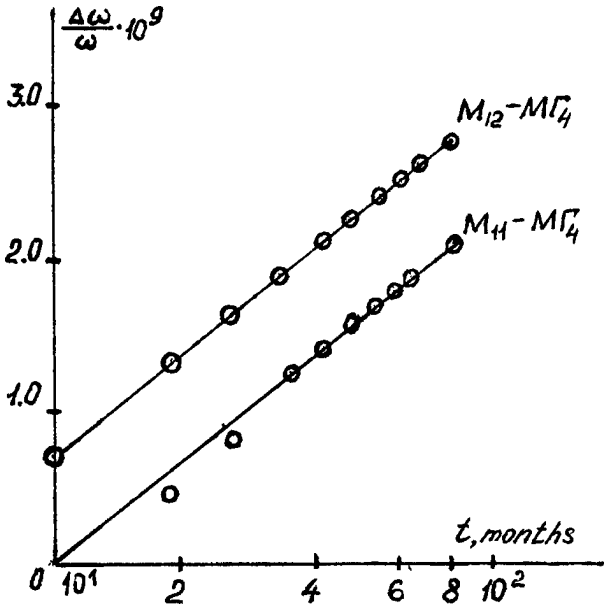


Figure 1 - Dependence of irreversible frequency variation of precision quartz crystal units vs time.

Piezoelectric elements - quartz crystal bars have been suspended by means of thin kapron fibres and have been driven at a freqency of 100 kHz (2nd mechanical overtone).

Comparision of crystal oscillators M_{11} and M_{12} has been made by using a molecular oscillator.

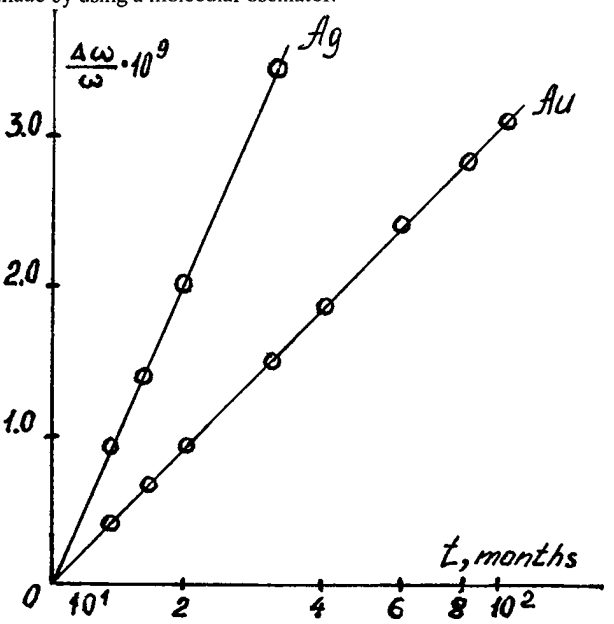


Figure 2 - Dependence of irreversible frequency variation of precision quartz crystal units vs time.

Piezoelectric elements - quartz crystal lenses driven at the frequency of 2 MHz (5th mechanical overtone).

Comparison of quartz crystal oscillators has been made by means of Hydrogen Standard.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

STUDY ON CHARACTERIZATION OF FREQUENCY STABILITY IN TIME DOMAIN

Sun Hongwei

Shaanxi Astronomical Observatory, Academia Sinica
Lintong, 710600 Shaanxi, P. R. of China

Abstract

In this paper, it is discussed that the relationships and physical meaning between Allan variance and standard variance, as well Allan variance defect. In addition, we proposed a method of frequency stability characteriztion in time domain.

Introduction

In time and frequency measurements, since the colored noises exist in atomic clocks, frequency instability does not observe normal distribution like general random error. Its statistics average value changes via time. With addition of calculating *times, its error sum is divergent. Hence, classical statistics method, i.e., standard* variance can not characterize frequency stability. Usually, people use Allan variance. In the article, we studied the problem about characteriztion of frequency stability in *time domain, and drawn some results.*

Analysis of relationships between variances

In error theory, standard variance is often used to describe random variable departure scope relative true average value, defination is

$$\sigma^2=(1/f_0)^2\langle(f_i-\bar{f})^2\rangle \qquad (1)$$

where f_0 is nominal frequency; f_i is ith *time measure frequency;* $\langle\ \rangle$ *denotes* average value of infinite time measurement, and

$$\bar{f}=\lim_{n\to\infty}\frac{1}{N}\sum_{i=1}^{N}f_i$$

Actually, measurement times is impossible infinite, can we substitute it by following formula

$$\sigma_N^2=[1/(N-1)f_0^2]\sum_{i=1}^{N}(f_i-\bar{f}_N)^2 \qquad (2)$$

where:

$$\bar{f}_N=(1/N)\sum_{j=1}^{N}f_i$$

0-7803-1945-1/94 $3.00 © 1994 IEEE

As a available statistics variable, the more measurement times, the more precise and the lesser the error. Thus we can use finite times to replace approximately infinite situation.

For frequency standard, as there is flicker noise, (1) is divergent, standard variance can not be replaced with (2). Next, we observe (2) from another aspect.

Marking:

$$\sigma^2(N,T,\tau)=[1/(N-1)f_0^2]\sum_{i=1}^{N}(f_i-\bar{f}_N)^2$$

where N is measure times; τ is sample time; $T-\tau$ is measure time interval.

Assuming that we measure M groups frequency value, and there are N times measure in each group. The standard variance in jth group is

$$\sigma_j^2(N,T,\tau)=[1/(N-1)f_0^2]\sum_{i=1}^{N}[f_i(j)-\bar{f}_N(j)]^2$$

Using $\sigma_j^2(N,T,\tau)$, j = 1, 2, ..., M, we may calculate its average value:

Marking:

$$[\sigma^2(N,T,\tau)]_m=(1/M)\sum_{j=1}^{M}\sigma_j^2(N,T,\tau)$$

Thus we get two sequences:

$$\sigma^2(2,T,\tau),\sigma^2(3,T,\tau),\ldots,$$

$$\sigma^2(N,T,\tau),\ldots \quad (3)$$

$$[\sigma^2(N,T,\tau)]_1,[\sigma^2(N,T,\tau)]_2,\ldots,$$

$$[\sigma^2(N,T,\tau)]_m,\ldots \quad (4)$$

sequence (3) limit value is standard variance, (4) limit value is Allan variance. On condition that there are flicker noises, (3) is divergent, which shows that (2) can not be thought an approximate value to the standard variance. While sequence (4) convergent, which shows that (2) may be thought:

$$[\sigma^2(N,T,\tau)]_1$$

descibing Allan variance. Hence, we have drawn that if the N is limited, standard variance differ from Allan variance; if N is limitless, limit value of Allan variance equals standard variance. So, broadly speaking, standard variance is a particular case of Allan variance.

We can also prove that if the fluctuation frequency is independent random variable, of which standard variance exist, it equals Allan variance. But, some random variable can not be regarded as independent, though its standard variance exists.

Both variances succinct meaning

Standard variance is a departure relative true average value, Allan variance is departure relative average value of N times measure, and value in difference time is a departure around difference average value, that is, the departure is same, but the average value is fluctuation, which is right caused by the chronic change component of flicker frequency noise.

Allan variance defect

From the analysis previously, we are aware of that Allan variance is born of standard variance, in order to overcome divergent problem of standard variance.

We know, when an expectation and a standard variance of random variable exist, and the mersure times is limited, standard variance of unbiased estimate is

$$\sigma^2=[1/(N-1)f_0^2]\sum_{i=1}^{N}(f_i-\bar{f}_N)^2$$

This is famous Bessel formula. But, it is directly quoted by Allan variance definition, although the premise changes. Several problems about mathematics, physical and logic in characterizing stability by Allan variance derive from the weak point.

As standard variance dose not exist in atomic frequency standard, we use a statistics according to groups, we have proved the formula

$$\sigma^2=(1/M)\sum_{i=1}^{N}(1/N)(f_{ji}-\bar{f}_{jN})^2(1/f_0^2)$$

being optimum unbiased estimate. In this formula a coefficient is 1/N, while in Allan variance is 1/(N-1). Only using 1/N can we get the optimum unbiased estimate. Only M may use finite to take the place of infinite, N can not. When the N is relatively big, the difference between 1/N and 1/(N-1) is small. Nevertheless, in two-sample Allan variance N equals 2, this difference is very large.

We also know,

$\langle(Y_2-Y_1)^2\rangle$

denotes a statistics for the difference of two value in successive measurement.

$\langle[(Y_2-Y_1)/2]^2\rangle$

denote a statistics for departure of two value relative average value.

In Allan variance definition,

$\langle(Y_2-Y_1)^2/2\rangle=\langle[(Y_2-Y_1)/\sqrt{2}]^2\rangle$

statistics variable is two value in successive two value difference divided by $\sqrt{2}$. This is what statistics, its physical meaning is obscure.

Allan variance is a departure degree of statistics value relative group average value, while in two-sample Allan variance the meaning should be a departure degree of statistics value relative average value in two values. If in this definition the cofficient were 1/N, the particular case belongs to general case, there is no mistake in logic. But, it is 1/(N-1), so the definition dose not conform to logic.

A proposal

In many occasions of frequency standard application, we hope to know a set of frequency diserete characters. We can adopt a statistics method according to group. It is similiar to that of Allan variance, first,

to alter N-1 in Allan variance into N; second, to assume sample time to equal corresponding relative big average time, statistics time may tend to limitless. It is shown that the definition is no problem in maths.

Supposing we want to determine a frequency value, this is just as to determine the earth position in the Galaxy. It is not convenient to locate that directly. We may firstly find out the solar system position in the Galaxy, then to determine the earth position in the solar system. We compare average frequency of relative big time unit to the solar system, small time unit to the earth, namely, we consider the frequency of relative small sample time only in average frequency of relative big time unit. For example, the statistics in second frequency will obtain aid by minute average frequency, its meaning is second frequency fluctuation extent to surround minute average frequency, and all that, we may also know minute frequency fluctuation extent to surround hour average frequency; hour frequency fluctuation extent to surround day mean frequency and so on, thus we can prophesy second frequency value extent in any moment. However, Allan variance dose not possess the advantage.

Reference

[1] J.A.Barnes, et al., " Characterization of Frequency Stability " , IEEE Trans. Instrum. Meas., val. IM-20, No. 2.1971.

[2] P.Lesage and C.Audoin, "Characterization of frequency stability: uncertainty due to the finite number of measurements", IEEE Trans. Instrum. Meas., vol. IM-22, pp 157-161, June 1973.

[3] D. B. Percival, "The U. S.Naval Observatory Clock Time Scale", IEEE Trans. Instrum.Meas., val. IM-27, No.4, 1978.

[4] D. W. Allan, "Time and Frequency Metrology : Current Status and Future Consideration", Proc.5th European Frequency and Time Forum, pp. 1-9, Besancon, 1991.

[5] J. Groslambert, J. -J. Gagnepain, "A new filtered Allan variance and its application to the identification of phase and frequency noise sources", Proc.43rd Ann. Freq. Control Symp., pp.326-330, Denver, 1989.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DESIGN AND OPTIMIZATION OF LOW-NOISE OSCILLATORS USING NONLINEAR CAD TOOLS

Ulrich L. Rohde, Chao-Ren Change, Jason Gerber

Compact Software, Inc.
201 McLean Boulevard Paterson, NJ 07504

Abstract

The design of low phase noise oscillators has been both cultured for a long time and frequently labeled as "black magic". In this paper, I want to first show a novel mathematical approach which considers all noise contribution and allows both the calculation and optimization of the phase noise of an oscillator with arbitrary topologies. After explaining the physics and mathematical background, we will show three examples: a 10 MHz crystal oscillator, an 800 MHz VCO and a 39 GHz millimeter-wave oscillator [1].

Introduction

This paper presents the use of a novel algorithm for the computation of near-carrier noise in free-running oscillators using the harmonic balance (HB) technique. Traditional approaches relying on frequency conversion analysis are not sufficient to describe the complex physical behavior of a noisy oscillator. The accuracy of the approach is based on the dynamic range of the harmonic balance simulator and the quality of the parameter extraction for the active device. The algorithm described has also been verified with several examples up to millimeter wave. This is the only algorithm which provides a complete and rigorous treatment of noise analysis for autonomous circuits.

Noise Generation in Oscillators

The qualitative picture of noise generation in oscillators is very well known (D.B. Leeson, 1966). The physical effects of random fluctuations taking place in the circuit are different depending on their spectral allocation with respect to the carrier:

- Noise components at low frequency deviations: Frequency modulation of the carrier; Mean-square frequency fluctuation proportional to the available noise power

- Noise components at high frequency deviations: Phase modulation of the carrier; Mean-square phase fluctuation proportional to the available noise power

This paper will demonstrate that the same conclusions can be quantitatively derived from the HB equations for an autonomous circuit.

Equivalent Representation of a Noisy Nonlinear Circuit

A general noisy nonlinear network can be described by the equivalent circuit shown in Fig.1. The circuit is divided into linear and nonlinear subnetworks as noise-free multiports. Noise generation is accounted for by connecting a set of noise voltage and noise current sources at the ports of the linear subnetwork [2-10].

0-7803-1945-1/94 $3.00 © 1994 IEEE

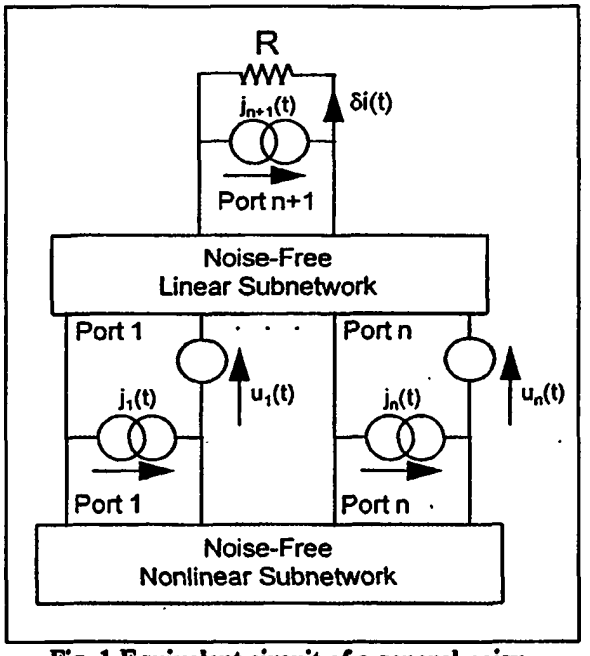


Fig. 1 Equivalent circuit of a general noisy nonlinear network.

Frequency Conversion Approach

The circuit supports a large-signal time-periodic steady state of fundamental angular frequency ω_0 (carrier). Noise signals are small perturbations superimposed on the steady state, represented by families of pseudo-sinusoids located at the sidebands of the carrier harmonics. Therefore, the noise performance of the circuit is determined by the exchange of the power among the sidebands of the unperturbed steady state through frequency conversion in the nonlinear subnetwork. Due to the perturbative assumption, the nonlinear subnetwork can be replaced with a multifrequency linear multiport described by a conversion matrix. The flow of noise signals can be computed by means of conventional linear circuit techniques.

The frequency conversion approach frequently used has the following limitations:

The frequency-conversion approach is not sufficient to predict the noise performance of an autonomous circuit. The spectral density of the output noise power, and consequently the PM noise computed by the conversion analysis are proportional to the available power of the noise sources.

- In the presence of both thermal and flicker noise sources, PM noise: raises as ω^{-1} for $\omega \to 0$; tends to a finite limit for $\omega \to \infty$.
- Frequency conversion analysis: correctly predicts the far-carrier noise behavior of an oscillator, and in particular the oscillator noise floor; does not provide results consistent with the physical observations at low deviations from the carrier.

This inconsistency can be removed by adding the modulation noise analysis. In order to determine the far-away noise using the autonomous circuit perturbation analysis, the following applies:

The circuit supports a large-signal time-periodic autonomous regime. The circuit is perturbed by a set of small sources located at the carrier harmonics and at the sidebands at a deviation ω from carrier harmonics. The perturbation of the circuit state $(\delta X_B, \delta X_H)$ is given by the uncoupled sets of equations:

$$\frac{\partial E_B}{\partial E_B}\delta X_B = J_B(\omega) \qquad (1)$$

$$\frac{\partial E_H}{\partial E_H}\delta X_H = J_H(\omega) \qquad (2)$$

where:

- E_B, E_H = vectors of HB errors
- X_B, X_H = vectors of state-variable (SV) harmonics (since the circuit is autonomous, one of the entries X is replaced by the fundamental frequency ω_0)
- J_B, J_H = vectors of forcing terms
- The subscripts B and H denote sidebands and carrier harmonics, respectively.

For a spot noise analysis at a frequency ω, the noise sources can be interpreted in either of two ways:

- Pseudo-sinusoids with random amplitude and phase located at the sidebands. Noise generation is described by Eq. (1) which is essentially a frequency conversion equation relating the sideband harmonics of the state variables and of the noise sources. This description is exactly equivalent to the one provided by the frequency-conversion approach. This mechanism is referred to as ***conversion noise*** [11-13].

Modulation Noise

Now we introduce the modulation noise. Sinusoids located at the carrier harmonics, randomly phase- and amplitude-modulated by pseudo-sinusoidal noise at frequency ω. Noise generation is described by Eq. (2), which describes noise-induced jitter of the circuit-state, represented by the vector δX_H. The modulated perturbing signals are represented by replacing the entries of J_H with the complex modulation laws. This mechanism is referred to as ***modulation noise***. One of the entries of δX_H is $\delta\omega_0$ where $\delta\omega_0(\omega)$= phasor of the pseudo-sinusoidal components of the fundamental frequency fluctuations in a 1 Hz band at frequency ω. Eq. (2) provides a frequency jitter with a mean square value proportional to the available noise power. In the presence of both thermal and flicker noise, PM noise raises as ω^{-3} for $\omega \to 0$ and tends to 0 for $\omega \to \infty$. Modulation noise analysis correctly describes the noise behavior of an oscillator at low deviations from the carrier and does not provide results consistent with physical observations at high deviations from the carrier. The combination of both phenomenon explains the noise in the oscillator shown in Fig. 2 where the near-carrier noise dominates below ω_X and far-carrier noise dominates above ω_X.

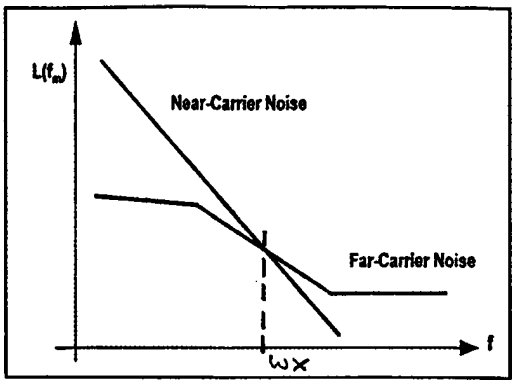


Fig. 2 - Oscillator Noise Components

Fig. 3 (itemized form) shows the noise sources as they are applied at the IF. We have artificially defined the low oscillator output as IF. This applies to the conversion matrix calculation.

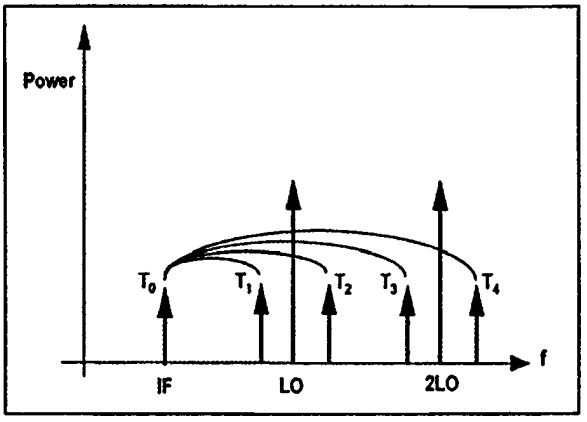


Fig. 3 - Noise Sources where the noise at each sideband contributes to the output noise at the IF through frequency conversion.

Fig. 4 shows the total contribution which have to be taken into consideration for calculation of the noise at the output. The accuracy of the calculation of the phase noise depends highly on the quality of the parameter extraction for the nonlinear device particularly the high frequency phenomena have to be properly modeled. In addition, the flicker noise contribution is essential.

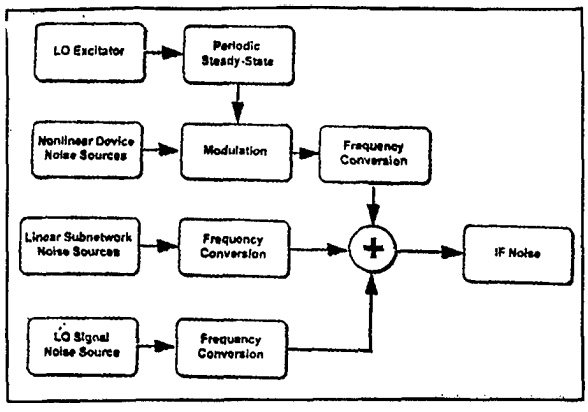


Fig. 4 - Noise Mechanisms

Fig. 5 shows the measured phase noise at 10 kHz of the carrier of an 800 MHz oscillator and the simulated result. The good agreement between measured and predicted data is worth noticing. At currents above 5 mA, the simulation becomes accurate as some breakdown effects occur which were not part of the simulation.

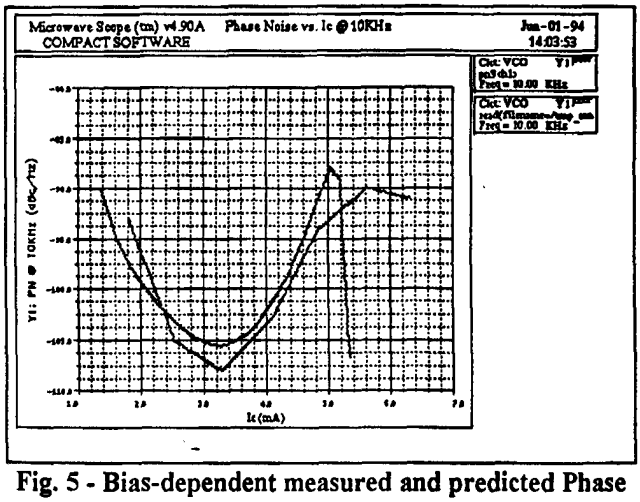


Fig. 5 - Bias-dependent measured and predicted Phase Noise at 10 kHz of the carrier of an 800 MHz oscillator

Conversion Noise Analysis

The actual mathematics used to calculate the noise result were incorporated in the harmonic balance simulator as follows:

k-th harmonic PM-noise (3)

$$\left\langle \left|\delta\Phi_k(\omega)\right|^2 \right\rangle = \frac{N_k(\omega) - N_{-k}(\omega) - 2\,\mathrm{Re}\left[C_k(\omega)\right]}{R\left|I_k^{SS}\right|^2}$$

k-th harmonic AM-noise (4)

$$\left\langle \left|\delta A_k(\omega)\right|^2 \right\rangle = 2\frac{N_k(\omega) - N_{-k}(\omega) + 2\,\mathrm{Re}\left[C_k(\omega)\right]}{R\left|I_k^{SS}\right|^2}$$

k-th harmonic PM-AM correlation coefficient (5)

$$C_k^{PMAM}(\omega) = \left\langle \delta\Phi_k(\omega)\delta A_k(\omega)^* \right\rangle = -\sqrt{2}\frac{2\,\mathrm{Im}\left[C_k(\omega)\right] + j\left[N_k(\omega) - N_{-k}(\omega)\right]}{R\left|I_k^{SS}\right|^2}$$

where

- $N_k(\omega)$, $N_{-k}(\omega)$ = noise power spectral densities at the upper and lower sidebands of the k-th harmonic
- $C_k(\omega)$ = normalized correlation coefficient of the upper and lower sidebands of the k-th carrier harmonic
- R = load resistance
- I_k^{ss} = k-th harmonic of the steady-state current through the load.

Modulation Noise Analysis

k-th harmonic PM-noise (6)

$$\left\langle \left|\delta\Phi_k(\omega)\right|^2 \right\rangle = \frac{k^2}{\omega^2}\mathbf{T}_F\left\langle \mathbf{J}_H(\omega)\mathbf{J}_H^t(\omega)\right\rangle\mathbf{T}_F^t$$

k-th harmonic AM-noise (7)

$$\left\langle \left|\delta A_k(\omega)\right|^2 \right\rangle = \frac{2}{\left|I_k^{SS}\right|^2}\mathbf{T}_{Ak}\left\langle \mathbf{J}_H(\omega)\mathbf{J}_H^t(\omega)\right\rangle\mathbf{T}_{Ak}^t$$

k-th harmonic PM-AM correlation coefficient (8)

$$C_k^{PMAM}(\omega) = \left\langle \delta\Phi_k(\omega)\delta A_k(\omega)^* \right\rangle = \frac{k\sqrt{2}}{j\omega\left|I_k^{SS}\right|^2}\mathbf{T}_F\left\langle \mathbf{J}_H(\omega)\mathbf{J}_H^t(\omega)\right\rangle\mathbf{T}_{Ak}^t$$

where

- $J_H(\omega)$ = vector of Norton equivalent of the noise sources
- T_F = frequency transfer matrix
- R - load resistance
- I_k^{ss} = k-th harmonic of the steady-state current through the load

Experimental Variations

We will look at three examples where we compare predicted and measured data. Fig. 6 shows the abbreviated circuit of a 10 MHz crystal oscillator. It uses a high-precision, high-Q crystal

made by companies such as Blily. Oscillators like this are made by several companies and are intended for use as both frequency and low-phase-noise standards. In this particular case, the crystal oscillator being considered is part of the HP 3048 phase-noise measurement system.

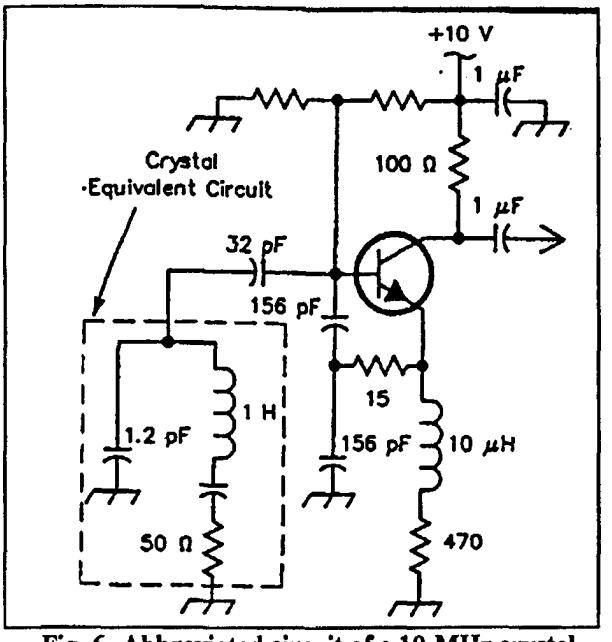


Fig. 6 -Abbreviated circuit of a 10-MHz crystal oscillator

Fig. 7 shows the measured phase noise of this frequency standard by HP, and Fig. 8 shows the predicted phase noise using the mathematical approached outlined above.

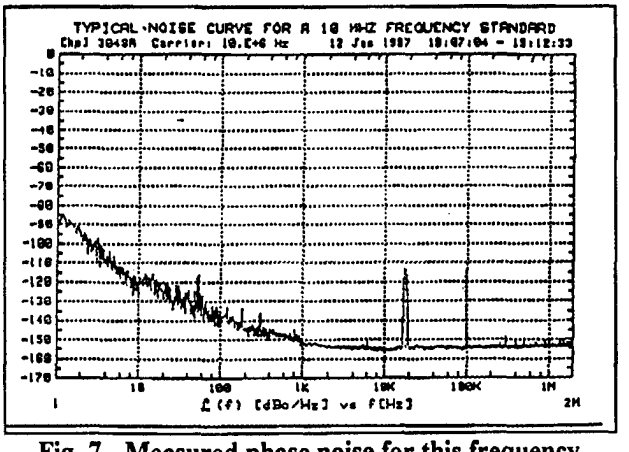


Fig. 7 - Measured phase noise for this frequency standard by HP

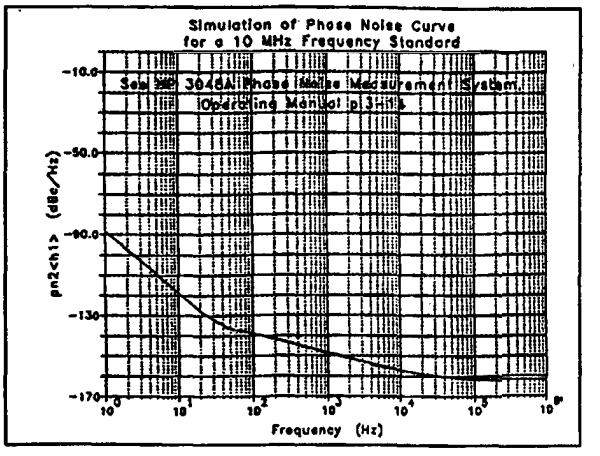


Fig. 8 - Simulated phase noise of the oscillator shown in Fig. 6

In corporation with Motorola, we incorporated an 800 MHz VCO. In this case we also did the parameter extraction with the Motorola transistor. Fig. 9 shows the Colpitts oscillator which uses RF feedback in the form of a 15Ω resistor and a capacitive voltage divider of 1pF each between base and the feedback resistor and the feedback resistor and ground. Also the tuned circuit is loosely coupled to this part of the transistor circuit.

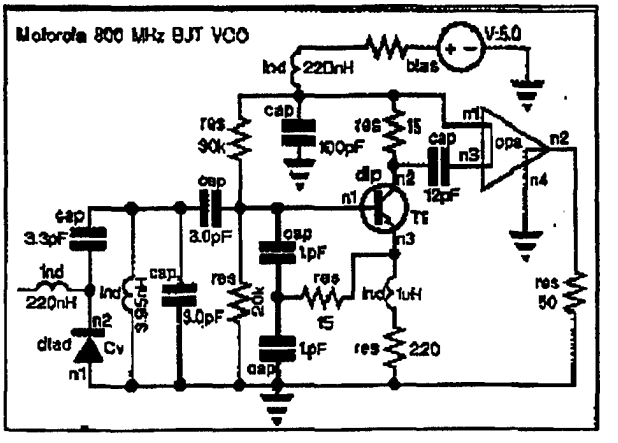


Fig. 9 - Colpitts Oscillator which uses RF feedback

Fig. 10 shows a comparison between predicted and measured phase noise for this oscillator.

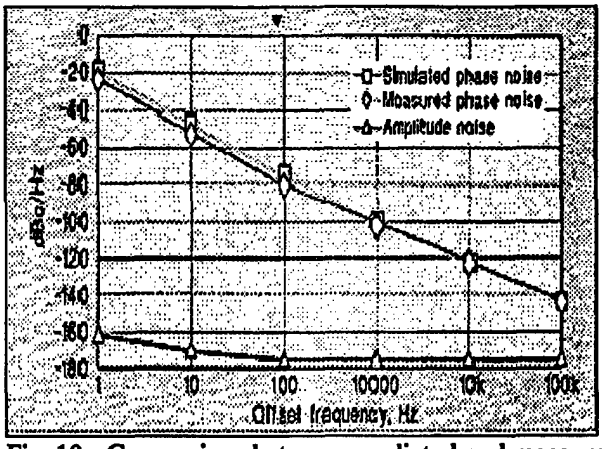


Fig. 10 - Comparison between predicted and measured phase noise for this oscillator

Finally, to extend the validity up to the millimeter-wave range, Fig. 11 shows a 39 GHz oscillator built by Texas Instruments and Fig. 12 shows the measured and predicted phase noise.

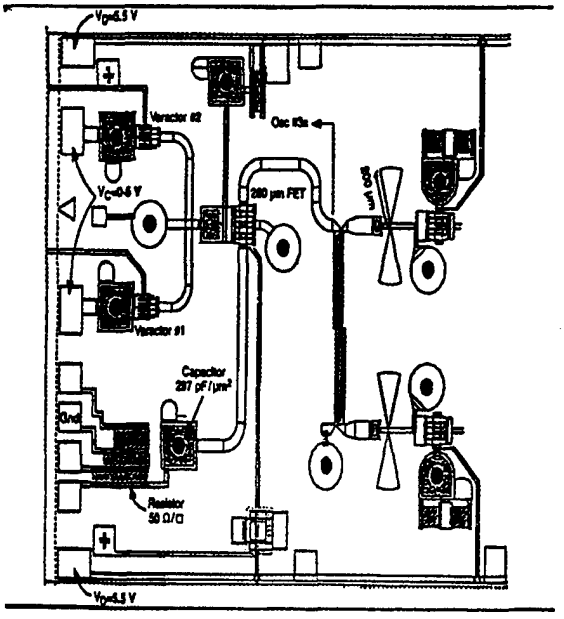


Fig. 11 -Layout of 39 GHz oscillator built by Texas Instruments and used as a test vehicle

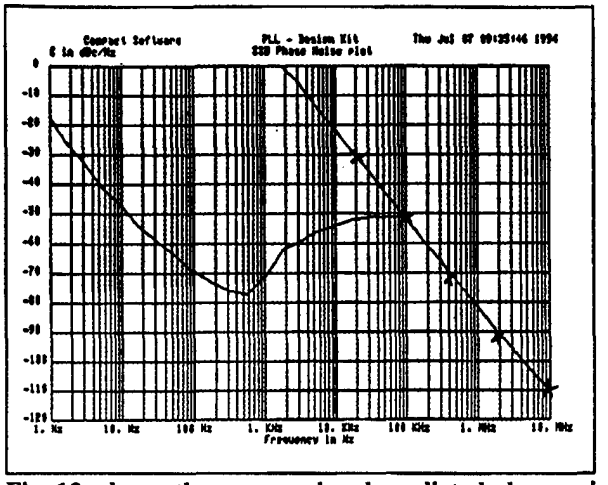


Fig. 12 -shows the measured and predicted phase noise of the TI oscillator including its phase noise when part of a PLL

Optimization of Phase Noise

By allowing the varying of the feedback and DC operating point, phase noise can be improved. The values where were allowed to vary were the capacitive feedback voltage divider, a negative feedback with starting value of 15Ω which reduces the AM-PM conversion and the emitter resistor which changes the DC bias and therefore the bias-depdnent flicker noise. Fig. 13 shows the improvement of the 800 MHz VCO as previously shown. While the close-in phase noise can be improved drastically by approximately 32 dB, the far-out noise at 20 MHz and further away is deteriorated. This is due to the resistive feedback which reduced the AM to PM conversion [14-16].

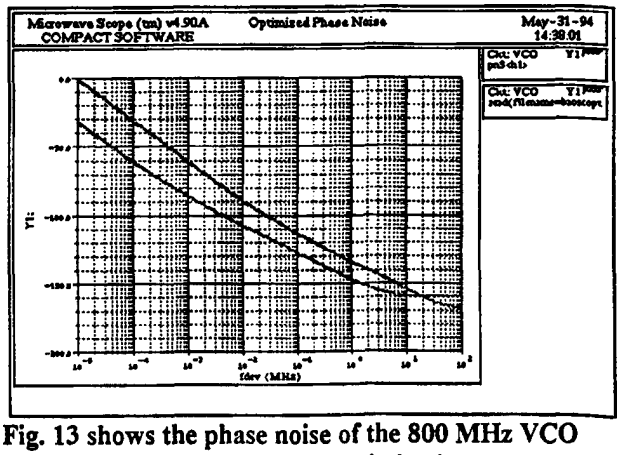


Fig. 13 shows the phase noise of the 800 MHz VCO before and after its optimization

Summary

This paper has demonstrated the use of harmonic balance simulation in the design of low phase noise oscillator. The simulator can be used to accurately predict close-in and far-out phase noise based on the contribution of the transistor, the loaded Q of the resonator, and in the case of a VCO, the contribution of the tuning diode.

This is the first time that a bias-dependent noise model for both FETs and bipolar transistors has been developed and used for this type of application where the optimization is based on the principle of allowing both the circuit components and the bias-point to be varied for best performance.

Practical examples ranging from a 10 MHz crystal oscillator to a 39 GHz millimeter-wave oscillator have been supplied. By demonstrating both the theoretical background and practical results, it has been shown that a reliable method for designing and optimizing for low phase noise has been successfully understood and incorporated in our commercial HB simulator, Microwave Harmonica.

Acknowledgement

This work, partly sponsored by the US Department of Defense MIMIC program, was performed in cooperation with Professor Vittorio Rizzoli's group of the Department of Electrical Engineering, Information and Systems Engineering at the University of Bologna, Italy and the Motorola Portable Products Division, Florida.

References

[1] U.L. Rohde, "All About Phase Noise in Oscillators", QEX

[2] U.L. Rohde, C.R. Chang and J. Gerber, "Parameter Extraction for Large Signal Noise Models and Simulation of Noise in Large Signal Circuits Like Mixers and Oscillators", Proceedings of the 23rd European Microwave conference, Madrid, Spain, Sept. 6-9, 1993.

[3] C.R. Chang, "Mixer Noise Analysis Using the Enhanced Microwave Harmonica", Compact Software Transmission Line News, June, 1992, Vol. 6, No. 2 pp. 4-9.

[4] T. Antognetti and G. Massobrio, Semi-Conductor Device Modeling with SPICE, McGraw-Hill, New York, NY, pp. 91 (1988).

[5] R.J. Hawkins, "Limitation of Nielsen's and Related Noise Equations Applied to Microwave Bipolar Transistors, and a New Expression for the Frequency and Current Dependent Noise Figure", Solid State Electronics, 20, pp. 191-196 (1977).

[6] Tzu-Hwa Hus, and Craig P. Snapp, "Low Noise Microwave Bipolar Transistor Sub-Hall-Micrometer Emitter Width" IEEE Transactions Electron Devices, Vol. ED-25, pp. 723-730, June 1978.

[7] R.A. Pucel and U.L. Rohde, "An Accurate Expression for the Noise Resistance R_n of a Bipolar Transistor for Use with the Hawkins Noise Model", IEEE Microwave and Guided Wave Letters, Vol. 3, No. 2, February, 1993, pp. 35-37.

[8] G. Vendelin, A.M. Pavio, and U.L. Rohde, Microwave Circuit Design: Using Linear and Nonlinear Techniques, John Wiley and Sons, New York (1990).

[9] R.A. Pucel, W. Struble, Robert Hallgren and U.L. Rohde, "A General Noise De-embedding Procedure for Packaged Two-Port Linear Active Devices", IEEE Transactions on Microwave Theory and Techniques, Vol. 40, No. 11, Nov. 1993, pp 2013-2024.

[10] U.L. Rohde, "Improved Noise Modeling of GaAs FETS, Parts I and II: Using an Enhanced Equivalent Circuit Technique", Microwave Journal, November, 1991, pp 87-101 and December 1991, pp 87-95, respectively.

[11] V. Rizzoli, F. Mastri and C. Cecchefti, "Computer-Aided Noise Analysis of MESFET and HEMT Mixers", IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-37, pp 1401-1410, Sept. 1989.

[12] V. Rizzoli and A. Lippadni, "Computer-Aided Noise Analysis of Linear Multiport Networks of Arbitrary Topology", IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-33, Dec. 1985, pp 1507-1512.

[13] V. Rizzoli, F. Mastri, and D. Masotti, "General-Purpose Noise Analysis of Forced Nonlinear Microwave Circuits", published in Military Microwave, 1992.

[14] U.L. Rohde, "Key Components of Modern Receiver Design, VHF Conference, Weinheim, Germany, Sept. 17-18, 1993.

[15] C.R. Chang and U.L. Rohde "The Accurate Simulation of Oscillator and PLL Phase Noise in RF Sources", Wireless '94 Symposium, Santa Clara, CA Feb. 15-18, 1994.

[16] U.L. Rohde, "Oscillator Design for Lowest Phase Noise", Microwave Engineering Europe, May, 1994, pp 31-40.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

NONLINEAR OSCILLATOR DESIGN FOR MAXIMUM POWER

M.C.E. Yagoub, H. Baudrand, *Senior Member IEEE*

Laboratoire d'Electronique, ENSEEIHT, 2 rue Camichel, 31071 TOULOUSE -FRANCE-

ABSTRACT

In this paper a new approach to the design of nonlinear microwave oscillator is presented which allows optimum large-signal performance of the oscillator to be predicted.

The procedure consists in determining a surface including all the combinations of allowed powers available at the terminal ports of the nonlinear device in order to deduce its optimum loads and large-signal efficiency.

The technique has been applied to practical diode oscillator circuits. Comparisons with previously published results from the technical literature and with experimental data permit to validate the method.

I - INTRODUCTION

Most of the available approaches to the design of nonlinear microwave circuits are based on the principle of dividing equivalent circuit into two groups those which vary under large-signal conditions (nonlinear subnetwork) and those which do not vary (linear subnetwork) **[1]-[4]**.

Usually, the analysis requires the knowledge of the embedding impedances seen by the nonlinear subnetwork at a finite number of frequencies (equivalent Thevenin circuit of the linear subnetwork). So, the topology of the linear subnetwork is fixed *a priori* and for optimum performance, the analysis loop is completed by an outer optimization loop.

Moreover, in the case of diode oscillator design, the retained topology must take into account the criteria for oscillation which is traditionally performed using Kurokawa's approach where the nonlinear device is modeled as a one-port in which the real part of the impedance is negative **[5]**.

The proposed method, already applied to diode frequency multipliers and mixers **[6]**, is an optimum approach which calculates the extremum allowed powers at terminal ports of the nonlinear subnetwork.

The procedure consists in determining in the power space, a surface including all these powers and then permits to deduce the optimum embedding impedances without any assumption for objective functions choice and independently of the linear subnetwork.

This approach can be assimilated to an optimum large-signal analysis. A small-signal analysis must then be performed in accordance with the oscillation condition.

The design of practical diode oscillators has allowed to obtain computed results which have shown good agreement with those given previously in the references and experimental data.

II - PRINCIPLE OF THE METHOD

In the case of large periodic signals, any nonlinear device can be simulated as a multifrequency multiport network where each port (k) corresponds to the kth harmonic of the fundamental input frequency (a port "0" can be added for bias).

For a finite number n of harmonics, if V_k and I_k represent respectively the voltage and current at a port (k), the characteristics of such a component can be expressed by means of the following (n+1) equations :

$$I_k = f_k(V_o, ..., V_n) \qquad k = 0, ..., n \qquad (1)$$

Let P_k be the available power at the input of kth port. It is said that power is allowed at port (k) when there exist active or passive loads at each port so that the network can absorb the power P_k.

The aim of this method is to define, in the power space, a volume inside which the points have coordinates equal to the allowed powers at different ports of the nonlinear lossy network (In the lossless case, this surface is described by the well-known Manley-Rowe relations).

This volume is surrounded by a surface called "characteristic surface". The approved approach for its determination is to consider beforehand the tangent planes to the surface.

0-7803-1945-1/94 $3.00 © 1994 IEEE

In order to present a graphic illustration of the proposed method, let us consider the case of a two-port network (fig.1).

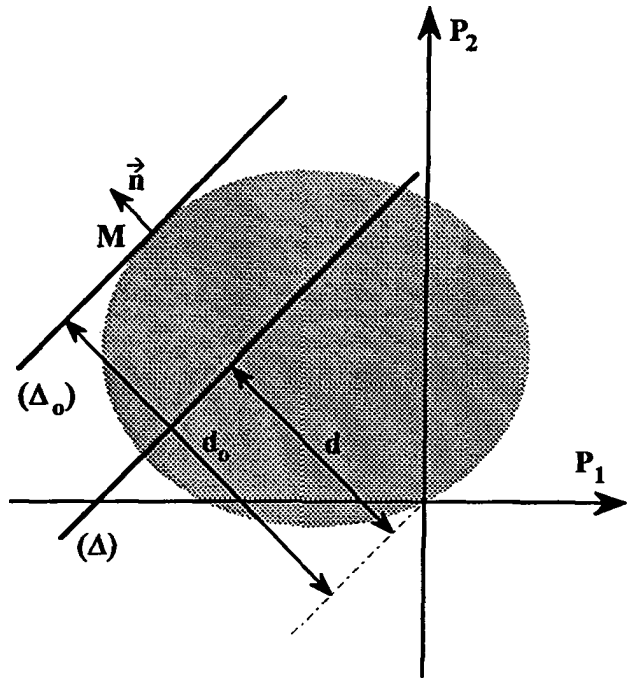


Fig.1: Representation of the characteristic surface for a two-port network.

The equation of any plane (Δ) cutting the surface can be written as:

$$\sum_{k=0}^{n} \lambda_k P_k = d(V_o,..,V_n) \qquad (2)$$

where **d** is the distance to the origin and the λ_i the coordinates of the normal $\vec{n}$ to the plane.

The final purpose is to examine if there exist points of (Δ) corresponding to allowed powers. For this, each port (k) of the nonlinear network was loaded by a voltage generator, of value V_k, which delivers the current I_k and the power P_k.

Any variation of voltage implies a variation of the distance d. Now on the contour of the characteristic surface, at contact point M_o, this distance has a maximum d_o. So, the points of this contour represent allowed powers independent of any variation of the voltages δV_k. These powers are the extremum allowed powers at terminal ports of the network.

These maximum powers verify the equation of any tangent plane (Δ_o) to the characteristic surface:

$$\sum_{k=0}^{n} \lambda_k P_k = d_o(V_o,..,V_n)$$
$$\Rightarrow \sum_{k=0}^{n} \lambda_k \delta P_k = 0 \qquad (3)$$

Moreover, from the relations (1), it is possible to determine an admittance matrix **[Y]** relating the vector currents **[δI]** and voltages **[δV]** :

$$[\delta I_0,...,\delta I_n]^t = [Y][\delta V_0,...,\delta V_n]^t$$
$$\Rightarrow \quad [\delta I] = [Y][\delta V] \qquad (4)$$

where $[\ ]^t$ represents the transpose matrix.

The combination of relations (3) and (4) gives, independently of any arbitrary variation of voltage δV_k (at any port of the multiport network), the following nonlinear system :

$$\mathbf{Re}\left\{ [I]^+[\lambda]+[V]^+[\lambda][Y] \right\} = [0] \qquad (5)$$

where **Re** denotes the real part and $[\ ]^+$ the conjugate transpose matrix.

This system describes all the combinations of extremum allowed powers available at the terminal ports of the nonlinear subnetwork.

To each set of values (λ_o, ..., λ_n) corresponds a solution of voltage vector and therefore a combination of extremum allowed powers. So, the intersection points of the characteristic surface with the tangent planes are obtained by variation of the λ_k.

Moreover, with the convexity of the surface **[6]**, this solution is unique.

III - VALIDATION OF THE METHOD

The method has been implemented in a symbolic programming language (MAPLE), which makes it possible to determine algebraic expressions **[7]**. This implementation can directly translate into FORTRAN code the nonlinear system and then it has been resolved on a personal computer.

To demonstrate the viability of the technique, the described method has been applied first to analyse a practical two Resonant Tunneling Diodes - RTD - oscillator circuit **[8]**.

Since the equivalent large-signal circuit and current-voltage relationship of RTD can also be used for Tunnel Diodes, it is useful to compare RTD's with the more familiar Tunnel diode.

For more accuracy, the equivalent circuit model of each RTD is the same as the one taken by Yang *et al.* [8] (as illustrated in fig.2), where a capacitance (C) in shunt with a conductance (G) is in series with a parasitic series resistance (R_S).

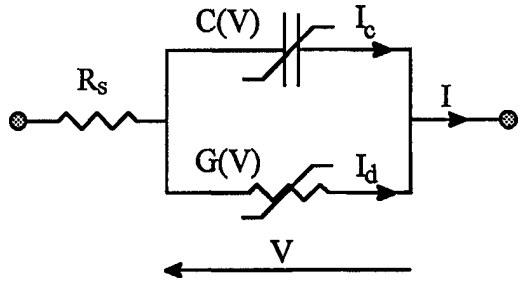


Fig.2 : Equivalent circuit of a single resonant tunneling diode.

The total current (I) is composed of a displacement current (I_c) due to the junction capacitance and a conduction current (I_d) [8], [9], i.e.:

$$I_c(V) = C_{jo}\left\{1-\frac{V(t)}{\Phi}\right\}^{-\gamma}\frac{dV(t)}{dt} \qquad (6)$$

$$I_d(V) = \frac{(I_{pd}-I_{vd})(V(t)-V_{vd})^4}{(V_{vd}-V_{pd})^5}\cdot \left\{5(V(t)-V_{pd})-(V(t)-V_{vd})\right\}+I_{vd} \qquad (7)$$

where the subscripts "pd" and "vd" represent respectively the peak and valley values of the diode current-voltage characteristics.

The total voltage is given as the sum of DC and RF voltage, so the relations (1) can be solved to find out the steady-state voltage across the diode.

In order to present a three-dimensional characteristic surface illustration, we have shown, in the fig.3, the variation of the powers for bias, fundamental frequency (ω) and second harmonic (2ω).

The curve of output power P_1 of the fundamental frequency against the bias input power P_{dc} is obtained as the locus of the intersection points of this surface with the plane ($OP_{dc}P_1$) when the other powers are set to zero (fig.4) - in the three-dimensional illustration, the plane is located at P(2ω) = 0 -.

The tangent line (D) parallel to OP_{dc} gives the maximum output power P_{1max} for the optimum input power $P_{dc\ opt}$.

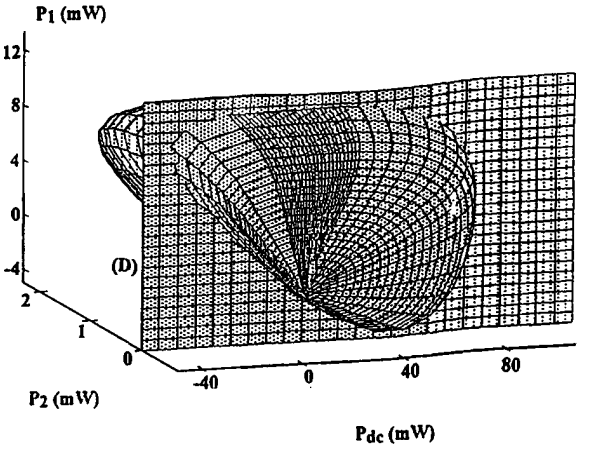


Fig.3 : Characteristic surface for the three first powers corresponding to bias power [P_{dc}] fundamental frequency power [P_1 (ω)] and second harmonic power [P_2 (2ω)] with the plane $P_2 = 0$.

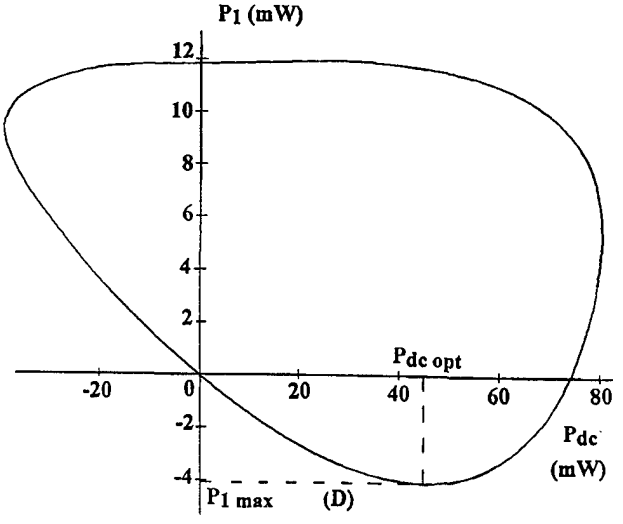


Fig.4 : Variation of the output power [P_1 (ω)] in function of the bias input power [P_{dc}] with the power [P_2 (2ω)] = 0.

The results, shown in table I, are in agreement with those reported by [8].

	this method	ref [8]
$P_{dc\ opt}$ (mW)	44.86	44.65
$P_{l\ max}$ (mW)	4.06	3.84
Efficiency (%)	9.04	8.60

Table I: Comparisons between the computed results and those given in [8].

In a second step, the procedure was used to realize a 12 GHz GaAs tunnel diode oscillator. The specified data of the used diode are:

$$C_{jo}\sqrt{\Phi} = 1.2\ \text{pF}.\ V^{1/2} \qquad V_{bias} = 0.12\ V$$

$$V_{vd} = 0.42\ V \qquad V_{pd} = 0.10\ V$$

The experimental optimum efficiency, obtained by variation of the bias voltage around the optimum computed data, confirm the validity of the theory (table II). Hence it is expected that oscillator design based on this method would be quite close to the first approach.

	computed	measured
$P_{dc\ opt}$ (mW)	0.75	0.74
$P_{l\ max}$ (mW)	0.31	0.28
Efficiency (%)	41.33	37.84

Table II: Computed and measured values for the designed diode oscillator.

IV - CONCLUSION

A new analytical approach for optimum oscillator design is described. The method has been verified for large-signal diode oscillator circuits and the obtained results are in good agreement with those cited from the literature and experimental measurements.

This agreement shows that the proposed optimum method provides an interesting tool for the design of a variety of nonlinear circuits.

REFERENCES

[1] V. Rizzoli, A. Neri, " State of the art and present trends in nonlinear microwave CAD techniques, " *IEEE Trans. Microwave Theory Tech.*, vol. MTT-36, pp 343-365, Feb. 1988.

[2] Y. Xuan, C. M. Snowden, " A generalized approach to the design of microwave oscillators, " *IEEE Trans. Microwave Theory Tech.*, vol. MTT-35, pp 1340-1347, Dec. 1987.

[3] V. Rizzoli, A. Costanzo, C. Cecchetti, " Numerical optimization of microwave oscillators and VCOs, " in *Proc. IEEE MTT-S Int. Microwave Symp.* Dig., Atlanta, 1993, pp 629-631.

[4] R.D. Martinez, R.C. Compton, " A general approach for the S-parameter design of oscillators with 1 and 2-port active devices, " *IEEE Trans. Microwave Theory Tech.*, vol. MTT-40, pp 569-574, March 1992.

[5] K.K. Kurokawa, " Some basic characteristics of broad band negative resistance oscillator circuits, " *Bell Sys. Tech. J.*, vol. 48, no. 6, pp 1937-1955, July 1969.

[6] M.C.E. Yagoub, H. Baudrand, " Optimum design of nonlinear microwave circuits, " *IEEE Trans. Microwave Theory Tech.*, to be published.

[7] B.W. Char, K.O. Geddes, G.H. Gonnet, B.L. Leong, M.B. Monagan, S.M. WATT, *Maple V Library reference manual*, New-York: Springer, 1991.

[8] C.C. Yang, D.S. Pan, " Theoretical investigations of a proposed series integration of resonant tunneling diodes for millimeter-wave power generation, " *IEEE Trans. Microwave Theory Tech.*, vol. MTT-40, pp 434-441, March 1992.

[9] J.A. Narud, S.S. Meyer " A polynomial approximation for the tunnel diode characteristic, " *IEEE Trans.*, vol. CT-10, n°4, p 520, 1963.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

COMPUTER AIDED DESIGN OF QUARTZ CRYSTAL OSCILLATORS

R. Brendel*, G. Marianneau*, T. Blin*, M. Brunet**

* Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon - 32 avenue de l'Observatoire - 25000 Besançon - France
** Centre National d'Etudes Spatiales - Dept AE/TTL/TF - 18 avenue E. Belin - 31055 Toulouse Cedex - France

Abstract

The last development of a simulation program specialized in the quartz crystal oscillator analysis is presented in this paper. The simulator actually being developed is using the full nonlinear Barkhausen criterion method. It consists in finding the frequency ω_0 and the amplitude u_0 which nullify both the real and imaginary parts of a characteristic complex polynomial $P(u,j\omega)$ describing the oscillator behavior. The main nonlinearities come from the amplifying transistor described by using large signal admittance parameters $y(u)$ obtained by means of an analog circuit simulator (SPICE). The paper presents the method used to derive and code the characteristic polynomial coefficients. This method has been successfully implemented for a Colpitts oscillator and is currently being used to build an oscillator library covering the most widely used structures. The validity and the predictive power of the model have been checked by using a circuit board and the comparison between experimental results and simulation is presented and discussed.

Introduction

A preliminary version of the nonlinear simulation program for quartz crystal oscillators has been presented a few years ago at the 45th Frequency Control Symposium [1]. The principle of that kind of simulator, its advantages with respect to other programs and its main features have then been explained. Since then, many efforts have been expended to improve and wide the program. For example, the simulator initially developed for a main frame computer is now running on a PC computer. Furthermore, new features have been added to the software, for example the excitation power of the resonator is now available. Also, the first version was limited to only one oscillator circuit, the present version is being completed by an oscillator library covering the most widely used structures as found in the literature [2, 3, 4]. Each component of the library is in fact an independent module called by the main program.

The analysis and coding of a new oscillator circuit starts with a preliminary manual analysis necessary to derive the oscillation condition. In the steady state, this condition is expressed under the form of a characteristic polynomial in the harmonic variable $j\omega$ the root of which is the oscillation frequency ω_0. If this preliminary analysis is not carefully performed, it may result in an unnecessarily high polynomial degree due to a lack of simplification giving rise to a too large amount of code and an inaccurate solution [5]. Then it is essential to formulate the oscillation condition under a general reduced form which can be used for any oscillator structure.

In the previous version of the program, the nonlinear behavior of the transistor was described by a set of tables coming from an electrical simulator (SPICE), these tables set up once and for all contained the admittance parameter value as a function of the signal amplitude for various bias and temperature conditions. The actual value of the y-parameters was calculated by interpolating in these tables. Although this method can be used for a single oscillator circuit, it no longer suits the present purpose because of the wide range of transistors and their possible bias condition. To overcome this difficulty, the large signal parameters are now calculated at run time : as soon as the bias and temperature condition is known, SPICE supersedes the program and simulates the large signal transistor behavior for this condition. Then, the program supersedes SPICE back and uses the obtained results to set up the large signal admittance parameters.

At the same time the software was developed, a circuit board of a test oscillator was built to check for the validity of the simulation results. First experiments have shown strong discrepancies between experimental and expected results, after careful examination, it appears that the calculated resonator drive level is much higher than usually assumed. At this level it is no longer justified to neglect the amplitude-frequency effect also called isochronism defect. The program has been modified to take this effect into account and the calculated results are now in good agreement with the experiment.

0-7803-1945-1/94 $3.00 © 1994 IEEE

Reduced forms of an oscillator

As a consequence of the Thevenin's theorem it can be shown that by using simple admittance and impedance transforms like series or parallel combination or Kennely's transforms, the dynamical equivalent circuit of any oscillator using a single transistor can be reduced to one of the equivalent T- or Π- reduced forms represented in Figs. 1 and 2.

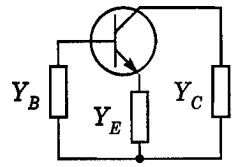


Fig. 1 : *T-reduced form of an oscillator*

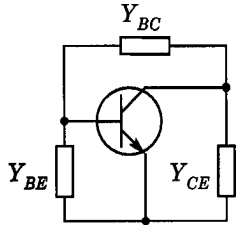


Fig. 2 : *Π-reduced form of an oscillator*

Moreover, the Π-form can be obtained from the T-form by using Kennely's transforms :

$$Y_{BE} = \frac{Y_B \cdot Y_E}{\Sigma_y}, \quad Y_{CE} = \frac{Y_C \cdot Y_E}{\Sigma_y}, \quad Y_{BC} = \frac{Y_B \cdot Y_C}{\Sigma_y} \tag{1}$$

$$\Sigma_y = Y_B + Y_E + Y_C$$

The admittances Y_B, Y_C, Y_E are functions of the circuit components like resistances, capacitances and inductances under the form of rational functions in the Laplace's variable s.

The transistor itself is represented by its large signal admittance parameters equivalent circuit in which the y-parameters are function of the signal amplitude [4]. We thus obtain the dynamical equivalent Π-form represented in Fig. 3.

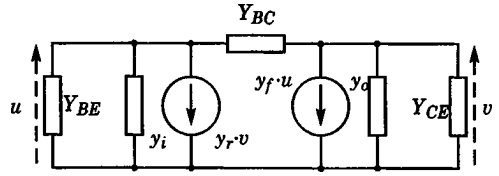


Fig. 3 : *Dynamical equivalent Π-form*

Eventually, by using simple transforms (2) the circuit can be reduced to the dynamical form represented in Fig. 4.

$$\begin{aligned} y'_i &= y_i + Y_{BE} + Y_{BC} \\ y'_f &= y_f - Y_{BC} \\ y'_r &= y_r - Y_{BC} \\ y'_o &= y_o + Y_{CE} + Y_{BC} \end{aligned} \tag{2}$$

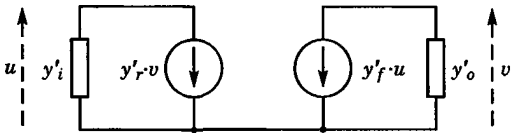


Fig. 4 : *Dynamical reduced form of an oscillator*

Oscillation condition

By applying Kirchhoff's law to both input and output parts of the circuit (Fig. 4) we obtain the following system

$$\begin{aligned} y'_i u + y'_r v &= 0 \\ y'_f u + y'_o v &= 0 \end{aligned} \tag{3}$$

The system (3) admits a nontrivial solution only if its determinant is null, this gives the oscillation condition of the circuit

$$y'_i y'_o - y'_r y'_f = 0 \tag{4}$$

Substituting y' in (4) by their expression (2) leads to another form of the oscillation condition

$$\Delta + Y_{BC} \cdot \sigma + y_i Y_{CE} + y_o Y_{BE} + Y_{CE} Y_{BC}$$

$$+ Y_{BE} Y_{BC} + Y_{BE} Y_{CE} = 0$$

$$\Delta = y_i\ y_o - y_r\ y_f$$
$$\sigma = y_i + y_o + y_f + y_r \qquad (5)$$

Δ is the determinant and σ is the sum of the transistor equivalent circuit y-parameters as represented in Fig. 3.

It should be noted here that the transistor y-parameter have a real and an imaginary parts so that each of them can be considered as a parallel combination of a conductance and a capacitance both being nonlinear functions of the signal amplitude. We can write in a general form

$$y_n = g_n + s\,.\,C_n \quad (n = i, f, r, 0) \qquad (6)$$

From their definition recalled in Eq. (7), it results that y_i and y_f depends on the input voltage u while y_r and y_o depends on the output voltage v (see Fig. 5).

$$y_i = \left(\frac{i_i}{u}\right)_{v=0} \qquad y_r = \left(\frac{i_i}{v}\right)_{u=0}$$
$$y_f = \left(\frac{i_o}{u}\right)_{v=0} \qquad y_o = \left(\frac{i_o}{v}\right)_{u=0} \qquad (7)$$

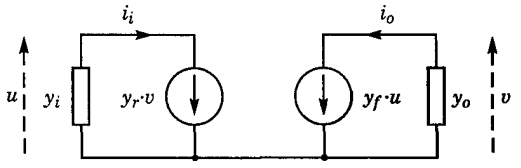


Fig. 5 : *y-parameter representation of a transistor*

A third equivalent form of the oscillation condition can be obtained by substituting the admittances Y_{BE}, Y_{CE} and Y_{BC} by their expression (1)

$$\Delta.\Sigma_y + \sigma.Y_B.Y_C + y_i\,.\,Y_C.Y_E + y_o.Y_B.Y_E + Y_B.Y_C.Y_E = 0 \qquad (8)$$

Characteristic polynomial

Let's demonstrate the derivation method on a simple example. Figure 6 represents a simplified form of a Colpitts oscillator, for the sake of simplicity, the resonator equivalent circuit will be reduced to the series resonant circuit (Cq, Lq, Rq) and the transistor reverse admittance parameter y_r will be neglected so that, by comparing figures 2 and 6 we readily obtain relations (9) while Δ and σ of equations (5) reduce to (10).

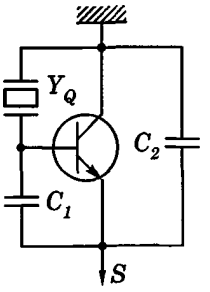


Fig. 6 : *Simplified form of a Colpitts oscillator*

$$Y_{BC} = Y_Q = \frac{s\,.\,C_q}{1 + s.C_q\,R_q + s^2.L_q\,C_q}$$
$$Y_{BE} = s\,.\,C_1$$
$$Y_{CE} = s\,.\,C_2 \qquad (9)$$
$$\Delta = y_i\,.\,y_o$$
$$\sigma = y_i + y_o + y_f$$

Substituting Eqs. (6) into (10) leads to the following expressions :

$$\Delta = \delta_o + s\,.\,\delta_1 + s^2\,\delta_2$$
$$\sigma = \sigma_o + s\,.\,\sigma_1 \qquad (10)$$

with

$$\delta_o = g_i\ g_o$$
$$\delta_1 = C_i\ g_o + C_o\ g_i$$
$$\delta_2 = C_i\ C_o \qquad (11)$$
$$\sigma_o = g_i + g_o + g_f$$
$$\sigma_1 = C_i + C_o + C_f$$

and substituting back Eqs. (9), (10) and (11) into the oscillation condition (5) gives the polynomial in the Laplace's variable s

$$a_o + s\,a_1 + s^2\,a_2 + s^3\,a_3 + s^4\,a_4 = 0 \quad (12)$$

in which the following relations have successively been used :

$$\begin{aligned}
\beta_1 &= \delta_1 + C_2\,g_i + C_1\,g_o \\
\beta_2 &= \delta_2 + C_2\,C_i + C_1\,C_o + C_1\,C_2 \\
\gamma_1 &= \sigma_o\,C_q \\
\gamma_2 &= (\sigma_1 + C_1 + C_2)\,C_q \\
a_o &= \delta_o \\
a_1 &= \delta_o\,C_q\,R_q + \beta_1 + \gamma_1 \\
a_2 &= \delta_o\,L_q\,C_q + \beta_1\,C_q\,R_q + \beta_2 + \gamma_2 \\
a_3 &= \beta_1\,L_q\,C_q + \beta_2\,C_q\,R_q \\
a_4 &= \beta_2\,L_q\,C_q
\end{aligned} \quad (13)$$

Equation (12) represents the characteristic polynomial of the oscillator, as we can see, each coefficient of this polynomial is expressed as a function of the component value of the circuit. It is easy to understand that the degree and the complexity of the polynomial coefficients are strongly depending on the component number. Even in the simple case presented here, the full development of the polynomial coefficients results in intricate expressions hardly manageable without error. Nevertheless, by using computer aided symbolic calculation [6], it has been possible to automate the derivation and coding in high level language of the polynomial coefficients so as to minimize the risk of error. In a general way, the analysis of an oscillator structure starts with admittance expressions like those of Eqs. (9) and (10), next, the admitances are combined step by step so as to reduce the circuit to its T- or Π-form the corresponding equivalent admittances being expressed as fractional functions of the previous step. Eventually, the characteristic polynomial coefficients are obtained under the following general form analog to Eq. (12)

$$\sum_{k=0}^{K} a_k\,s^k = 0 \quad (14)$$

In a similar manner, it is possible to express the excitation power of the resonator as a fractional function of the circuit admittances and the Laplace's variable s so that it is possible to take the amplitude-frequency effect into account, this calculation also needs the transistor gain A_v which can be obtained from either one of Eqs. (3)

$$A_v = \frac{v}{u} = -\frac{y_i'}{y_r'} = -\frac{y_f'}{y_o'} \quad (15)$$

Substituting y' by their expression (2) leads to :

$$A_v = -\frac{y_i + Y_{BE} + Y_{BC}}{y_r - Y_{BC}} = -\frac{y_f - Y_{BC}}{y_o + Y_{CE} + Y_{BC}} \quad (16)$$

As the transistor y-parameters, A_v depends on the signal amplitude u.

Oscillation frequency and amplitude

When the steady state is reached, the oscillation characteristics are described by Eq. (14) in which the Laplace's variable s is changed into the harmonic variable $j\omega$, this splits Eq. (14) into two equations corresponding to the real and imaginary parts :

$$\sum_{m=0}^{M} a_m(u).\omega^m \;,\; \sum_{n=0}^{N} \beta_n(u).\omega^n = 0 \quad (17)$$

It is obvious that these two equations cannot be simultaneously satisfied if the coefficients are constant. In fact, some of them, namely those which include the transistor y-parameters, depend on the signal amplitude u. So that the problem amounts to find the pair (u_o, ω_o) satisfying both equations (17). This technique implies to know the nonlinear behavior of the transistor. In the present version, the eight required nonlinear functions, that is the real and imaginary parts of the four complex y-parameters are obtained by using the electrical simulator SPICE [7]. Before the oscillator analysis begins, the program gives SPICE the transistor bias conditions and asks it to perform a set of transient sinusoidal analyses with increasing signal amplitudes. This enables the program to calculate the y-parameter as a function of the amplitude by processing the SPICE's output data as described in refs [1,9]. (See the flow chart represented in Fig. 9).

Once the transistor nonlinear behavior is known, the solution of equations (17) is obtained by following the same procedure as in ref. [1] with the addition of the resonator frequency-amplitude effect as described in the next section.

Resonator frequency-amplitude effect

It is well known that the resonant frequency of a quartz resonator depends on the drive level.

For instance, a 5 MHz AT-cut crystal exhibits a fractional frequency change of about $0.2/A^2$ while the change of an SC-cut is roughly ten times smaller [2]. Figure 7 shows the amplitude response curves of a 5 MHz fifth overtone AT-cut resonator for different

levels of drive [8] and Fig. 8 represents the isochronism defect, that is the fractional resonant frequency change as a function of the excitation power, of a 10 MHz third overtone SC-cut resonator.

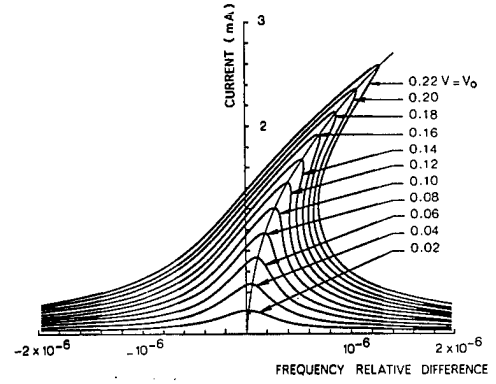


Fig. 7 : *Amplitude response curves of a 5 MHz 5th overtone AT-cut resonator for different level of drive (after [8])*

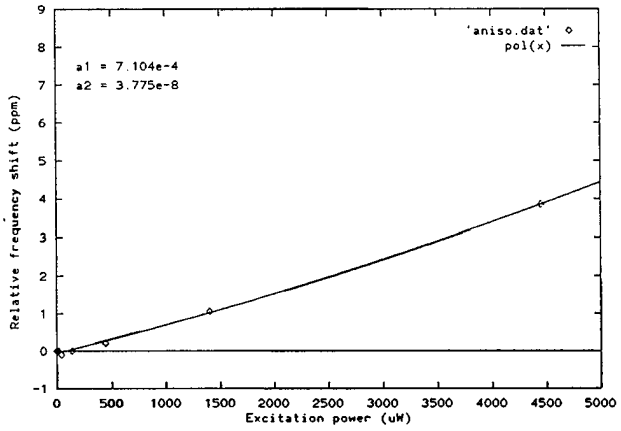


Fig. 8 : *Isochronism defect of a 10 MHz 3rd overtone SC-cut resonator*

This feature is introduced in the program by means of the resonant frequency of the resonator which is expressed as a function of the excitation power

$$f_q(P) = f_q(0)\,(1 + a_1 P + a_2 P^2) \qquad (18)$$

$f_q(0)$ represents the resonator frequency at very low drive level while a_1 and a_2 are the coefficients required by the program to describe the anisochronism defect. In the present case, they have been obtained from a second order polynomial regression of the experimental data as shown in Fig. 8.

Program features

Figure 9 presents the general flow chart of the present version of the program and Fig. 10 details a bit more the last step, that is the calculation of the oscillation characteristics. Let's demonstrate the calculation principle on an example. Figure 11 represents the Nyquist's plane of a Colpitts oscillator characteristic polynomials, i.e. the parametric representation of Eqs. (17), each curve represents the imaginary part-vs-real part with respect to the frequency for different amplitude values.

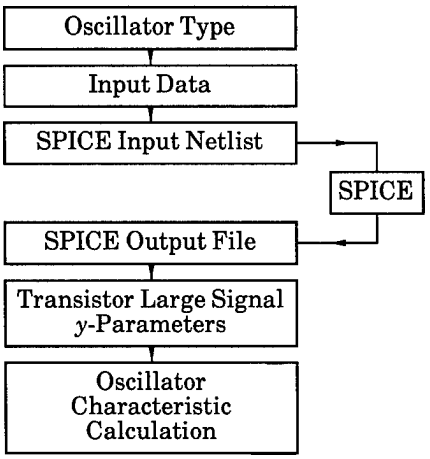


Fig. 9 : *General flow chart of the program*

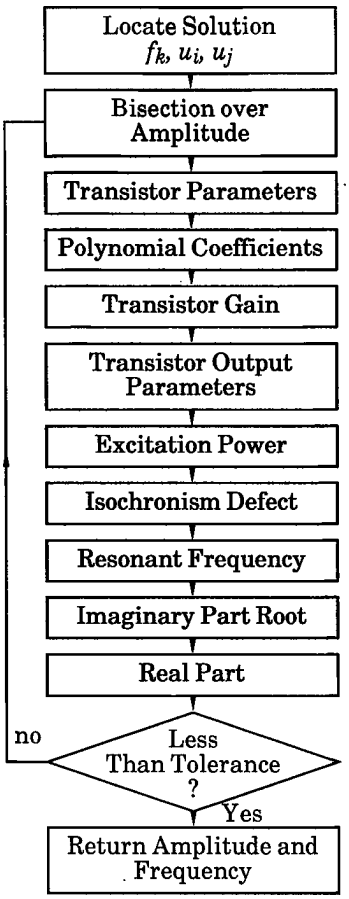


Fig. 10 : *Calculation of the oscillator characteristics*

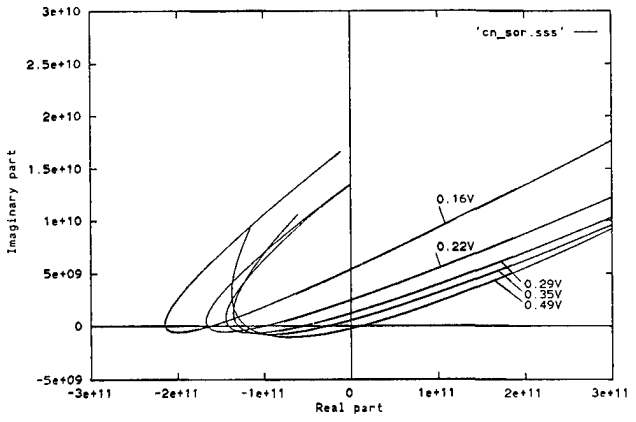


Fig. 11 : *Colpitts oscillator characteristic polynomials*

We can see that as the amplitude increases, the characteristic polynomials get closer and closer to the origin until they pass beyond it, then, the expected polynomial is located between the two polynomials passing on each side of the origin and the corresponding amplitudes u_i and u_j can be used as starting points for the bisection process. In the same way, the frequency f_k of the nearest polynomial point from the origin can be used as starting point for the imaginary part root calculation algorithm. Once the solution has been located this way, it is improved by using the algorithm described on the flow-chart in Fig. 10.

Because the characteristic polynomial coefficients are expressed as functions of all the circuit components, it is possible for the program to calculate the influence of a change of any component value on the oscillation amplitude and frequency as well as on the resonator excitation level. In the same way, it is possible to get the sensitivity of these quantities to small component variation. In addition of the rated value the user can specify the tolerance and the temperature coefficient of each component in the input data file. This allows the program to calculate the induced dispersion and to perform worst case analysis. The temperature behavior of the resonator is described by the four coefficients of the frequency temperature curve cubic regression so that the frequency variation can be calculated over a given temperature range.

Experimental verification

The only way to get confidence in the simulation program consists in comparing the predicted values with experimental results. To this end, the Colpitts oscillator represented in Fig. 12 has been built on a circuit board so that some components can be easily altered.

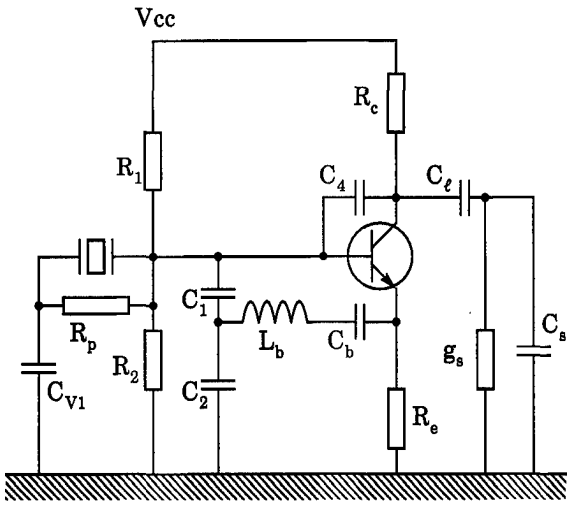


Fig. 12 : *Test Colpitts oscillator*

The main difficulty in this experiment is to keep a good enough frequency stability so that tiny deviations induced by small component variations can be evidenced. The largest part of frequency fluctuations comes from the influence of temperature variations on the quartz resonator.

So as to prevent the frequency measurement accuracy from these perturbating effects, the resonator is kept at ist turn-over point in a temperature controlled oven, the remaining circuitry being placed nearby outside. Figure 13 shows the experimental setup used as well as the block diagram of the frequency measurement system.

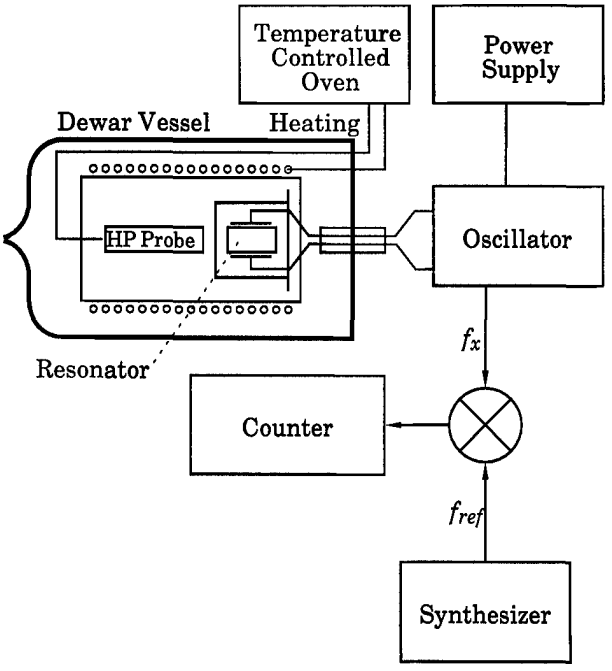


Fig. 13 : *Experimental setup*

The change of a component value is performed by switching on or off a set of components put in parallel, this procedure permits to increase or decrease the rated value of the components in the range 0 to 10 %. The results obtained by changing some selected components of the circuit represented in Fig. 12 are shown in Figs. 14 to 17, in these figures, the small squares are experimental data while solid lines are simulation results. the diamond-marked lines are obtained by giving the resonator parallel capacitance its nominal 1.2 pF value in the simulation input data file. But because of the stray capacitance due to the distance from the circuit to the resonator, the overall parallel capacitance has been measured to be about 10 times larger, another simulation with a 11 pF value has been performed and the plus-marked lines show the results thus obtained.

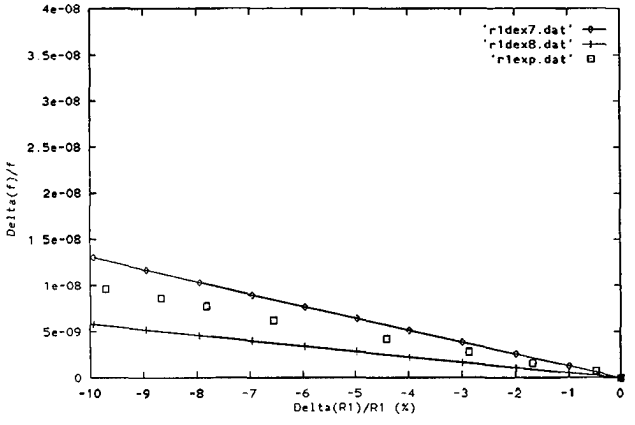


Fig. 14 : *Simulated and experimental fractional frequency change when varying the transistor base resistance (see Fig. 12)*

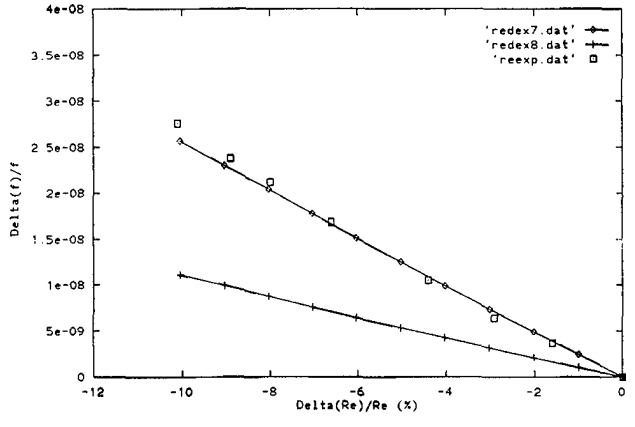


Fig. 15 : *Same as Fig. 14 when varying the transistor emitter resistance*

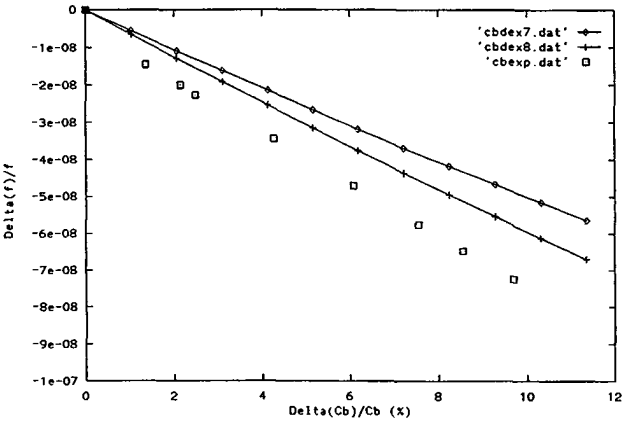


Fig. 16 : *Same as Fig. 14 when varying the selection mode capacitance*

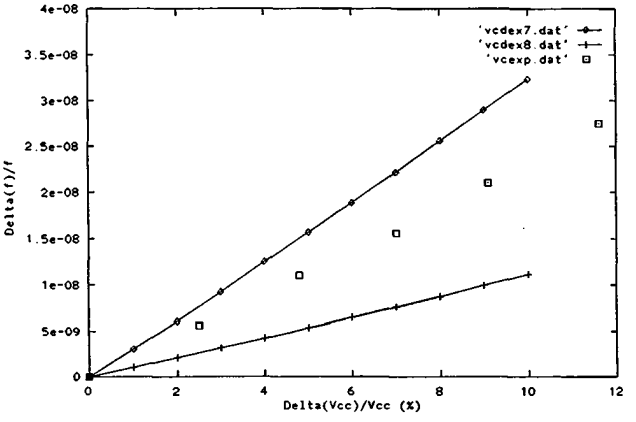


Fig. 17 : *Same as Fig. 14 when varying power supply*

The experimental results show that in all cases the sense of variation and the order of magnitude of the frequency shift have been properly predicted. This was not always the case before the resonator amplitude-frequency effect was taken into account. Nevertheless, it can be seen that the value of a component, for instance the resonator parallel capacitance, can noteacibly modify the sensitivity with respect to another component.

So, it is possible that some parasitic components like stray capacitances which are not taken into account in the model or not accurately measurable might affect in some way the sensitivity with respect to other components. Hence, the accuracy of the simulation more likely stems from the component value uncertainties rather than from the lack of precision of the model.

Conclusion

The nonlinear analysis used to describe the transistor and the oscillator behavior enables the program to take into account important effects like the frequency amplitude effect.

Experiments have shown that the simulation program thus developed can accurately predict the real oscillator features. Remaining discrepancies between simulated and experimental results might be due to the lack of knowledge of the real component values due to parasitic components the value of which is difficult to assess. So as to extend the scope of the program, an oscillator library is actually being implemented. Nevertheless, this solution is not quite satisfactory because the user is still constrained by the necessarily limited number of designated structures.

A more ambitious project dealing with a topological analysis of any oscillator structure described by a netlist is now undertaken. The idea is to automate the initial analysis by parsing the input netlist and identifying the impedance transforms to be performed in order to put the initial structure to its reduced form. Applying the general oscillation condition gives the symbolic form of the characteristic polynomial coefficients which can eventually be used to analyze the oscillator behavior.

Acknowledgements

This work has been supported by CNES and DGA under contract # 962/CNES/91/1476/00.

References

[1] R. Brendel, F. Djian, E. Robert, "High precision nonlinear computer modelling technique for quartz crystal oscillators", Proc. 45th Annual Symposium on Frequency Control, pp. 341-351 (1991).

[2] B. Parzen, A. Ballato, "Design of crystal and other harmonic oscillators", John Wiley (1983).

[3] R.J. Matthys, "Crystal oscillator circuits", J. Wiley (1983).

[4] M.E. Frerking, "Crystal oscillator design and temperature compensation", Van Nostrand Reinhold Co (1978).

[5] R. Brendel, T. Blin, G. Marianneau, "Modélisation non linéaire des oscillateurs ultrastables pour applications aérospatiales", Final Report Contract # 962/CNES/91/1476/00 (1994).

[6] A.C. Hearn, "Reduce user's manual V.3.4", Rand Publication (1991).

[7] L.W. Nagel, "SPICE2 : A computer program to simulate semiconductor circuits", Memorandum ERL-M520, University of California, Berkeley (1975).

[8] J.-J. Gagnepain, R. Besson, "Nonlinear effect in piezoelectric quartz crystals", Physical Acoustics, vol. XI, Academic Press (1975).

[9] R. Brendel, F. Djian, J.-J. Gagnepain, "Modélisation des effets d'environnement sur les caractéristiques des oscillateurs à quartz ultrastables", Final Report Contract # 832/CNES/88/5374/00 (1991).

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A NEW 5 AND 10 MHz HIGH ISOLATION DISTRIBUTION AMPLIFIER

C. W. Nelson[▲], F. L. Walls[●], M. Sicarrdi[♦], A. De Marchi[♦]

[▲]SpectraDynamics,
530 South 44th Street, Boulder CO 80303

[●]National Institute of Standards and Technology,
325 Broadway, Boulder CO 80303

[♦]Dip. di Electronica, Politecnico di Torino,
C. so Duca degli Abruzzi 24, Torino, Italy

Abstract

Increasing performance demands made by precision timing have made NIST's present 5 MHz distribution amplifier system obsolete. A new design providing improved phase stability with temperature, harmonic purity, and phase noise is presented. By building on previous designs, a modified cascode amplifier was created with performance increases of more than 10 fold in phase noise, temperature coefficient and isolation. An input-output isolation of 140 dB and channel to channel isolation of greater than 125 dB was achieved. Phase noise performance of -152 dBc/Hz at 1 Hz with a noise floor of -170 dBc/Hz was also achieved. Input and output matching provide average return losses greater 30 dB. Harmonics are all -45 dBc or better at an output of +13 dBm, and the temperature coefficient of output phase is less then 1 ps /°C.

Introduction

The improved performance of new frequency standards as well as the advent of the trapped ion clocks, places a greater demand on signal distribution and measurement systems. In order to reach fractional frequency stabilities on the order of 10^{-16} in 10^4 seconds, timing errors due to transmission and measurement must be less then 1 ps. At 5 MHz this corresponds to holding phase variations to under 30 microradians for about 3 hours. The old design, summarized in table 1, fails this criteria mainly in aspects of temperature coefficient and timing errors due to voltage standing wave ratio (VSWR). In order to approach the error budget of 1 ps a new set of design criteria was created and is shown in table 2.

Table 1, Summary of old design performance

isolation	125 dB
phase noise at 1 Hz	-142 dBc/Hz
phase noise floor	-165 dBc/Hz
temperature coefficient	12 - 15 ps/°C
input/output return loss	25 - 13 dB
harmonics	-20 dBc
power consumption	160 mW

Table 2, Performance goal of new design.

isolation	120 dB
phase noise at 1 Hz	-145 dBc/Hz
phase noise floor	-165 dBc/Hz
temperature coefficient	1 ps/°C
input/output return loss	35 dB
harmonics	-45 dBc
power consumption	<160 mW

The development and performance of an amplifier that meets these requirements will be described.

History

The amplifier presented here has a long evolution beginning in 1976 with the work of Gray and Glaze [1]. The amplifier shown in figure 1, consisting of two common base bipolar junction transistors (BJT) driven by a common-emitter stage, still shows impressive isolation and phase noise performance. This design is presently used in about 100, five channel units for NIST's atomic clock ensemble. Building on this design, De Marchi, et al. [2] produced a very high performance amplifier with a 350 MHz bandwidth and isolation as high as 150 dB. This was achieved using the circuit shown in figure 2. The amplifier used three alternating npn and pnp transistors with their bases tied directly to ground. This required a bipolar power supply and a separate bias

0-7803-1945-1/94 $3.00 © 1994 IEEE

current for each transistor. The signal was coupled out via a common emitter stage. By using only SMT technology, parasitics were greatly reduced and the very high bandwidth was realized. The high isolation is also attributed to having the bases tied directly to ground, effectively forming a shield between the collector and emitter, thus reducing capacative coupling .

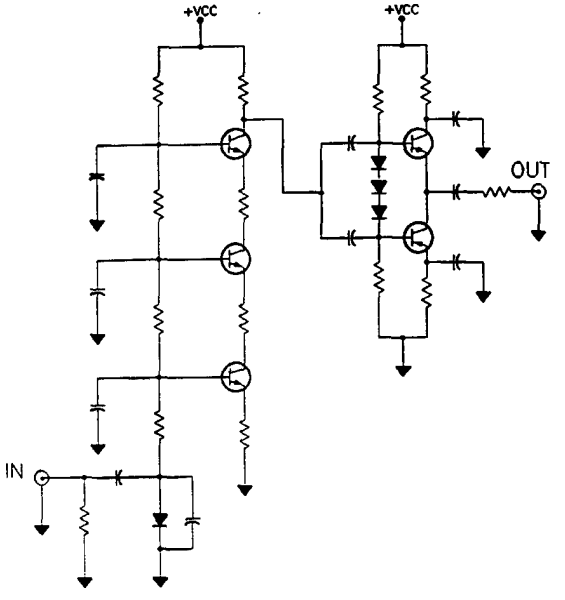


Figure 1. NIST's present isolation amplifier

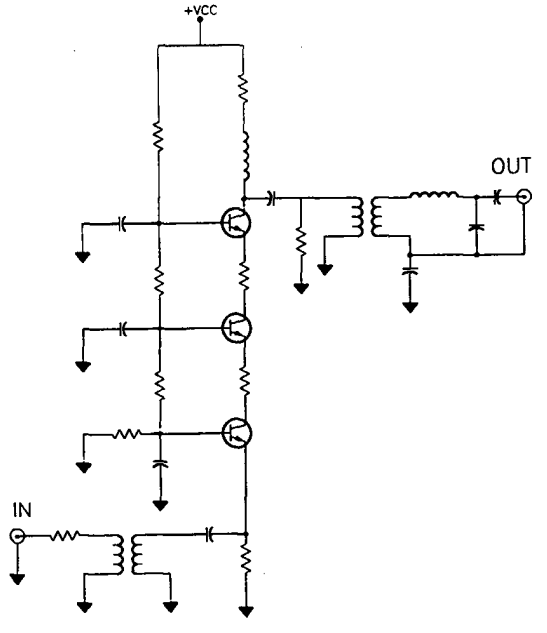


Figure 2. De Marchi's isolation amplifier.

Circuit Description

The power constraint placed on the amplifier by the existing rack system ruled out De Marchi's design, which consumed five watts per channel. The following circuit was created by combining aspects of both Gray and De Marchi designs.

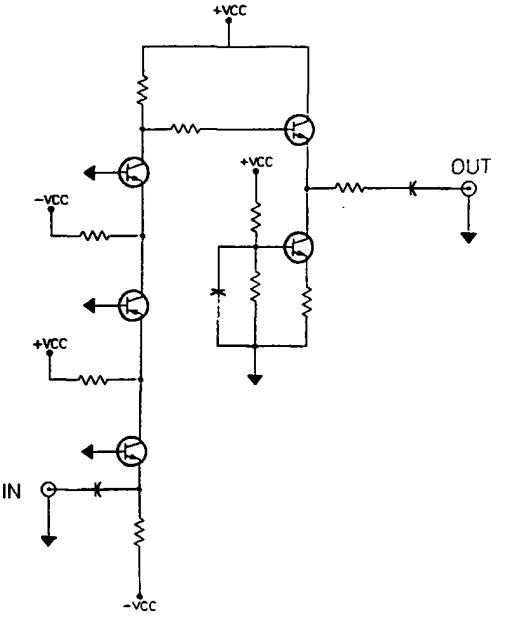


Figure 3. NIST's new isolation amplifier

De Marchi's approach of three common base stages was combined with Gray's single current and voltage bias network. This required tying the bases to ground via large capacitors to maintain isolation. This approach dropped the power requirement for a single channel from five to less than one watt. To improve the harmonic distortion and temperature coefficients, the output drivers of the previous designs were removed. In order to achieve current gain in this common-base configuration, it was required to couple in with transformers. This coupling also enabled very precise impedance matches. The output is also transformer coupled and DC isolated. Simple L-type networks were used for input and output matches. The configuration of the voltage bias network supplying the transistor bases forms a very effective active noise filter, which is responsible for the extremely low phase noise of the circuit. Nickel-plated steel shields were soldered around each channel to help improve the isolation. By moving back to the single bias current configuration, the amplifier's power consumption was greatly reduced, at a cost of reducing the operating bandwidth.

Performance

The amplifier was configured to accept 5 and 10 MHz at +13 dBm and to provide five +13 dBm outputs. The plot of gain vs. frequency (figure 4) shows a 1 dB bandwidth of 22 MHz.

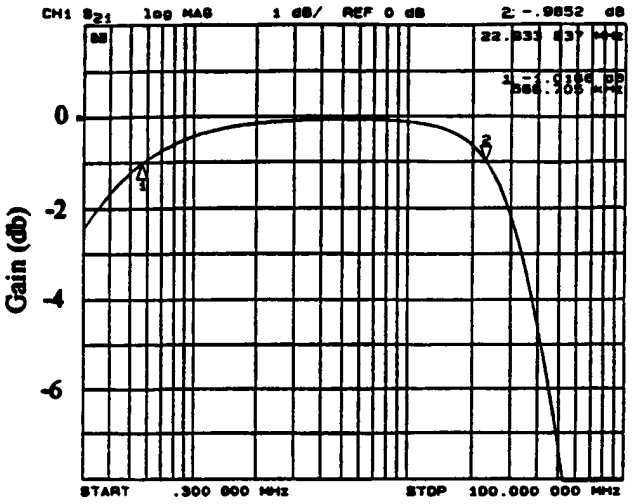


Figure 4. Gain vs. Frequency

Isolation was measured by injecting a +13 dBm signal at one port and measuring it at another. Reverse isolation was determined to be 144 dB, isolation between adjacent outputs is 125 dB, and isolation between non-adjacent outputs is 135 dB. We also attempted to measure isolation by opening and closing a port at a certain frequency while observing the other channels on a cross-correlation time measurement system. The effect could not be distinguished from the noise floor of the system.

Phase noise shown in figure 5, measured using a cross-correlation technique, shows a 1 Hz intercept of -152 dBc/Hz and a flicker floor of 170 dBc/Hz. The 1/f corner appears to be about 40 Hz. The droop at less than 2 Hz is due to the measurement being ac coupled. DC coupled spans of 20 Hz were not practical due to the high numbers of averages required for the cross-correlation and the possibility of overloading the analyzer.

By integration over the 1/f and f^0 noise components one can obtain an estimate of the short term fractional frequency stability of the amplifier (figure 6). This of course assumes a continued 1/f noise process inside of 1 Hz. This result could not be experimentally verified because NIST is presently unable to measure stabiltites of 10^{-16} at 5 Mhz and measurement times of 100 seconds.

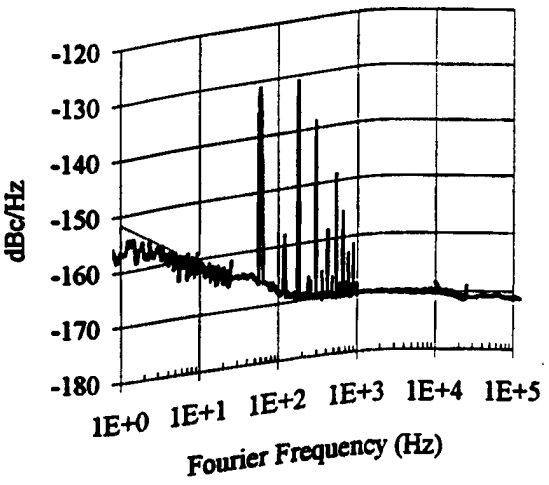


Figure 5. Phase noise of isolation amplifier

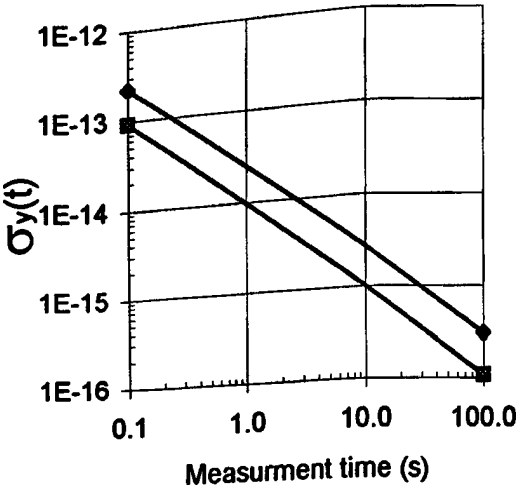


Figure 6. Estimated fractional frequency stability for a 1 Khz Bandwidth

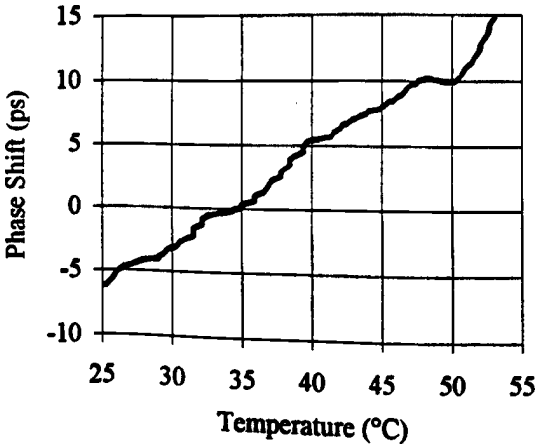


Figure 7. Phase-temperature coefficients

Phase stability vs. temperature was measured using the usual phase bridge technique[3]. In order to reduce noise the mixer output was integrated with a time constant of 10 seconds. Over a 25 °C range we observed a change in delay of only 750 fs/°C (figure 7). The range tested included 10 °C above and below the nominal temperature.

Special attention was paid to input and output impedance matches. Equation 1 shows the timing errors due to a transmission system [3].

$$\delta t = \frac{L}{\beta c} + \frac{1}{4\nu_o}\frac{\rho_S \rho_L}{\eta}\sin\phi \qquad (1)$$

where δt is the timing error, L,β,c have to do with the transmission media, ν_o is the frequency, ρ_S and ρ_L are the reflection coefficients of the source and load, η is the round trip attenuation and $\sin\phi$ is the angle of the twice reflected signal at the load. Table 3 shows the VSWR of the old isoamps and the worst case timing errors (η=1 and $\sin\phi$ =1) for two identical amplifiers connected in series. Table 4 shows the same for the new isolation amplifier.

Table 3 Timing errors of old amplifier due to VSWR

	input VSWR	output VSWR	error
5 MHz	1.52	1.19	900 ps
10 MHz	1.46	1.11	250 ps

Table 4 Timing errors of new amplifier due to VSWR

	input VSWR	output VSWR	error
5 MHz	1.002	1.04	2 ps
10 MHz	1.01	1.15	10 ps

The above tables show the extreme importance of impedance matching in precision timing applications. By carefully cutting all cable lengths to integer multiples of λ/2 one should be able to get the error due to VSWR down to 0.2 ps at 5 MHz.

With all transistor stages running in class A operation harmonic distortion was greatly reduced. A comparison between the old and new isoamps is shown in figures 8 and 9. These figures were both measured at +13 dBm output power. The new isoamp shows the second harmonic at -45 dBc and the third at -50 dBc. Running with an input of +15 dBm the amplifier runs with 0.25 dB compression and -35 dBc harmonic distortion.

Conclusion

With the performance improvements presented above, the new isolation amplifier can achieve the required 1 ps error budget. It is presently being installed at NIST to distribute signals from masers and other high stability sources.

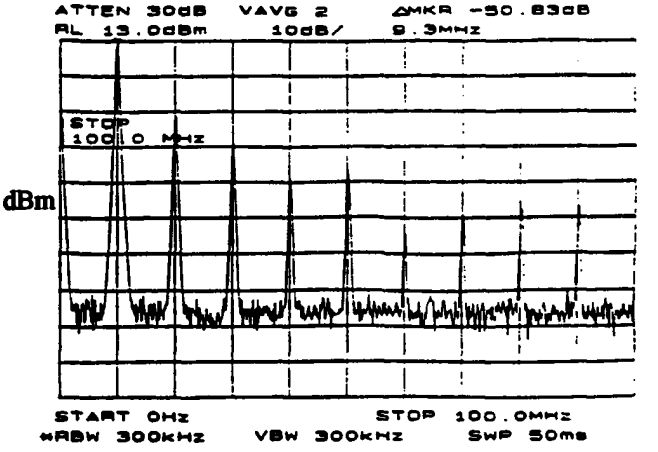


Figure 8. Harmonic distortion of old amplifier at an input level of +13 dBm

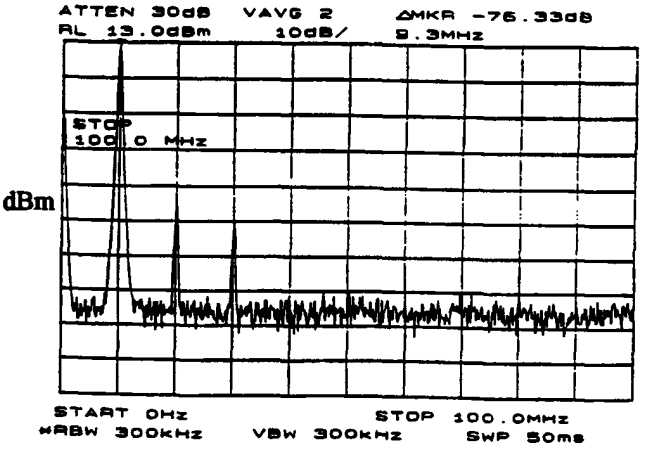


Figure 9. Harmonic distortion of new amplifier at an input level of +13 dBm

Acknowledgments

The authors would like to thank Steve Jefferts for design suggestions as well as Andrew Novick, Tony Dang and Cindy Osbourne for their help in prototype assembly.

References

[1] F. L. Walls, S. R. Stein, J. E. Gray, D. J. Glaze, "Design considerations in state-of-the-art signal processing and phase noise measurement systems", Proc. 30st Ann. Frequency Control Symp., 1976, pp. 269-274

[2] A. De Marchi, F. Mussion, M. Siccardi, "A High isolation low noise amplifier with near unity gain up to 100 MHz", Proc. 47st Ann. Frequency Control Symp., 1993, pp. 216-219

[3] L. M. Nelson, and F. L. Walls, "Environmental effects in mixers and frequency distribution systems", Proc. 46st Ann. Frequency Control Symp., 1992

A No-drift and less than 1 x 10^{-13} Long-term Stability Quartz Oscillator Using a GPS SA Filter

John A. Kusters & Kenneth Ho
Hewlett-Packard Co., Santa Clara, CA

Robin P. Giffard & Leonard S. Cutler
Hewlett-Packard Laboratories, Palo Alto, CA

David W. Allan
Allan's TIME, Fountain Green, UT

Introduction

Recently the time-domain spectrum of the Global Positioning Satellite (GPS) Selective Availability (SA) modulation has been characterized[1]. The continued growth of GPS, its utility, and the technology to receive and utilize the GPS signals have resulted in stable, multi-channel GPS timing receivers. These are now available from several vendors. Further, the GPS satellite constellation is essentially complete and has been declared operational, at least in its initial phase.

Coupled with the growth in hardware technology and understanding of GPS has been the development of algorithms which provide suitable enhancement to GPS timing signals, increasing their stability and accuracy. Using the above, we have constructed a near optimum SA filter which essentially removes the effects of SA on timing signals. The filter was designed with three aspects in mind:

1. Good performance and reliability obtained even in the presence of occasionally bad data and with restrictions on the GPS satellite viewing angles.

2. Improved performance for both time and frequency outputs by optimally filtering the SA Modulation to best match the stability performance of the internal oscillator.

3. Maximum performance results from minimum design and product cost.

The second and third aspects imply that oscillators other than cesium and/or rubidium should be considered.

The HP10811D/E SC-cut oscillator has been in continuous production for over 15 years, with a total production considerably over 100,000 to date. As we build only one crystal and one oscillator, continued value engineering[2] has resulted in exceptionally high yields and excellent performance, especially in aging and response to environmental conditions. As a result, the oscillator exhibits time-domain stability better than 1 x 10^{-11} well beyond 1000 seconds. This infers that the HP10811 is fully suitable for integration with GPS given that the decorrelation time of GPS has been determined to be about 400 seconds[1].

SA Filters

The SA Filter consists of algorithms implemented in a microprocessor to form a series of digital loops that have the overall effect of minimizing the effect of SA on the timing signals currently available from the GPS system.

To illustrate the effect of the SA Filter, Figure 1 shows typical data obtained by measuring the 1 pulse per second (pps) output of a six-channel GPS timing receiver against a similar pulse from an active ensemble of cesium standards. Actual data were taken each second and 100 consecutive measurements were averaged to produce the data shown. The pulse-to-pulse jitter of the ensemble is less than 100 ps and cannot be a source of the jitter seen in this figure. The slope of the data in Figure 1 represents a frequency difference of 9.4 x 10^{-14}

0-7803-1945-1/94 $3.00 © 1994 IEEE

between UTC(USNO) via GPS and the cesium ensemble.

Figure 2 shows two plots, the gray plot replicates the data shown in Figure 1 with mean slope removed. The rms deviation of this data is 31 ns, corresponding to the SA noise observed using 100 second averages of the receiver output. Six satellites were almost always available for tracking. The solid line shows the result of applying a filtering algorithm that is optimized for comparing GPS with a high quality cesium standard. The rms of the filter output is 1.7 ns, demonstrating the resolution that can be obtained with this technique for time transfer between cesium standards. The filter design is such that the output can be obtained in real time. This filter provides no improvement for dynamic positioning and assumes the coordinates are fixed and known. The coordinate solution obtained through the SA noise is adequate.

The effect of the filtering algorithm can be seen in Figure 3. The upper line shows the modified Allan Deviation (MDEV) of the time difference data before filtering. For times short compared with 10^5 seconds the noise is dominated by SA, and the slope is about -3/2, indicating a white phase-noise process. The lower line is the MDEV of the filtered data. The amplitude of the noise has been reduced approximately to the noise level expected from a cesium standard. At 2×10^5 seconds, outside the stop-band of the SA filter, the value of MDEV observed is of the same order as the noise typical of the steered GPS clock[3], and slightly larger than the expected noise of the ensemble, 1×10^{-14}.

The SA Filter has another attribute that makes the overall system response far more robust than a single receiver by itself. Because of the reduction in SA noise, coupled with the excellent long-term stability of the 10811 oscillator, the process is able to provide a calibration of the oscillator characteristics. Over a period of time, by tracking the oscillator's time and frequency offsets as well as its frequency drift, the effects of temperature can be discerned if they are repeatable as they are averaged over time. These model elements can be used if the system has to go into holdover mode. This procedure maintains the system with a given degree of synchronization and/or syntonization for a significantly longer period than could be predicted if the oscillator's performance were not so calibrated.

Overall, the combination of good GPS timing receivers, excellent oscillator performance, and the SA Filter algorithms permit performance while locked to GPS approaching that of low-end cesium standards and, while in holdover, to approach that of rubidium standards at a cost comparable to quartz standards.

Experimental Results

During the investigations of the SA Filter concepts, several equipment configurations were used. The difference between all of these was the manner in which the quartz oscillator was controlled. In the experiments, the oscillator was either steered and became the primary source of output, or was unsteered and was used as an internal time-base for a synthesizer.

Unsteered oscillators are more stable. Steered oscillators permit operation with simpler electronics.

Holdover Measurements

A goal of the experiment was to determine how much improvement in the aging of the oscillator was obtained by characterizing its model elements. To standardize on the method, and to continue collecting data during 'holdover,' we chose to post-process the data and predict holdover performance and compare to actual performance. As a follow-on to the data shown here, actual holdover performance is now being measured and validates our analytic method.

Unsteered Oscillators

Figure 4 is the generic block diagram of an unsteered oscillator system. The Frequency Translator may be implemented in many ways, most commonly some form of synthesizer using the quartz oscillator as its timebase. An essentially continuous measurement is made between the external reference and an appropriate signal from the oscillator. As the software knows what difference was measured, an appropriately compensated command can be sent to the Frequency Translator. Over a period of time, in comparison with the external reference, the SA filter provides the opportunity to characterize the oscillator's instability characteristics.

Figure 5 shows the results of the oscillator offset measured as described above. The oscillator had been in storage for several months before the start, and was used as we received it, with no pre-aging or syntonization. The slope seen in Figure 5 is mostly due to a frequency error of about 1.1 Hz. Frequency as computed from the phase data is shown in Figure 6. Immediately apparent is a large aging rate for the first few days. The gray plot in Figure 6 is the computed frequency showing the effects of SA noise. The solid line was obtained by processing the original data through an SA Filter

When the filtered data is examined to determine aging rates, the curve shown in Figure 7 results. A 12 hour moving window was used on the filtered data to produce this curve. The 12 hour window has the effect of enhancing any diurnal effects. A rather strong temperature dependent effect is seen.

To examine holdover response, filtered data from days 4 and 5 were used to characterize the oscillator aging. The filtered phase data was fit to a nearly parsimonious model (a linear term plus a log term). The data string used consisted of 10 points essentially uniformly distributed over the 24 hours. The holdover results are shown in Figure 8. In spite of strong temperature effects, the overall apparent aging was reduced by a factor of 20 or more when the data was taken during day 4. When the data was taken during day 5, the improvement was greater than 30. As the oscillator aged into specification, the performance continued to improve.

Steered Oscillators

Figure 9 shows the generic block diagram of a system employing a steered oscillator. The steering method and hardware were adapted directly from the HP5071A Cesium Standard. The actual steering command is a digital word passed through D/A converters and appropriate filters before steering the oscillator. The resolution of the numerical equivalent of the steering command is about 1×10^{-11} per steering unit. A plot of the steering commands is shown in Figure 10. A new steering command is computed by the system every 10 seconds. This plot also shows a non-optimal, non-parsimonious, prediction of future performance. This was obtained by doing a 2nd order regression to the data in the 24 hour period shown on the Figure as the "Characterization Period."

Again, holdover performance was determined from the difference between actual and predicted. The expected synchronization and syntonization errors are shown in the table in Figure 10. Of note is that this oscillator performed better than the oscillator used in the unsteered experiment above, and the temperature effects were dynamically compensated through a temperature sensor, with time lag and temperature coefficient factors learned in a prior experiment.

Unsteered oscillators are more stable and predictable than continuously steered oscillators. During the several experiments, we noted that while the very-long-term aging of steered and unsteered oscillators are essentially identical, the day-to-day results are considerably different. Usually unsteered oscillators show continuous monotonic aging without any unexpected results. Steered oscillators were anything but monotonic, tending to change sign of the aging every few days.

Although unsteered oscillators are more predictable, and hence have improved holdover, predictions made on the basis of steering commands produce acceptable performance, showing deviations in frequency less than 1×10^{-10} and time deviations on the order of 1-2 microseconds for periods up to 3 days after the oscillator has been characterized.

Figure 11 is a plot of the MDEV of a GPS Steered oscillator determined under two conditions. The first used the 1 pps from the system measured against the cesium ensemble with a time-interval counter. The system used is the HP5071A cesium long-term stability production measurement system. The second measured the 10 MHz output against an offset 10811 oscillator in a conventional heterodyne short-term stability system. The two measurement systems were set up to take data concurrently. A key point on the 1 pps measurement is the measurement noise seen from 1 second through 100 seconds. This is a consequence of the 1 pps circuitry that has a pulse-to-pulse jitter of about 30-40 ps, and a time-interval counter with a minimum resolution of the same order of magnitude. Hence the actual stability is represented by the bottom curve through sampling times of 10 seconds, and the upper curve for times beyond 200 seconds.

The overall curve illustrates several points. Even though steering commands are issued every 10 seconds, the digital noise is well below the output time-domain stability. The long-term stability at time samples greater than 10,000 seconds approaches that of cesium. The intermediate stability, although always below 1×10^{-11}, shows the effects of oscillator characteristics, and of the various control loops and their time constants.

Given the usually unknown condition of the oscillator, GPS receiver, and the various time delays in the SA Filter and control loops, the start-up time to full lock and tracking to GPS with GPS steered oscillators takes about 8 hours. This is shown in Figure 12. After 8 hours, the system is synchronized to within 100 ns, and the average frequency offset magnitude over the next 16 hours was less than 5×10^{-13}.

Summary

The data presented above indicate that it is possible to obtain near cesium performance in the long-term from an appropriate quartz oscillator when combined with current GPS timing receivers and a near optimal SA Filter algorithm. Further, the data shown verify that both steered and unsteered quartz oscillators give adequate performance.

References

[1] Allan, D.W. & Dewey, W.P., *Time-Domain Spectrum of GPS SA,* presented at the 1993 ION GPS-93 Conference.

[2] Kusters, J.A. & Adams, C.A., *Applications of Total Process Control Techniques in the Production of High Precision Quartz Resonators*, Proc. 39th Annual Symposium on Frequency Control, pg. 475 ff, 1985.

[3] Buisson J.A., private communication

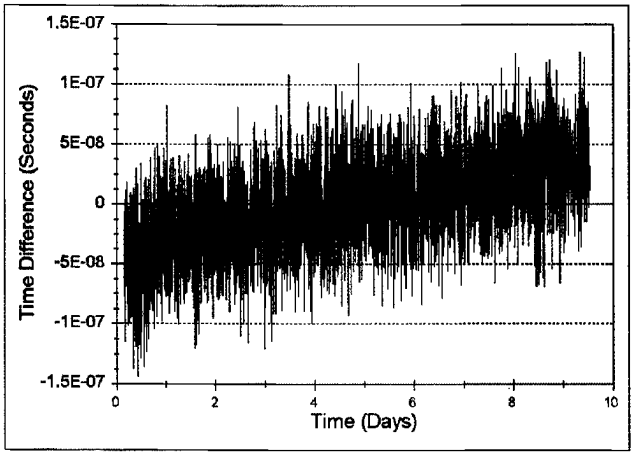


Figure 1 -- GPS Receiver vs. 5071A Cesium Ensemble 100 second averages of 1 pps data

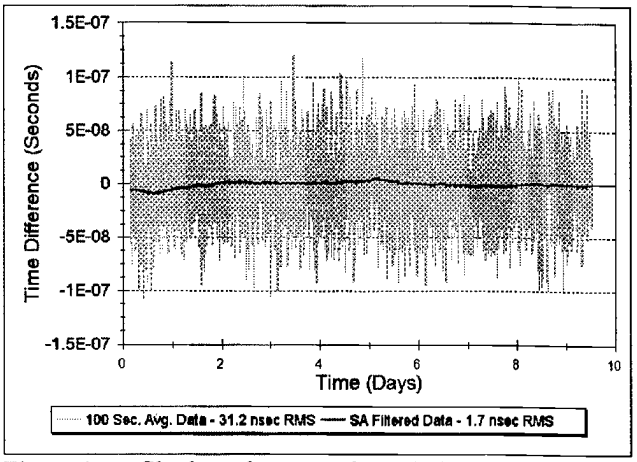


Figure 2 -- Shadow plot - raw data of Figure 1 with 9.4 x 10^{-14} offset removed. Solid line, results of applying SA filter to raw data

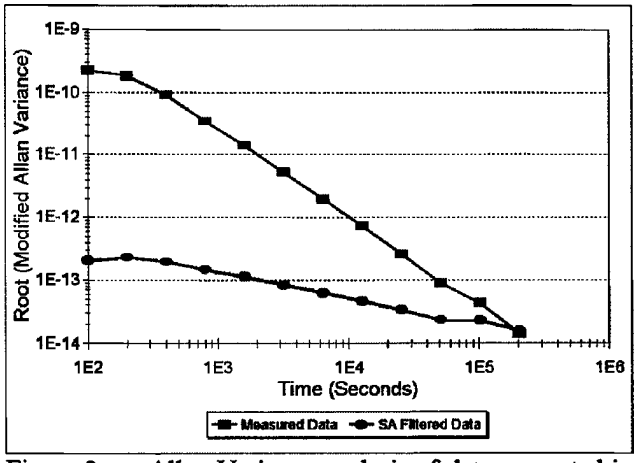


Figure 3 -- Allan Variance analysis of data presented in Figure 2.

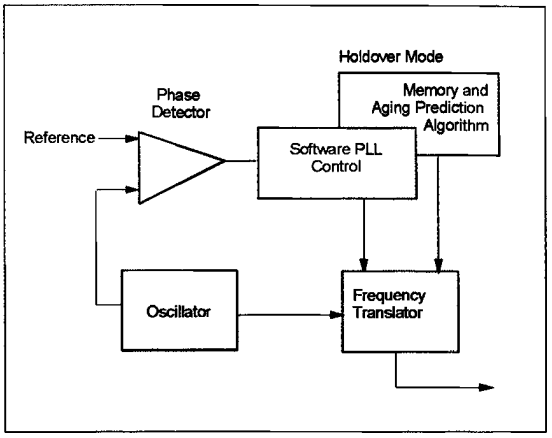


Figure 4 -- Block diagram, unsteered oscillator system

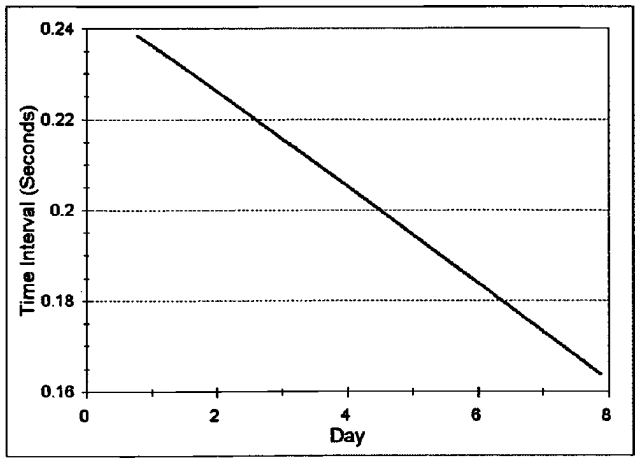


Figure 5 -- Phase measurement of unsteered oscillator vs. GPS receiver - 100 sec averages of 1 pps data.

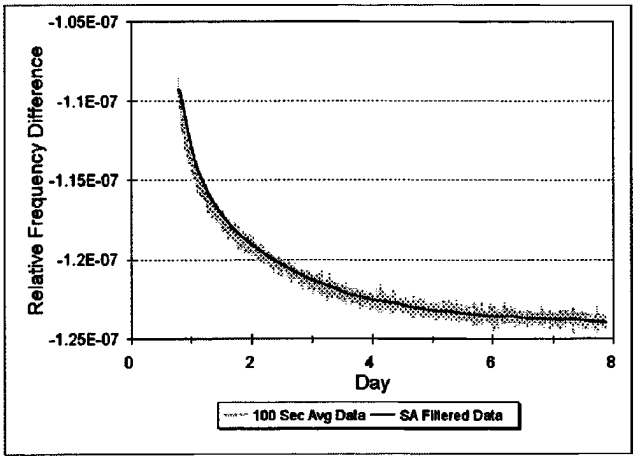


Figure 6 -- Frequency data - Figure 5. Shadowed data shows effects of SA. Solid line - results of applying SA filter to original data

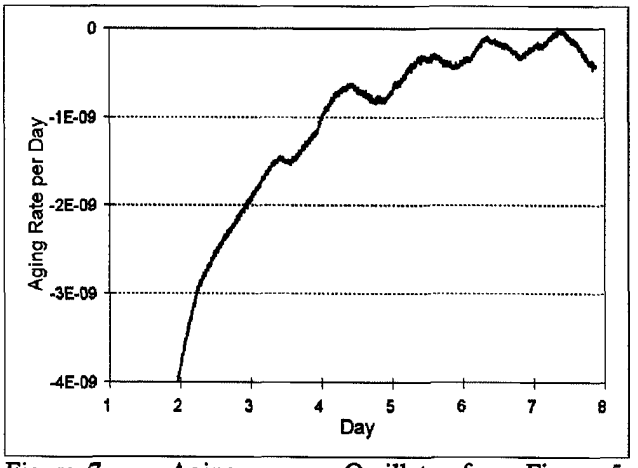


Figure 7 -- Aging curve - Oscillator from Figure 5. Curve derived from SA Filtered data with a 12 hour moving window to emphasize diurnal effects.

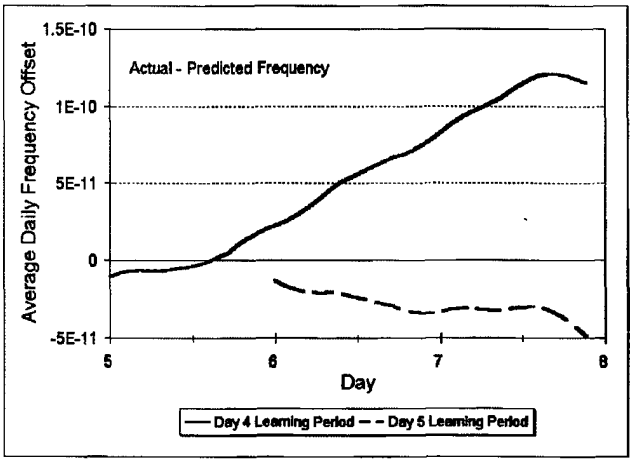


Figure 8 -- Holdover performance - Curve computed from SA Filtered data fit. Characterization period was 24 hours to reduce diurnal effects.

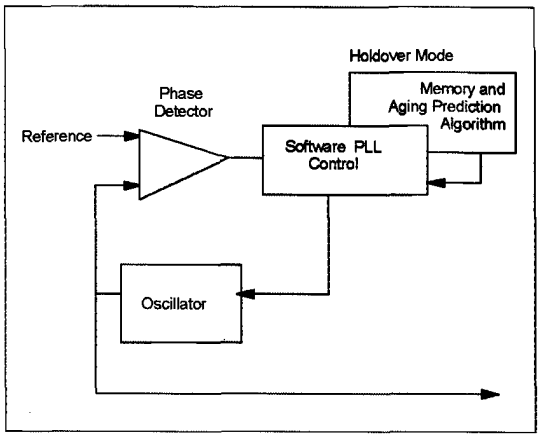


Figure 9 -- Block diagram, steered oscillator system

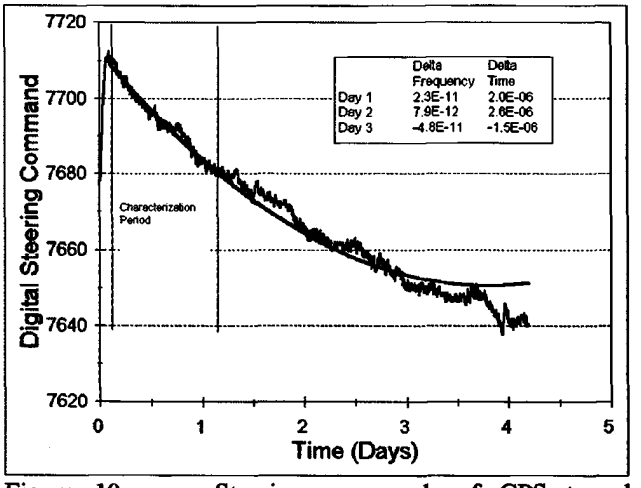


Figure 10 -- Steering commands of GPS-steered oscillator. Table gives frequency and time errors accumulated after end of characterization period.

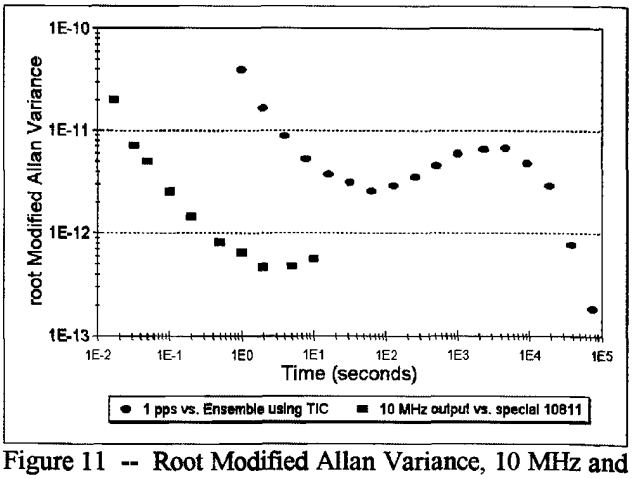


Figure 11 -- Root Modified Allan Variance, 10 MHz and 1 pps output of GPS Steered oscillator

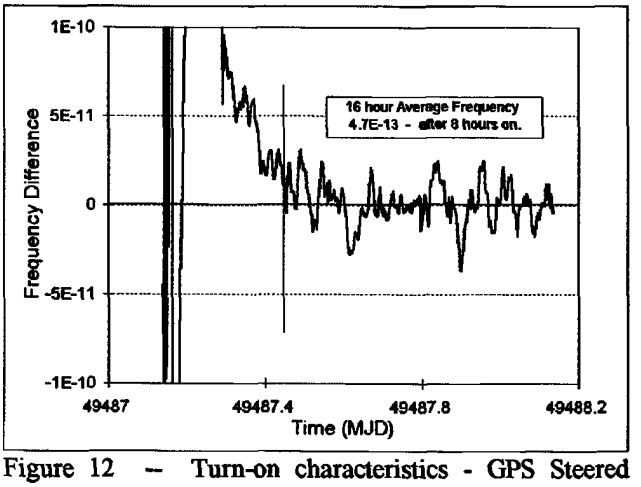


Figure 12 -- Turn-on characteristics - GPS Steered Oscillator. Frequency offset computed after 8 hours on.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DUAL-MODE CRYSTAL OSCILLATORS WITH RESONATORS EXCITED ON B AND C MODES

Anatoly V. Kosykh, Igor V. Abramson, Victor P. Bagaev

Omsk Riezoengineering Company Ltd.
P.o. Box 7357, 644020, Omsk, Russia

There is very attractive idea to use SC-cut resonators excited on B and C modes for frequency stabilization (in particular, for digital thermocompensation) but up to now these operating regimes are used very seldom. The cause consists in a difficulty to get stable excitation in wide temperature range. Due to development of original resonator exitation networks, of special geometry of piezoelement and due to original technology of its manufacture, we succeeded in developing of B and C modes in wide (-60...+85)øC temperature range. 70% yield of suitable units is attainable when resonators are exited on 10 MHz frequency using 1-st mechanical harmonic. The requirements to a circuit of a dual-mode oscillator (that is oscillator capable of simultaneous exiting on two asynchronus frequencis) are presented.

While this work a lot of experiments have been involved to design double-mode crystal meeting the requirements assined above. To avoid the dips of activity of the B-mode being inherent property of overton SC-cut crystals due to nonlinear interaction of the thickness modes the fundamental mode TD-cut (yxbl/23ø25'/34ø) crystal plate have been used here. Studies of both C and B-mode activity on the plano-plano plate geometry, the electrodes form and the mounting points location provided creation of miniature (7x8 mm) crystal plate having acceptable motional resistance on B and C-modes. Exploration of the modes activity behavior over a wide temperature range shown that the motional resistance remains within (8-12) Ohms for C-mode and (50-65) Ohms for B-mode. Practical realization of the dual-modes oscillator in the form of a hybrid LCI is described. The dual-mode crystal oscillator parameters are: output frequency - 10 MHz; supply voltage - 5V; consumption current - 5-7 mA; output levels - TTL; B-mode level in C-mode output spectrum -50 dB; regime unstability - (0.1...0.2)ppm/10% voltage change.

INTRODUCTION

Sixteen years have already passed since J. Kusters and colleagues proposed to use SC-cut resonators excited at B and C modes for the purposes of thermo-compensation [1]. Attractive as it is the idea has not yet found wide application.

With an acceptable F-T curve and an extremely low dynamic temperature coefficient in C-mode, a steep monotonous F-T curve in B-mode, a spatial coincidence of the both modes in SC-cut resonator (C-mode as a reference and B-mode as a thermo-sensor) promises to results in a very low dynamic thermocompensation error. All the above SC-cut peculiarities make possible to obtain high accuracy oscillators. The laboratory experiments (e.g.[2]) corroborate these theoretic presumptions.

But so far in Russia at least the progress has not yet passed over the experimental stage. One of the main reason is the complexity of obtaining stable dual-mode excitation and B-mode oscillation quenching observed in a narrow temperature interval. Investigations have shown that in certain temperature points a B-mode dynamic resistance rises abruptly as a result of B and C modes interaction and energy transfer. The fact that activity dips are not observed in every unit of 100% of resonators elevates some hope that there is the solution of the problem. But for many years this solution was not found and the researches proceeded to seek the alternative version of solution free from this limitation.

In 1989 a number of papers was published [3,4,5,6] devoted to problem of SC-cut resonators use in digital thermocompensated oscillators (DTCXO) with excitation on the 1-st and 3-rd C-mode harmonics. A frequency difference between the 3-rd (fc) and the tripled 1-st (fc) C-mode mechanical harmonics is used as a thermo-dependent oscillation. The idea itself is not new one. F-T curve linearity of the difference oscillation (fd = 3fc -fc) is well known from classic crystal oscillator literature (e.g. [7]). L. Marianovsky (Russia) was the first (to our opinion) who suggested to use the "difference oscillation" as a thermo-sensor signal in DTCXO [8]. He used AT-cut in his experiments and so could not have practical results as good as the authors of [3,4,5,6].

After analytical comparison we came to the conclusion that the simultaneous B- and C-mode excitation with extraction of their frequency difference was more promising and had undertaken our efforts on the oscillator design with this type of resonator. Two investigation directions were used:

1- resonator design with constant parameters in a temperature interval;
2- oscillator circuit development providing effective simultaneous excitation of two modes.

The investigations resulted in a high parameters dual-mode oscillator with rather good reproducibility. There are tuned circuits in oscillator design developed and they are suitable for microcircuit fabrication.

0-7803-1945-1/94 $3.00 © 1994 IEEE

1. MAIN SPECIFICATION OF A DUAL-FREQUENCY OSCILLATOR USED IN DTCXO

A dual-frequency oscillator intended to be used in a DTCXO should fully comply with the following specifications:

1. The span of the reference mode F-T curve in an operational temperature interval should be as low as possible.
2. The F-T curve of the thermosensor mode should be monotonous with the slope as large as possible.
3. The oscillation with reference mode frequency should be highly stable with low noise and discrete components levels.
4. Thermal time constants of the both modes should have similar values.
5. The reference mode should have low dynamic temperature frequency coefficient.

Keeping in mind the above specifications let us now compare resonators with B and C modes excitation and resonators with the 1-st and 3-rd harmonics of C-mode. Since the C-mode is used as a reference oscillation its parameters will be in both cases approximately the same. Principal distinctions lay in the parameters of the thermosensor oscillation. A circuit diagram for this oscillation forming in both cases is shown in Fig.1a and Fig. 1b. Parameter comparison for both versions is given in Table 1.

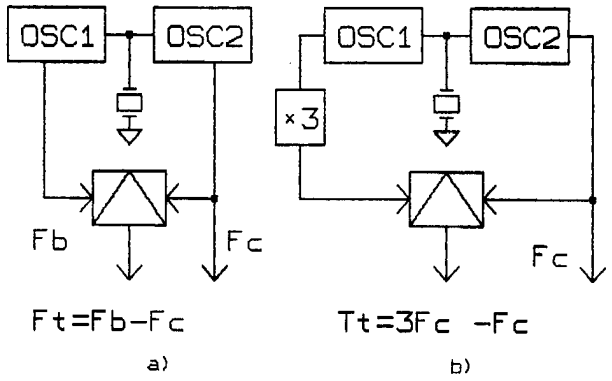


Fig.1. **Dual-mode oscillators block-diagrams.**

Table 1

variant	reference mode requency	F(T) MHz	F-T curve absolute slope	F-T curve relative slope, ppm
B&C mode	10 MHz	0.9	-300 Hz/øC	-330
1-st & 3-rd harm.	10 MHz	0.15	-14 Hz/øC	-95

Two variants of thermosensor signal treatment in DTCXO are possible:
- a frequency-meter configuration (See Fig. 2a);
- a period-meter configuration (See Fig. 2b).

Evidently in both cases the higher is the thermosensor characteristic slope S(T), the better are DTCXO's parameters. In particular, the relations are revealed of S(T) with dynamic

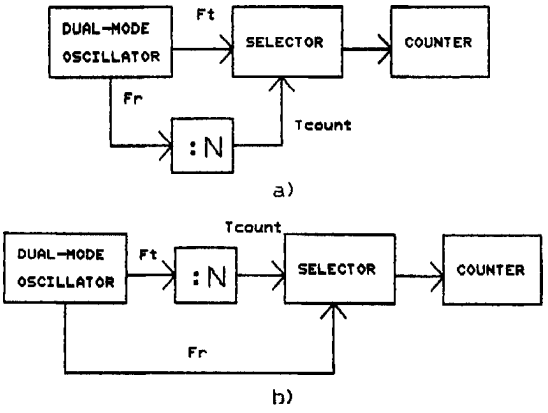


Fig.2. **Thermosensor signal treatment devices block-diagrams.**

temperature characteristics [9] and with a digital noise spectrum of an output oscillation [10]. Of primary importance is the ratio of the absolute slope S(T) to the absolute instability (power supply, long-term, short-term) of the oscillation F(T). It is easy to realize that the higher is the absolute slop S(T) the easier is (with other condition being equal) to secure their necessary ratio. Since the slope S(T) is 20 times higher in the 1-st variant (in absolute figures) in solving the problem of the B-mode stable excitation this variant is obviously preferable. The problem solutions will be considered in the following items.

2. METHODS OF DUAL-FREQUENCY OSCILLATIONS EXCITATIONS

The behavior of one and the same resonator is different in different oscillation circuits. It is especially evident in such complex cases as simultaneous excitation of two modes. Besides stable two-mode excitation a dual-frequency oscillator should have some other qualities:

- low consumption power;
- small overall dimensions;
- ease of tuning in manufacture;
- pure spectrum of the output reference oscillation.

From these initial positions the authors treat double frequency excitation methods familiar to them.

2.1. Dual-frequency oscillators with separated nonlinear elements

A dual-frequency oscillator of this type is a device of two independent oscillators coupled to each other via dual-mode crystal resonator. Each oscillator comprises a selective filter providing the oscillations to arise only at certain frequency. A block diagram of this oscillator is shown in Fig. 3. Good mutual filtration of excited modes is an obligatory condition of oscillations stability. With equal oscillations activity the selectivity of separation filters should not be worse than 20 - 30 dB. If one of the modes is more weak than the other one, the stronger mode level at the input of the oscillator exiting the weaker mode should be suppressed still more. In practice, LC-cirquits are used as selective elements, their quantity

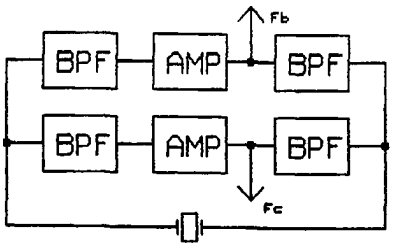


Fig.3. Dual-frequency oscillators with separated nonlinear element block-diagram.

being up to 3 or 4 [6]. The presence of adjustable tuned circuits makes worse the features of reproducibility, vibration stability and overall dimensions. There are opinions that to retain these qualities the separation filters could be made of piezomaterials, but this will make the oscillator more complex and costly.

2.2 Dual-frequency oscillators based on mutually synchronized oscillators

The seeking to get rid of tuned circuits and to obtain a diagram suitable for microelectronic implementation led to an idea to use mutually synchronised oscillators for the sake of filtration. It is known that in two mutually coupled oscillators synchronous oscillations are possible to arise under certain conditions. Synchronizm region is determined by the oscillator phase characteristic and by the coupling level. This coupled system exhibits striking properties and can be used for separation of dual-frequency oscillations. I. V. Boudiakov (Russia) suggested a diagram providing the excitation of the SC-cut resonator in B and C-modes. The diagram (See Fig. 4) comprises four oscillators mutually synchronized in pairs [11].

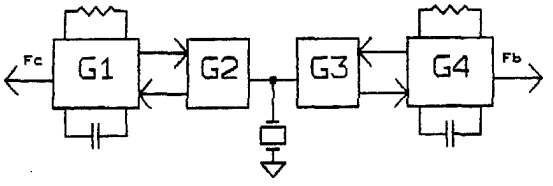


Fig.4. Dual-frequency oscillators based on mutually synchronized oscillators block-diagram.

The system of mutually coupled oscillators G1, G2 delivers the frequency of B-mode at the output and the system of oscillators G3, G4 - delivers the frequency of C-mode at the output. The G1 and G4 oscillators are of RC-type. Their natural oscillation frequencies are the same as the B-mode and C-mode frequencies respectively. An example of this diagram successful realization was given in [12] (See Fig. 5)

If the modes being excited are of near frequencies, a situation is possible due to temperature frequency drift of RC-oscillators when all four oscillators are acting at the same frequency. In this case a synchronization with LC -oscillator may be necessary instead of RC-oscillator synchronization.

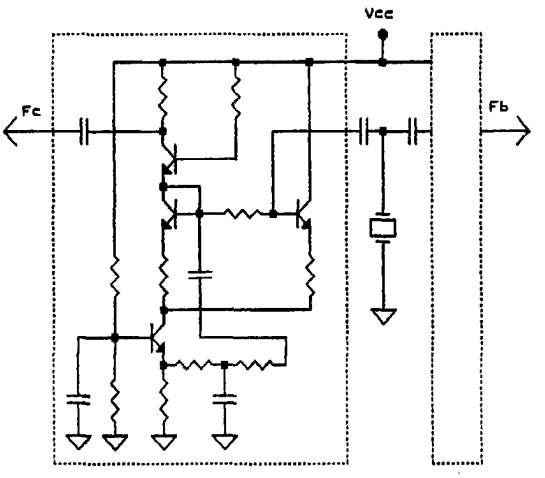


Fig.5. Mutually synchronized oscillators.

2.3. Dual-mode modulation type oscillators

When using B-mode not for frequency driving but for filtering, the change of its motional resistance shows itself in lesser degree. Pure filtering method does not give good thermosensor signal-to-noise ratio and, therefore, is not being used. Combined modulation-adaptation method developed by U. S. Shmaliy [13] and U. S. Ivantchenko [14] seems to be more interesting. Here C-mode is excited in simple self-oscillation regime. To obtain B-mode frequency, oscillation Fc is amplitude (or frequency) modulated by modulator MOD with oscillation Fb-Fc frequency coming from a certain VCO. In result, Fb frequency component appears in the output spectrum of the modulator. Resonator's B-mode is used as a frequency discriminator controlling VCO frequency. The feedback circuit being closed, VCO frequency always follows B-mode natural frequency, changing with temperature (See Fig. 6)

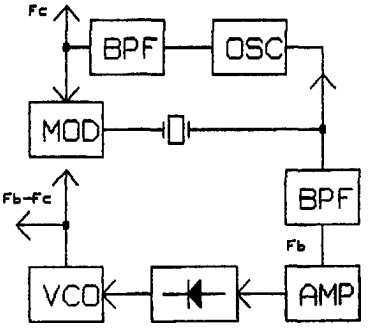


Fig.6. Dual-frequency modulation type oscillator block-diagram.

Adjusting modulation index we can set minimum possible Fb -signal level in the output spectrum, where still surely temperature signals Fb and Fc are formed. This method essentially diminish requirements to the constancy of B-mode motional resistance, but does not simplify the problem of mutual filtration of oscillations (in particular, of strong Fc -oscillation in the loop for modulating signal formation) It needs to complicate the circuit for their separation.

2.4. Automatically balanced modes, asynchronoysly excited dual-frequency oscillators

The dual-frequency oscillation methods considered above, regime rather complicated excitation and discrimination circuitry. At the same time there is a possibility to excite and maintain two asynchronous oscillations in single-transistor oscillator. V. Anisimov in his theoretical paper [15] proved the existence of this possibility provided that an amplifier element non-linearity is expressed by 5-th degree polynom:

$$1 = \alpha U - \gamma U^3 + \beta U^5$$

and there is a non-zero solution to the system of equations describing oscillation process in the oscillator. If there is a self-bias circuit and a cut-off angle is less than 90ø, such type of system can produce mild asynchronous self-oscillation.

The method of simultaneous excitation of two asynchronous oscillation in one self-oscillator is, from our point of view, the most economic and attractive and, therefore, all our efforts were aimed at its development. Non concentrating attention here on theoretical grounds, we shall consider practical aspects in detail.

Several self-oscillator cirquits can be chosen for a base network. V. Samoilenko (who was the first to begin works on dual-mode oscillators in our laboratory) used Colpitts [16] oscillating circuit. It was noticed that the process of dual-frequency excitation is influenced by transistor d.c. component, excited mode motional resistance and by parameters of self-bias circuits. It is convenient to carry out investigations of the oscillator using the network shown in Fig. 7.

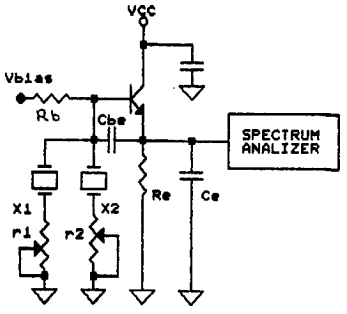
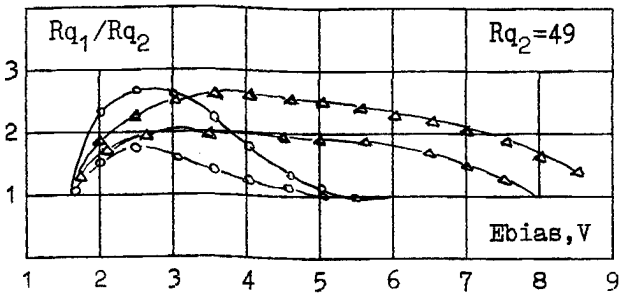


Fig.7. Dual-frequency oscillator network.

Since it is rather complicated to change mode motional resistance in actual dual-mode resonator, we recommend, while setting regions of stable dual-frequency excitation, to use separate monofrequent resonators with series-connected trimming resistors. Fig. 8 shows admissible ratios of excited mode motional resistances to base bias voltage at various values of self-bias resistance Rbias [17].

At high-resistance of the base circuit, the two-frequency regime region is more wide, but such type of oscillator is sensitive to replacement of transistors and to temperature changes. Good results can be obtained with low-resistance divider if to decrease, artificially, the amplitude of oscillation by way of a.c. loading



1 - Rb=680K; Re=1.8K 3 - Rb=43k; Re=1.8K
2 - Rb=680K; Re=720 4 - Rb=43K; Re=720

Fig.8. Allowed Rq1/Rq2 ratio vs. E bias.

the oscillator [17]. These oscillator circuits are shown in Fig. 9.

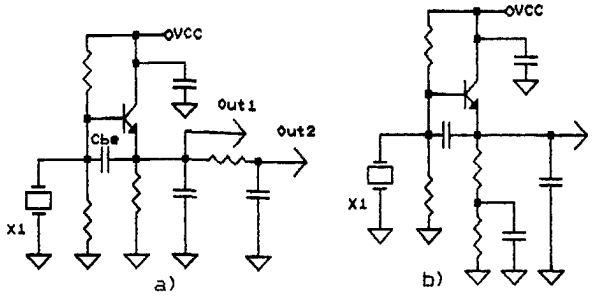


Fig.9. Improved dual - frequency oscillators network.

Relation of d.c. R(=) and a.c R(~) load resistances of the oscillator effects the oscillating region stability. As Fig. shows, this relation must be not less than 2.

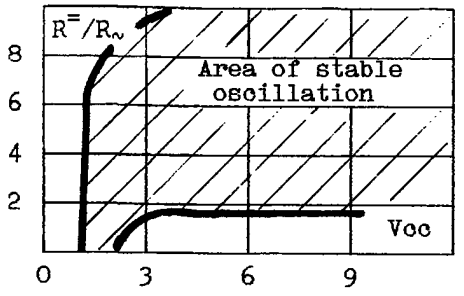


Fig.10. Allowed R=/R ratio vs. V

Fig. 8 and Fig. 10 give no idea about spectrum composition of output oscillation. Alternation of transistor regime, for instance, Eb causes the change of relation of Fc and Fb levels of excitation. For digital thermocompensation the most preferable regime is one at which B-mode level in output signal would be minimum. That kind of regime corresponds to the point where two-frequency oscillation transforms into one-frequency oscillation. To stay at this point, however, is not an easy task. When changing the supply voltage, or the temperature, the two-frequency oscillation can spontaneously pass to one-frequency oscillation. To hold the oscillator

in optimum regime it was proposed [16] to introduce a feedback circuit maintaining given excited modes level relation (Fig. 11).

Since the oscillator is a non-linear device, its output spectrum contains, apart frequencies Fb and Fc, their combinations. Low-frequency component Fd=Fb-Fc is the most interesting (because it is easily taken off Out 2 -point of the oscillator shown in Fig. 9). Difference frequency signal Fd can be filtered out with RC-filter, amplified, and, after detection, used to control two-frequency self-oscillator regime.

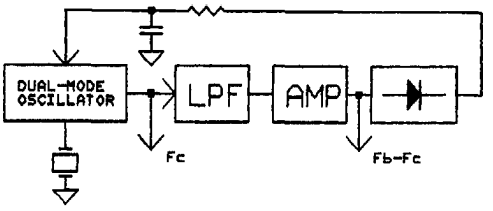


Fig. 11. Dual-frequency oscillator with authomatically balanced modes block diagram

An amplifies the difference signal to a level when its amplitude is enough to be used as thermo-sensitive output signal and for controlling oscillator d.c. regime. This type of oscillator was used in DTCXO [2].

Asynchronous excitation of two frequencies in single oscillator makes very strict demands to a resonator. Therefore, great attention was paid to its development.

3. DUAL-MODE RESONATOR

Evident advantages of double-mode oscillator using B-mode as a thermosensor have stimulated designing special dual-mode quartz resonator. Proceeding from peculiarities of the DTCXOs considered above we can define desired properties of the dual-mode resonators

- as small as possible C-mode frequency instability over the operating temperature range;
- absence of activity dips on both the C and the B-mode within the operating temperature range;
- appropriate ratio of the both modes' activity
- the C-resonance should be stronger than the B-resonance;
- no significant variations of the modes' activity ratio vs. an ambient temperature;
- the resonator should be controllable enough for adjustment within its frequency vs. temperature instability;
- minimum frequency hysteresis and aging rate with as small as possible dimension.

Taking into account all the above requirements a double-mode resonator design based on a SC-cut fundamental mode plano-plano piezoplate has been developed. It was shown in [18] that the fundamental mode SC-cut resonators have appropriate electrical parameters and no dips of activity when an optimal design of the piezoplate was used. Further investigations were aimed at minimizing the plate size to make it compatible with miniature crystal enclosers. The piezoplate measuring 7x8 mm and was mounted at two points located near the plate corners (Fig.12). The film electrodes diameter and the plate thickness provide sufficient "energy trapping" to achieve high Q-factor and to depress the coupling with spurious modes.

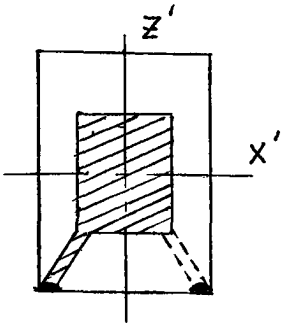


Fig.12. The dual-mode resonator plate configuration.

The length of the plate is oriented along Z' crystallographic axis to provide more sound attenuation of B-mode vibrations at the mounting points [19]. Electrical characteristics of the developed resonator are given in the Table 2.

Table 2. Electrical Characteristics of the Dual-mode Resonator

Mode	Frequency MHz	Motional resistance Ohms	Q-factor $\frac{3}{10}$	$\frac{C\ parallel}{C\ motional}$
C	10,0	10-15	~270	~1000
B	10,9	45-60		

As one can see from the Table 2 the resonator has the parameters to be adequate to dual-mode excitation requirements. The C and B modes resistance vs. temperature dependence is found to doesn't suffer sufficient distortions, however the B-mode resistance maximum value may exceed 2 times of the minimum value. So considerable variations in the modes activity are, of course, undesirable from the dual-mode excitation standpoint.

To reduce the B-mode resistance fluctuations a deposition of damping mass on the free edges of the plate has been used. The resulting resistance vs. a temperature is displayed in Fig. 13 in

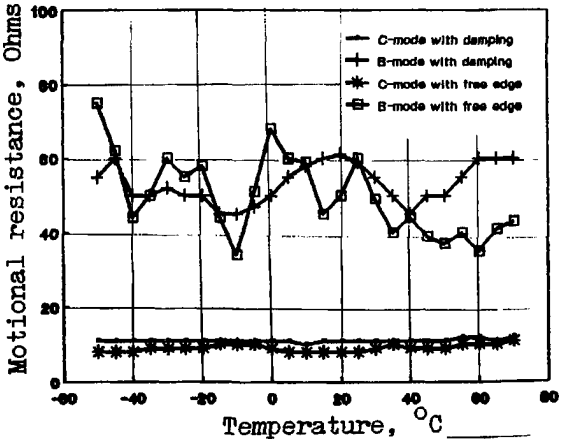


Fig.13. Motional resistance of the Band C-modes vs. temperature.

reference to the characteristic of the resonator without damping. From comparison of the data one can conclude that while the C-mode resistance behavior exhibits no noticeable difference, the B-mode resistance variations have been reduced from about 2 times to 1.4 times, under which the B and C- resistance average ratio remains without changes.

Frequency vs. temperature characteristics of the

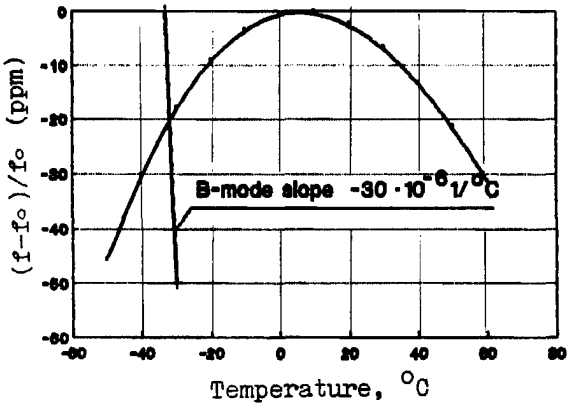


Fig.14. Frequency vs. temperature characteristics of the dual-mode resonator.

resonator in the C and B-modes are usual for SC-cut crystals (Fig. 14). To provide minimum frequency vs. temperature instability the lower turn over temperature should be arranged in the middle of an operating temperature range.

Aging rate of the dual-mode resonators was tested in steady operation state during 7 months at ambient temperature +60 C. Typical results are exhibited in Fig. 15. As one can see the final frequency shift doesn't exceed 4x10-7, during which for the last month it is less than 2x10-8. This allows to expect the yearly aging rate less than 5x10-7.

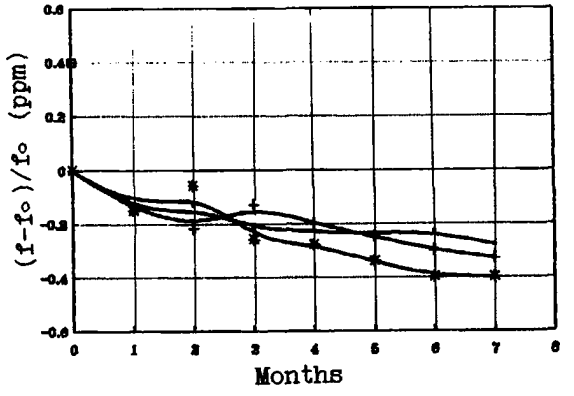


Fig.15. Ageing rate of the dual-mode resonators.

Frequency hysteresis of both main and thermosensitive modes is essential part of DTCXOs' instability when C-mode first harmonic and C-mode third overtone dual-mode excitation is used. In case of dual B and C-mode excitation the B mode frequency hysteresis doesn't matter due to large frequency vs. temperature slope of that mode.

Frequency hysteresis of the miniature dual-mode resonator for the C-mode was examined under ambient temperature variation on the cycle +20° → - 60° → +20° → +80° → +20° by steps of 20°C measuring meantime the resonator frequency. The frequency difference at some temperature points when the temperature first decreases (or rises) and then returns to the initial value is the frequency hysteresis.

For the resonators developed it was found to be less than 5x10-8 for the most units.

Taking into account all considered here one can conclude that the developed dual-mode resonators meet the most requirements and are promising for DTCXOs of (3-5)x10-7 frequency vs. temperature stability and less than 5x10-7 year's aging.

4. IMPLEMENTATION AND EXPERIMENTAL DATA

When developing a family of DTCXO [10], a task was set to create two-frequency oscillator circuit capable to excite resonators both on the 1-st (for type [2] oscillators), and on 3-rd mechanical harmonics (for DTCXO with frequency synthesizer. The test conducted have shown that the oscillator of Fig. 9 does not provide reliable excitation of dual-mode resonators on crystal mechanical harmonics. Besides, using this circuit we could not manage to excite resonators with great values of excited mode motional resistances, or with great multiplicity of these resistances. Self-oscillator on a single transistor has very few degrees of freedom. Changing capacitance of the load in the emitter circuit we can effect both dual-frequency oscillation exciting conditions and a choice of excited harmonic N. It is difficult to find here acceptable optimum.

In this connection an improved circuit [20] was proposed on the base of two-transistor oscillator (See Fig. 16a)

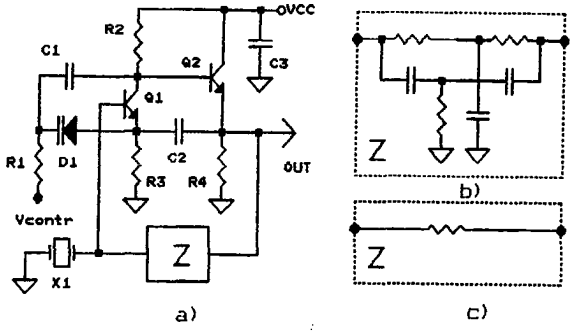


Fig.16. The two-transistors dual-frequency oscillator network.

Depending on crystal resonator parameters, the negative feedback circuit Z can be assembled by various ways (Fig. 16b, Fig. 16c).
Double "T" - bridge tuned to a frequency Fc promotes suppression of the relaxation to which this circuit is inclined sometimes. Adjustment of parameters of the two-frequency regime is carried out by changing oscillator regeneration index by way of alteration of varicap D1 capacitance. Refusal of excited modes level self-adjustment at by way of regime alteration by d.c., made it possible to improve oscillator frequency regime instability.

The dual-frequency oscillator has been implement of two hibrid LCIs: M1 and M2 (Fig. 17)

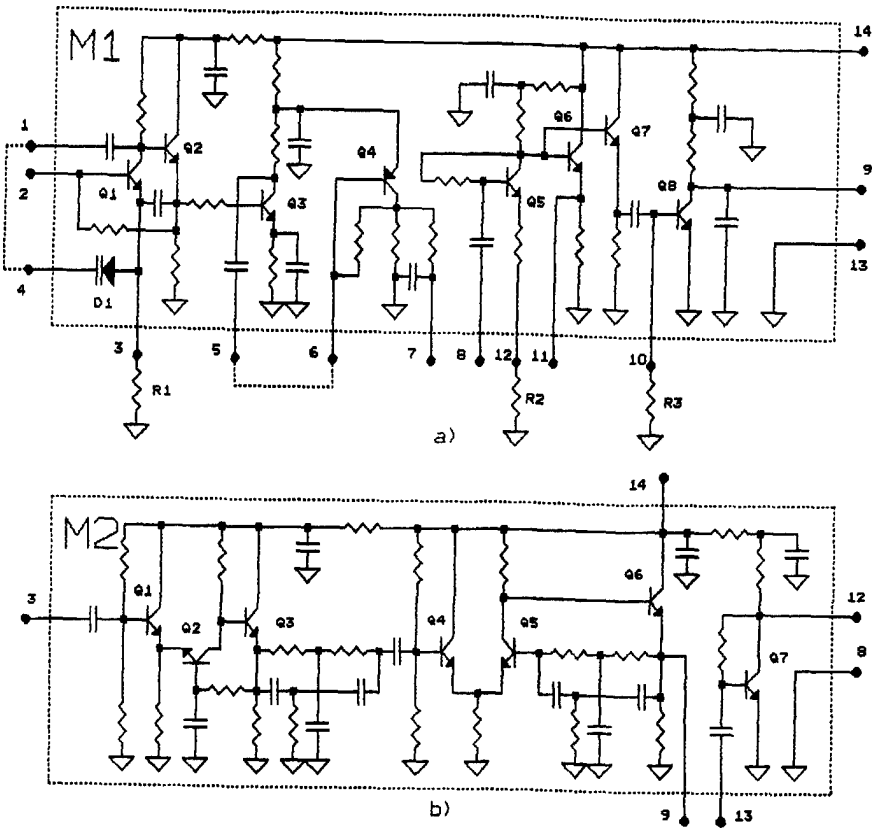


Fig.17. The dual-frequency oscillator microcirquit realization.

The M1 microcircuit comprises:

- the self-oscillator itself, assembled of transistors Q1 and Q2;
- Q3 is buffer;
- Q4 is two-frequency signal envelope detector;
- Q5, Q6, Q7 are difference frequency amplifier;
- Q8 is difference frequency amplitude detector.

The M2 microcirquit comprises, based on transistors Q1-Q6 difference frequency selective amplifier featuring a rise at the difference frequency and a drop - at the base frequency. An amplifier-former of the frequency Fc is assembled of transistor Q7. The circuit, of course, is not a simple one. Its complexity is, mainly, due to its amplifiers and buffer needed to generate stable pure spectrum signal. Fig. 18 shows the microcircuits interconnection. Usually self-oscillator operation adjustment is carried out by means of one resistor R2 (Fig. 17) changing control characteristic of modes level balancing circuit. Fig.19 illustrates how output oscillation spectrum composition is changing in this moment. The other parameters of two-frequency oscillator have been tabulated.

In this work the ideology of refusal of using LC-loops was realized on the level of hybrid LCI. The next stage of the planned developments was creation of a two-mode oscillator on semi-custom-made matrix LCI, but uncompleted works were held up because of economic depression.

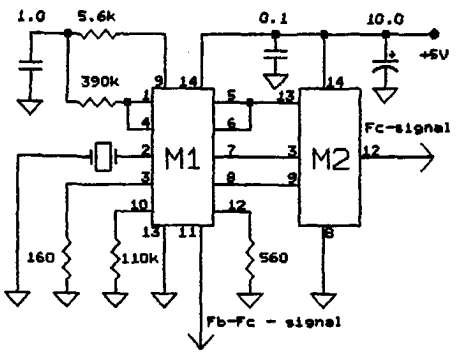


Fig.18. The dual-frequency oscillator microcirquits interconnection.

Table 3. Dual-mode oscillator parameters

Parameter	Volume
Frequency, MHz	10
Regime instability, ppm	0.3/10% Vcc change
Supply voltage, V	3.5...9
Consumption current, mA	5
Fb-mode level, dB	-40

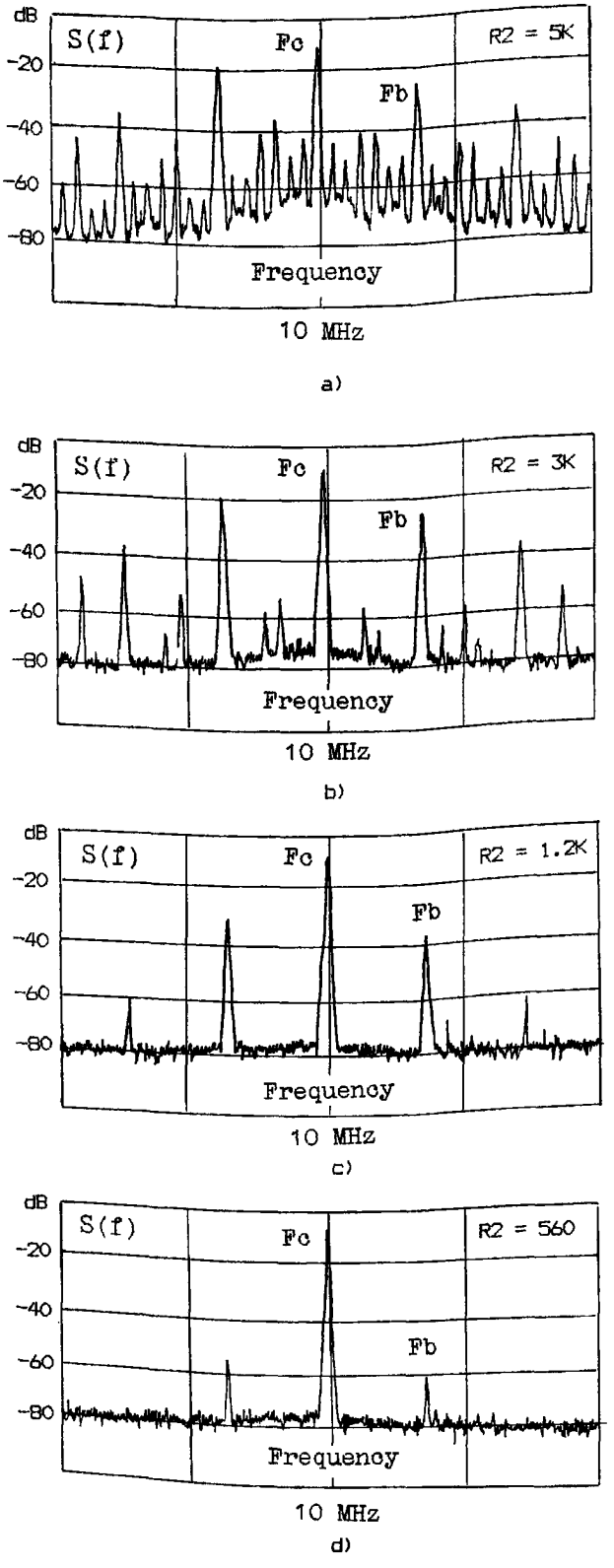


Fig. 19. The oscillator output signal spectrums.

CONCLUSION

The results of researches presented in this paper show that creation of two-frequency oscillators excited on B- and C- modes is possible. For the purpose of thermocompensation, the simultaneously excitation of B- and C-modes is more preferable, than the excitation of the 1-st and the 3-rd harmonics of C-mode. The most reliable results are obtained when using SC-cut resonators with 10 Mhz on 1-st mechanical harmonic. Results obtained with 10 MHz resonators excited on the 3-rd harmonic look promising. We had groups of resonators with good percentage of useful articles (without drops in the temperature range). In addition, the percentage of useful articles was about the same as for resonators of the 1-st harmonic before optimization of their design. Resonators of TD, SC, IT, DP (13 54 /35) -cuts were excited, and some interesting (not yet completed) results were obtained.

For excitation of dual-mode resonators special circuits suitable for microelectronics technology were designed.

All this gives grounds to hope that the dual-frequency oscillators excited on B-and C-modes will be widely used in DTCXO- technology.

REFERENCES

1. Kusters J.A., Leach J.G., Ficher M.S. Quartz Resonator Cut to Compensate For Static And Dynamic Thermal Transient. US Patent N 4079280 1978.

2. Багаев В.П., Косых А.В., Лепетаев А.Н., Самойленко В.Ф. Двухмодовый кварцевый генератор с цифровой термокомпенсацией (Dual-mode digitally compensated crystal oscillator). Электросвязь (USSR), 1986, N 3, p. 48-51.

3. S.S. Shodowski. Resonator Self-Temperature-Sensing Using a Dual-Harmonic-mode Crystal Oscillator. Proc. 43-rd A.F.C.S., 1989, p. 2-7.

4. R.L.Filler, J.R.Vig. Resonator For the Microcomputer Compensated Crystal Oscillator. Proc. 43-rd A.F.C.S., 1989, p.8-15.

5. M.Bloch, M.Meirs, J.Ho. The Microcomputer Compensated Crystal Oscillator (MCXO). Proc. 43-rd A.F.C.S.,1989, p. 16-19.

6. Benjaminson A., Stalling S.C. A Microcomputer-compensated Crystal Oscillator Usng a Dual-Mode Resonator. Proc. 43-rd A.F.C.S., 1989, p.20-26.

7. Альтшуллер Г.Б. Кварцевая стабилизация частоты. М.: Связь, 1974.

8. USSR Invention Certificate N 758472. Crystal oscillator. /L.S.Marianovski, G.V.Vasecky. 1980

9. Kosykh A., Ionov B., Vasiliev A. Temperature-dynamic characteristics and temperature-dynamic compensation of crystal oscillators. Proc. 46-th A.F.C.S., 1992. p. 425-433.

10. A. Kosykh, V. Bagaev, B. Ionov, A. Lepetaev, S. Zavialov, A. Vasiliev. The New Method of Statistic Piecewise-Linear Interpolation and its Application to DTCXO Creation. Proc. 47-th A.F.C.S., 1993

11. Будяков И.В., Липко С.И., Прокопенко В.Г. Основные структуры двухмодовых кварцевых генераторов. (The Base Structures of dual-mode oscillators) /Д.С.Л. ВИНИТИ N 5040-В86 от 14.07.86.

12. Исследование и разработка высокостабильных задающих генераторов с цифровой термокомпенсацией частоты на основе многослойных БИС. (Research Work Report: Development and Research of High-stable Digitally-compensated oscilletors based on multy-layers LCIs) /Отчет по НИР, N Госрегистрации 01850067133. Руководитель Л.К.Самойлов, Таганрог, 1987.

13. Ю.С.Шмалий, Ю.И.Евдокименко. Модуляционный метод адаптации опорных кварцевых генераторов (Modulation Method of Adaptation of Reference Crystal Oscillators). В кн. Стабилизация частоты. Часть 1, ВИМИ, стр. 75-78. (USSR).

14. USSR Invention Certificate N 1709492. Устройство термокомпенсации кварцевого генератора (Crystal oscillator thermocompensation device). Ю.С.Иванченко, В.Ф.Самойленко, И.И.Буймала //Открытия, изобретения, 1992, N 4.

15. Анисимов В.В. О возбуждении бигармонических колебаний в генераторах с двумя степенями свободы. (On Excitation Biharmonic Oscillation in Oscillators Having Two Degree of Freedom). Вестник МГУ, N 1, 1956, стр. 137-146.

16. USSR Invention Certificate N 1145450. Термокомпенсированный кварцевый генератор. (Thermocompensated Crystal Oscillator). //Самойленко В.Ф. Изобретения, открытия. N 10, 1985.

17. Отчет по НИР "Разработка и исследование аппаратуры "Дельта-1" (Research Work Report: "Development and Research of Delta-1 Equipment") Руководитель В.П.Багаев. 1983.

18. Abramson I. Two-mode quart resonator for digital temperature compensated quartz oscillators. Proc. 46 A.F.C.S., 1992.

19. USSR Invention Certificate N 1573524. Двухчастотный кварцевый генератор. (Dual-frequency Crystal Oscillator) В. П. Багаев, С.А. Завьялов./Бюллетень изобретений N 23, 1990 г.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

TEMPERATURE COMPENSATION OF CRYSTAL OSCILLATORS USING MICROCONTROLLER - μCTCXO

Dejan Habić and Dragan Vasiljević

Faculty of Electrical Engineering, University of Belgrade
Bulevar Revolucije 73, P.O.Box 816, 11 000 Belgrade, Yugoslavia

Abstract

The design of compensation circuit in microcontroller temperature compensated crystal oscillator (μCTCXO) and design of production line (ACCL) for oscillator automatic calibration are presented.

The communication property of μCTCXO is introduced in order to provide automatic calibration. The shortened successive approximation algorithm gives short calibration time and reliable convergence of oscillator adjustment by the ACCL system.

The μCTCXO is realized as a low-cost device encapsulated In a single cubic inch volume. The 0.5ppm accuracy of frequency in -40 °C/+85 °C temperature range is obtained by a 9-bit A/D converter, 10-bit D/A output and 512 bytes of software placed in an EEPROM.

Including the standard laboratory frequency counter, temperature chamber and PC host computer the ACCL system requires only the additional analog multiplexer and interface board. The calibration software of 250 Kbytes is executed by the host computer. The ACCL system can compensate many oscillators simultaneously without selection of components and without human control in only one temperature run during production and exploitation.

Introduction

Analog temperature compensated crystal oscillators (TCXO) have a temperature sensing compensation network generating a control voltage applied to a varactor placed in series with oscillator's crystal. A compensation network provides frequency pulling to cancel oscillator frequency temperature drift. Compensation network adjustment requires three to five temperature runs. In order to accelerate production, computer aided methods may be used in compensation network element selection process [1,2].

Advances in integrated circuit technology introduced digital temperature compensation techniques (DTCXO). In comparison to analog TCXOs, they provide higher accuracy for wide temperature ranges, and simplify the adjustment procedure with fewer temperature runs.

The initial DTCXO solutions were based on a look-up table saved in a nonvolatile memory incorporated in the compensation circuit [3,4,5,6]. An almost linear temperature sensor generates voltage which is then converted by an A/D converter having an output which serves as the memory address. The selected memory location content corresponds to the required varactor voltage to compensate oscillator frequency drift [1,2]. A D/A converter converts back the digital memory contents to the analog varactor voltage.

The microcomputer compensated crystal oscillators (MCXO) were obtained in further development of digital compensation. Nonvolatile memory is replaced by microcomputer which improves compensation flexibility. The A/D converter gives temperature in digital form. This data are used by a microcomputer to generate varactor voltage using a small look-up table and interpolation calculation [7].

The above described digital compensation techniques are similar to the analog compensation producing a temperature dependent voltage for varactor. The dual-mode MCXO oscillator is introduced to improve compensation accuracy [8,9,10]. This circuit is based on the direct digital frequency synthesis. The difference between scaled overtone and fundamental oscillating frequency is used as the gate time for period measuring counter which counts scaled overtone frequency. The obtained number represents actual oscillator temperature. Using oscillator

0-7803-1945-1/94 $3.00 © 1994 IEEE

frequency to temperature dependence data taken during temperature calibration, the microcomputer calculates the number of pulses to be deleted from the output signal in order to obtain constant frequency [8]. The dual mode MCXO realization produces high precision compensation.

The purpose of this paper is to introduce a concept of microcontroller compensated crystal oscillator suitable for fully automated production. The oscillator compensation circuit is accomplished by a microcontroller (μCTCXO). It is a small, low-power and cheap device with rich resources and property to communicate, which is necessary for oscillator calibration by an automatic computer controlled line (ACCL). The proposed concept provides unsupervised fully automated temperature compensation of crystal oscillators without component selection in only one temperature run. Many oscillators may be calibrated/recalibrated simultaneously in production/exploitation. The proposed compensation concept supports both digital compensation approaches: dual mode MCXO and classical, temperature dependent voltage generation for VCO input control.

This paper describes the design of μCTCXO and ACCL system. Temperature compensation circuit of the μCTCXO is elaborated in the first section. The automated production line ACCL is presented in the second section. Finally, experimental results are reported in the third section.

Design of μCTCXO

The μCTCXO temperature compensation circuit (temperature dependent voltage generation approach) is shown in Figure 1. The temperature sensor voltage is digitized by an A/D converter. The CPU accepts temperature data, finds the corresponding data from the look-up table taken during calibration, and makes the interpolation to calculate the oscillator's VCO input voltage required to obtain nominal output frequency. The voltage is converted back to analog form by a D/A converter.

The general purpose 8-bit microcontrollers have all necessary resources for the realization of the described compensation circuit. The CPU renders programmability and calculations. The look-up table is saved in a non-volatile memory. The analog values are coupled to the microcomputer by A/D and D/A converters. Microcontroller resources such as frequency dividers, timer/counters and arithmetics/logic functions may be used to realize dual mode MCXO in combination with external mixers and filters. Finally, microcontroller ability to communicate allows calibration of many oscillators in a single temperature run by a fully automated ACCL line. Oscillator identification is obtained by a hardware key connected to the microcontroller's input digital port.

The μCTCXO circuit has two operating mode: calibration mode and autonomous modes.

The simple software which controls autonomous operation of the μCTCXO includes routines for control of the A/D and D/A subsystems and for search and interpolation of data from the look-up table. Compensation accuracy is determined by quality of components.

Communication between the μCTCXO and ACCL system during calibration is realized through an asynchronous serial RS232 interface. Calibration software

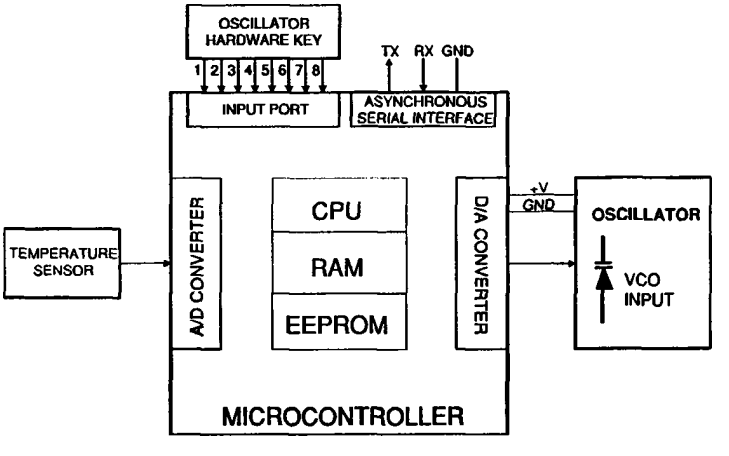


Fig.1 μCTCXO compensation circuit.

is a part of the automated production line ACCL and not of the μCTCXO itself.

Design of automated production line ACCL

The ACCL system is shown in Fig.2. The host computer controls the temperature chamber and frequency counter operation through a standard IEEE 488 bus interface to provide temperature adjustment and frequency measurement. The interface board is used as an adapter between the host computer's ISA bus and the asynchronous serial RS232 interface intended for communication with μCTCXOs and the analog multiplexer. Three wire serial connection allows simple handling of a large number of μCTCXOs. The intelligent analog multiplexer provides connection of oscillators signal output to the frequency counter.

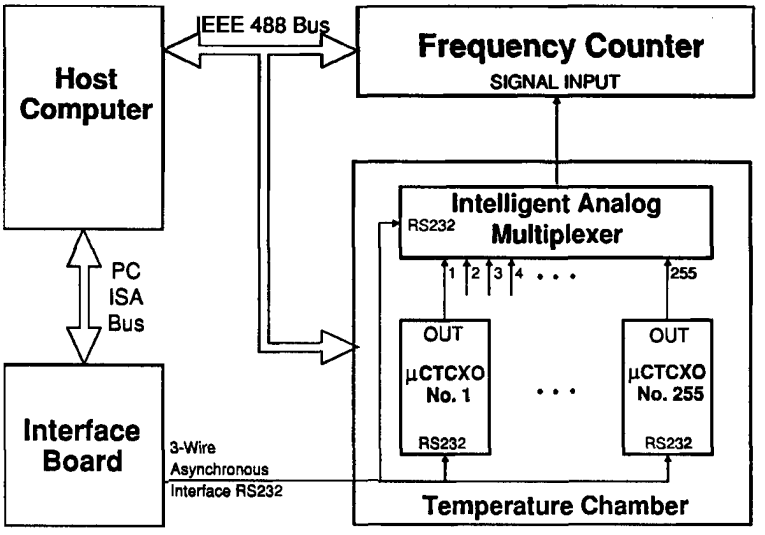


Fig.2 ACCL system.

The ACCL system operates under control of the calibration software executed on the host computer. The calibration process includes three tasks: communication, calibration and programming.

The communication is started when host computer downloads the program for communication to all μCTCXO devices simultaneously. Then host computer and selected μCTCXO exchange the address defined by the μCTCXO's hardware key. Signal out of the selected oscillator is connected to the frequency counter using analogue multiplexer.

The calibration process is open when the host computer gives information about desired temperature to the chamber using 488 bus. By the same bus host computer reads frequency counter. When oscillators frequencies are stabilized, the host decides that the oscillators transient thermal regime is finished and the temperature is held constant. Then the adjustment of selected μCTCXO is started. Using asynchronous serial interface the host computer reads the A/D converter and attempts to pull the oscillator frequency to the nominal value through adjustment of the D/A converter input in the first iteration. Now the 488 interface is activated again. The host computer reads the frequency counter. When the oscillator frequency is stabilized, the host computer compares it with the target frequency. If the difference is satisfactorily small, the adjustment is finished. If not, the host computer changes the D/A converter input and the process is continued until target frequency is achieved. When adjustment is finished, the host computer defines the new temperature point, and the calibration at this temperature is carried out in the same way. The number of temperature points is determined by the uncompensated oscillators frequency to temperature characteristics $(\Delta f/f)/\Delta T$ and the desired accuracy of compensation [1,7]. Obviously, the processing time is proportional to the number of temperature points.

When calibration is finished, the host computer programs μCTCXO's EEPROM with contents of look-up table and program for oscillator operation in autonomous regime.

Two problems had to be solved during the development of the described algorithm. The first problem is the determination of the initial value for the D/A input in order to obtain minimum number of approximation steps and shortest adjustment time. The right choice presumes a value as close as possible to the varactor voltage which gives nominal oscillator frequency. The initial value can be estimated on the basis of experimentally determined data on influence of every bit of the D/A converter on the oscillator frequency change Δf for each μCTCXO family. These data are placed in the host computer which uses it together with the difference Δf ' between the measured operating and nominal frequency in order to estimate the starting value of the D/A converter input.

The second problem is how to select D/A input in the second and other further steps in order to obtain convergence of the adjustment procedure. The successive approximation algorithm, common in A/D conversion, is a good choice due to the reliable and fast convergence. The proposed procedure represents a shortened successive approximation algorithm.

When many μCTCXO devices are adjusted simultaneously, the input to each oscillator's D/A should be defined as soon as a new temperature point is selected. In this way, all μCTCXOs will settle their frequencies simultaneously and adjustment time is shortened.

Experiments

The compensation circuit of μCTCXO device is realized by a Motorola 68HC11 microcontroller and by a KTY silicon temperature sensor, Fig.1. The passive second order RC filter is added to complete pulse-width D/A converter circuit. The μCTCXO autonomous mode operation is controlled by a 512 bytes program including 95 bytes look-up table.

The frequency stability better than +/-0.5ppm in the range of operating temperatures of (-40 °C , +85 °C) is obtained by a 9 bit A/D converter and 10 bit pulse-width D/A converter, Fig.3. The look -up table is obtained in 30 points for AT cut crystals with uncompensated frequency to temperature curve with maximum frequency deviation Δf/f=+/-25ppm and maximum slope (Δf/f)/ΔT=1.8ppm/ °C.

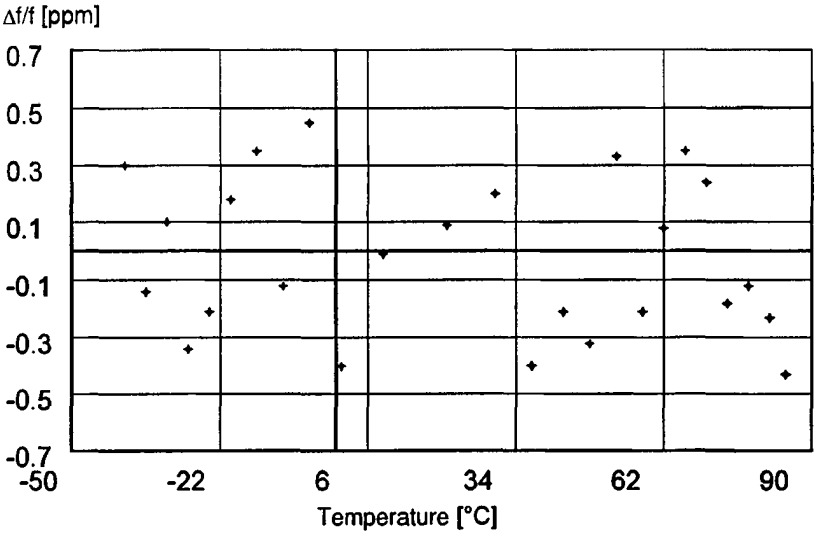


Fig.3 Frequency to temperature characteristic of a compensated μCTCXO.

The compensation circuit of the μCTCXO is built by general purpose and low-cost components. The entire μCTCXO is encapsulated in a standard 36mmX27mmX19mm size case. The compensation circuit current consumption is low (10mA/5V) in a wide range of operating temperatures (-40 °C , +85°C).

The ACCL line includes standard instruments such as laboratory frequency counter and temperature chamber with IEEE 488 bus. A PC is used as a host computer (Fig.2.). The ACCL line also includes a specially designed interface board and an intelligent analog multiplexer. A 68HC11 microcontroller and a RAM buffer memory make up the interface board. The analog multiplexer has the same microcontroller and an analog multiplexer tree with 255 inputs. The ACCL system is under control of a 250 Kbytes software under DOS.

The operation of fully automated ACCL system during calibration is illustrated in Fig.4. The number of attempts to obtain nominal frequency at a selected constant temperature depends on the closeness of the varactor voltage initial value and its final value. The diagram also illustrates the convergence of the shortened successive approximation algorithm.

The ACCL system calibration time for three μCTCXO devices is shown in Fig.5. The dark rectangle represents the time interval needed for temperature stabilization in the temperature chamber and oscillators. The gray rectangles designate time consumed by the ACCL software during execution of the shortened successive approximation algorithm and time spent for oscillator settling after change of varactor voltage. The main problem in design of ACCL system is to discover efficient calibration algorithm to save adjustment time.

Conclusions

This paper presents the design of a temperature compensation network in microcontroller compensated crystal oscillators (μCTCXO) and the design of production line (ACCL) for automatic oscillator calibration.

The μCTCXO circuit is built by general purpose low cost components. It is a small, intelligent unit able to communicate with the outside world.

The proposed concept of ACCL system provides fully automated, unsupervised temperature compensation of crystal oscillators without component selection in a single temperature run. Many oscillators may be calibrated simultaneously in production and later, during exploitation. Both digital compensation approaches are supported: dual mode MCXO and classical, temperature dependent voltage generation for VCO input control. Ef-

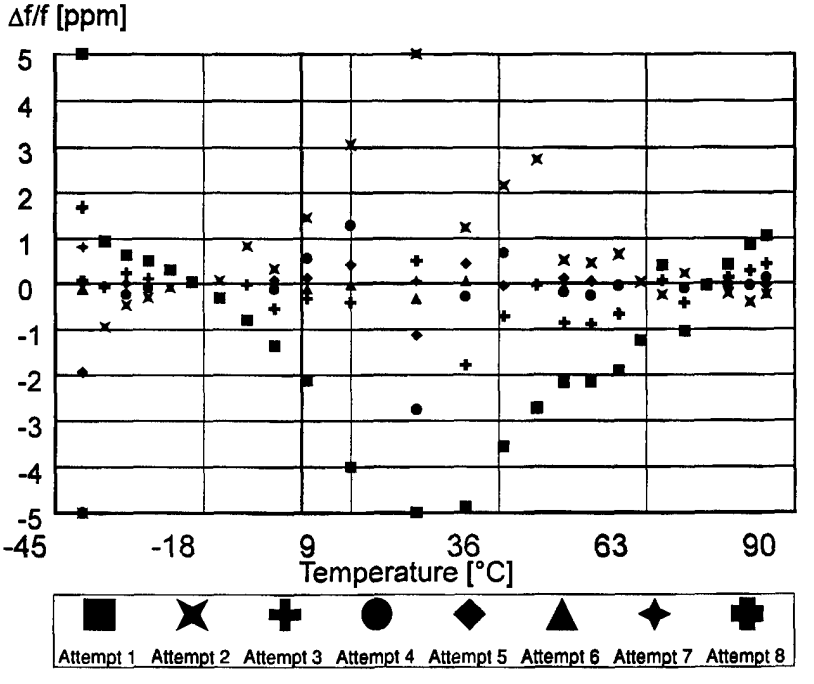


Fig.4 Operation of ACCL system during calibration.

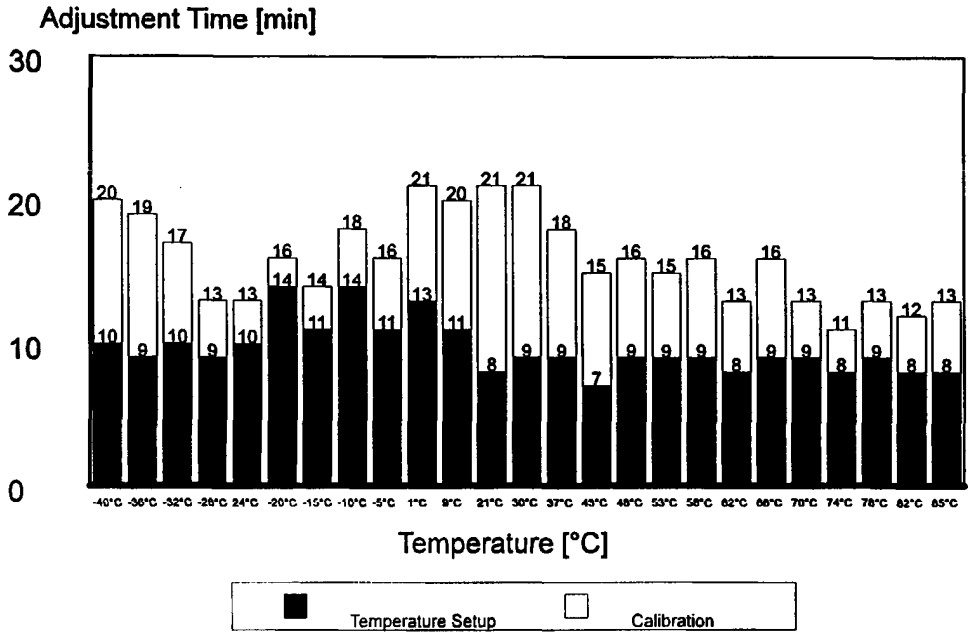


Fig.5 Time consumption of ACCL algorithm.

ficient shortened successive approximation algorithm provides fast and reliable convergence. Upgrading of a standard nonautomatic calibration system to an ACCL system is inexpensive, since few low-cost additional components are needed.

Currently, we are investigating further calibration algorithm improvements to shorten the calibration time even more.

References

[1] M.E.Frerking, "Method of Temperature Compensation," in Proceedings of the 36th Annual Frequency Control Symposium , 1982, pp. 564-570.

[2] V.Candelier, G.Caret and A.Debaisieux, "Low Profile High Stability Digital TCXO; Ultra Low Power Consumption TCXO," in Proceedings of the 43rd Annual Frequency Control Symposium , 1989, pp. 51-53.

[3] R.Rubach, "Dual Mode Digitally Temperature Compensated Crystal Oscillator," in Proceedings of the 36th Annual Frequency Control Symposium, 1982, pp. 571-575.

[4] E.K.Miguel, "A Temperature Compensated SC Cut Quartz Crystal Oscillator," in Proceedings of the 36th Annual Frequency Control Symposium , 1982, pp. 576-585.

[5] Takehiko Uno and Yoshio Shimoda, "A New Digital TCXO Circuit Using a Capacitor-Switch Array," in Proceedings of the 37th Annual Frequency Control Symposium , 1983, pp. 434-441.

[6] Ž.Aleksić and D.Vasiljević, "Digital Temperature Compensation of Crystal Oscillators Using Temperature Switches," in Proceedings of the 40th Annual Frequency Control Symposium , 1986, pp. 340-343.

[7] T.Miyayama, Y.Ikeda and S.Okano "A New Digitally Temperature Compensated Crystal Oscillator for a Mobile Telephone System," in Proceedings of the 42nd Annual Frequency Control Symposium , 1988, pp. 327-333.

[8] A.Benjaminson and S.C.Stallings, "A Microcomputer-Compensated Crystal Oscillator Using A Dual Mode Resonator ," in Proceedings of the 43rd Annual Frequency Control Symposium , 1989, pp. 20-26.

[9] R.L.Filler, J.A.Messina and V.J.Rosati, "Frequency-Temperature and Ageing Performance of Microcomputer-Compensated Crystal Oscillators," in Proceedings of the 43rd Annual Frequency Control Symposium , 1989, pp. 27-33.

[10] M.Bloch, M.Meirs, J.Ho, J.R.Vig and S.Schodowski, "Low Power Timekeeping ," in Proceedings of the 43rd Annual Frequency Control Symposium , 1989, pp. 34-36.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

TCXO'S EMPLOYING NS-GT CUT QUARTZ CRYSTAL RESONATORS FOR CELLULAR TELEPHONES

Eiki Takahashi, Osamu Teshigawara, Kazuo Yamashita

Japan Radio Co.,Ltd.
2-1-4 Kamifukuoka, Saitama 356, Japan

Abstract

This paper describes temperature compensated crystal oscillators (TCXO's) employing new shape GT cut quartz crystal resonators (NS-GT cut resonators) to be used in cellular telephones. The NS-GT-TCXO has an oscillation frequency of 2.4MHz and a frequency stability of ±1.5ppm within a temperature range of -20°C to +70°C using a simplified temperature compensation method. These superior characteristics can be achieved with low current consumption (approximately 0.8mA) and compactness (11.5mm x 8mm x 4mm). Additionally, the phase-noise characteristics of the NS-GT-TCXO have been evaluated. The phase-noise value of the phase-locked loop circuit using the NS-GT-TCXO as a reference oscillator is approximately -70dBc/Hz at 50Hz offset from the 614.4MHz carrier. This meets the requirements for use in cellular telephones. These results indicate that the miniaturization and low power consumption of the NS-GT-TCXO could have an important impact on the development of future cellular telephone systems.

Introduction

TCXO's employing AT cut quartz crystal resonators (AT-TCXO's) are commonly used as reference oscillators for the synthesizers of cellular telephones. However, due to the many components and a complex temperature compensation technique required by the AT-TCXO, it has been difficult to realize further miniaturization and lower power consumption, while at the same time holding down costs. On the other hand, NS-GT cut resonators exhibit excellent frequency stability for temperature. As reported in reference [1], crystal oscillators employing NS-GT cut resonators having a frequency stability of ±2.5ppm for a temperature range of -30°C to +85°C with no external temperature compensation can easily be mass produced. Moreover, the second and the third order frequency temperature coefficients of NS-GT cut resonators are extremely small compared to its first order frequency temperature coefficient. NS-GT-TCXO's for cellular telephones (requiring a stability of ±1.5ppm within a temperature range of -20°C to +70°C) can easily be achieved by compensating only the first order frequency temperature coefficient.

The objective of this paper is to clarify that the proposed method for constructing TCXO's is suitable for miniaturization, low power consumption, and cost effectiveness. The method for temperature compensation is to utilize the temperature characteristics of ceramic capacitors. Because the ceramic capacitors exhibits linear capacitance change for temperature, the first order frequency temperature coefficient of NS-GT cut resonators can be compensated by applying the ceramic capacitors to load capacitance.

In addition, this paper describes development of the NS-GT-TCXO's for mass production. For this purpose, the selection of the proper load capacitance, the classification of the NS-GT cut resonators, and the reduction of the variation of temperature compensation are required.

We have also evaluated the phase-noise performance of the NS-GT-TCXO. It is because the phase-noise performance near the carrier is of special importance in digital cellular telephone systems. The favorable results have been obtained for this requirement.

Temperature Compensation

Figure 1 shows the measured frequency temperature characteristics of an AT cut resonator,

0-7803-1945-1/94 $3.00 © 1994 IEEE

an NS-GT cut resonator, and the NS-GT-TCXO. The AT cut resonator is unstable for temperature and its third order frequency temperature coefficient is large. On the contrary, the NS-GT cut resonator has excellent frequency stability and almost linear characteristics. However, it is not stable enough for cellular telephones. In other words, it is too costly to select the required resonators among the mass produced resonators. The proposed method is to compensate only the first order frequency temperature coefficient of the NS-GT cut resonator. As a result, the superior characteristics of the NS-GT-TCXO are obtained with cost effectiveness.

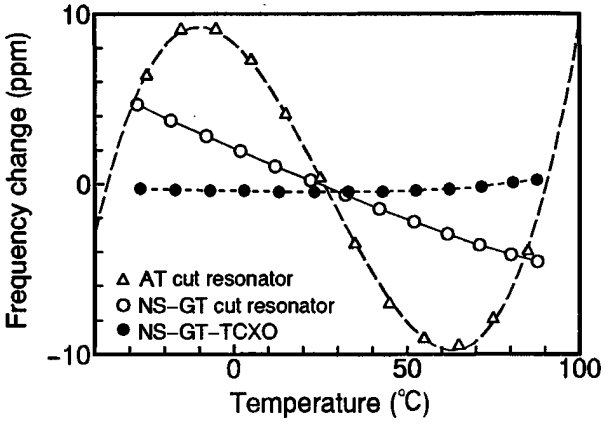


Figure 1. Measured frequency temperature characteristics of an AT cut resonator, an NS-GT cut resonator, and the NS-GT-TCXO.

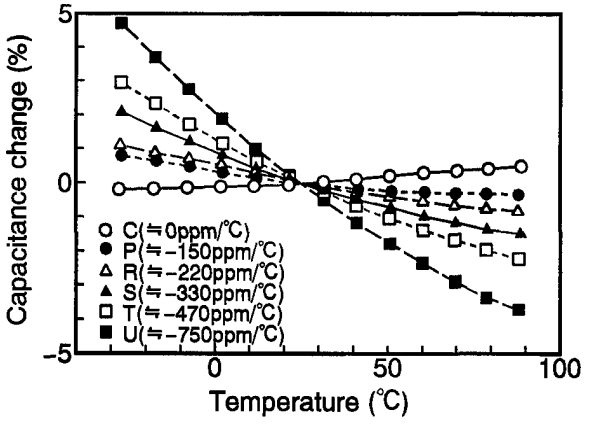


Figure 2. Measured capacitance temperature characteristics of the ceramic capacitors used for the NS-GT-TCXO. The C, P, R, S, T, and U show the kinds of the ceramic capacitors.

Temperature compensation is achieved by taking advantage of the capacitance temperature characteristics of the ceramic capacitors used for load capacitance. Figure 2 shows the measured capacitance temperature characteristics of the ceramic capacitors used for the NS-GT-TCXO. These capacitors are widely used in consumer products and meet EIA standard requirements. As is apparent from Fig. 2, most capacitors have linear and negative capacitance temperature characteristics. NS-GT cut resonators having only a negative value of the first order frequency temperature coefficient can be compensated with the capacitors.

Figure 3 shows the range of the first order frequency temperature coefficients of the resonators used for the NS-GT-TCXO. The overall range is limited by the compensation capabilities of the ceramic capacitors and the pulling sensitivity of resonators at any given load capacitance. Because the entire range (approximately 5ppm for a temperature range of -20°C to +70°C) is not critical for NS-GT cut resonators [1], these resonators can be cost effectively produced. The range is sub-divided into four groups (A, B, C, and D) by the first order frequency temperature coefficient. The number of groups and their widths are determined by the desired frequency temperature stability (which, in this case, is ±1.5ppm for a temperature range of -20°C to +70°C). Consideration of the tolerance of the temperature compensation and effects of the second and the third order frequency temperature coefficients are also relevant factors in determining the width of each group.

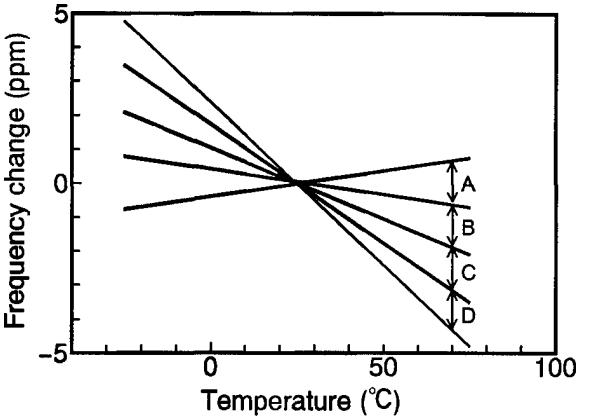


Figure 3. Range of the first order frequency temperature coefficients of the resonators used for the NS-GT-TCXO.

Circuit Design

Circuit Construction

Figure 4 shows the circuit construction of the NS-GT-TCXO. The NS-GT-TCXO consists of CMOS inverters, a trimmer capacitor, a varacter, ceramic capacitors, resistors, and an NS-GT cut resonator. The CMOS inverter (IC_1) and the feedback resistor (R_F) comprise an amplifier for the oscillator. The other CMOS inverter (IC_2) is used as a buffer. The drain output resistor (R_D) controls the drive level of the resonator. Gate load capacitance (C_G) consists of the ceramic capacitor (C_1) and the trimmer capacitor (C_T). Drain load capacitance (C_D) consists of the two ceramic capacitors (C_2, C_3) and the varacter (C_V). The bias resistor (R_C) is used for applying a control voltage (VC) to the varacter.

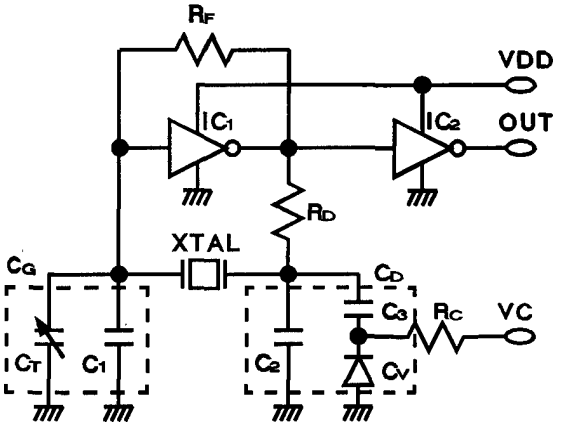


Figure 4. Circuit construction of the NS-GT-TCXO.

Selection of Load Capacitance

In the case of the input and the output reactance of the inverters are negligible, the total load capacitance C_L can be given as

$$C_L = \frac{C_G C_D}{C_G + C_D} \quad (1)$$

where,

$$C_G = C_T + C_1 \quad (2)$$

$$C_D = C_2 + \frac{C_V C_3}{C_V + C_3} \quad (3)$$

Because the load-capacitance change produces a frequency change, the frequency adjustment and the frequency control can be performed by changing the load capacitance. Conversely, load capacitance affects the frequency temperature characteristics of NS-GT cut resonators [2,3]. Figure 5 shows the variation of the first order frequency temperature coefficient versus load capacitance. The variation of the coefficient largely depends on the load capacitance when it is less than 10pF. On the contrary, the variation of the coefficient is almost independent of the load capacitance when it is 10pF or more. Therefore, a load capacitance of 15 pF has been selected.

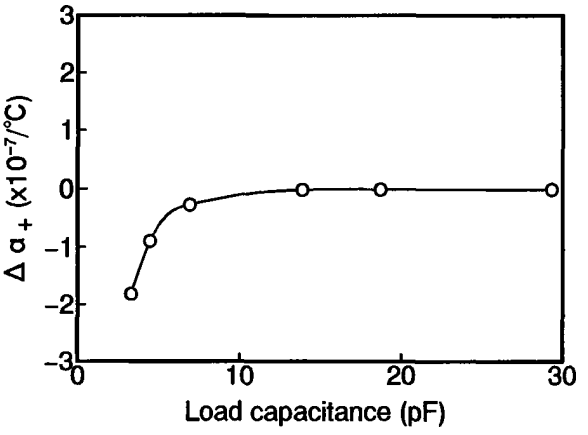


Figure 5. The variation of the first order frequency temperature coefficient of an NS-GT cut resonator ($\Delta\alpha_+$) versus load capacitance. Courtesy of Seiko Electronic Components Co.,Ltd.

Frequency Control Function

A voltage frequency-control function is one of the requirements of cellular telephones. The varacter C_V is used for this purpose. However, the capacitance temperature characteristics of the varacter depends on a control voltage. Figure 6 shows the measured capacitance temperature characteristics of the varacter when the control voltage is taken as a parameter. This adaptability of the varacter produces the variations in temperature compensation of the control voltage. This variation of temperature compensation increases as the frequency-control sensitivity becomes greater. Our cellular telephones have required a frequency-control sensitivity of 1.5ppm/V. As a result, the variation of temperature compensation for a temperature range of -20°C to +70°C has been less than 0.1ppm for a control voltage of 1V to 3V. This is negligible for the target stability.

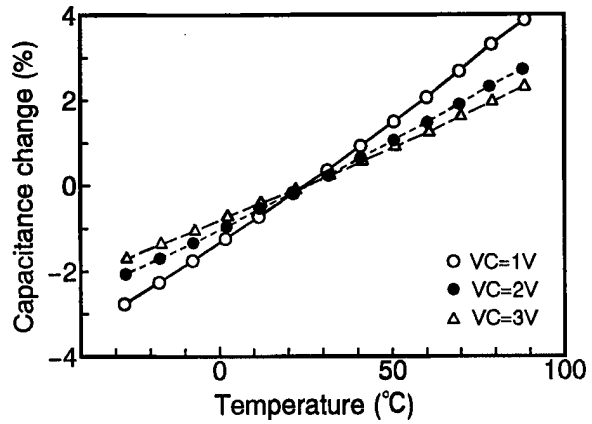


Figure 6. Measured capacitance temperature characteristics of the varacter when the control voltage is taken as a parameter.

Frequency Adjustment Function

A frequency adjustment function is required to correct for all of the reactance tolerances of the components and the resonance-frequency tolerance of the resonator. Although the frequency control function of the varacter can include frequency adjustment by increasing the frequency-control sensitivity, the variation of temperature compensation by a control voltage increases as described in the previous section. Therefore, a trimmer capacitor has been used for the frequency adjustment. A frequency adjustment range of ±15ppm has been decided to cover a resonator frequency tolerance of ±10ppm.

The capacitance change of the trimmer capacitor affects the voltage frequency-control sensitivity. This effect increases in the case of the trimmer capacitor and the varacter are located on the same side of either the gate or the drain load capacitance. To avoid this effect, the trimmer capacitor and the varacter are located on different sides (see Fig. 4).

Temperature Compensation Technique

As shown in Fig. 3, the range of the first order frequency temperature coefficients of NS-GT cut resonators is sub-divided into four groups based on temperature compensation. Therefore, four kinds of circuits, each having different capacitance temperature coefficients of load capacitance are required. Each circuit is designed to have a capacitance temperature coefficient which produces an equal and opposite frequency change of the center of each group shown in Fig. 3. The desired capacitance temperature coefficient is obtained by combining the capacitance temperature coefficients of the three ceramic capacitors C_1, C_2, and C_3.

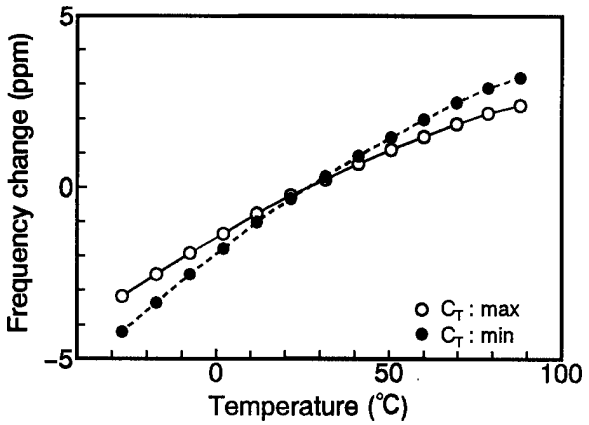


Figure 7. The frequency change induced by temperature compensation when the capacitance of the trimmer capacitor is the maximum value and the minimum value in an improperly designed NS-GT-TCXO.

The most important technique of this temperature compensation is to reduce the variations in temperature compensation. The variation mainly results from the load capacitance change induced by the trimmer capacitor. Figure 7 shows the amount of temperature compensation when the capacitance of the trimmer capacitor is the maximum value and the minimum value in an improperly designed NS-GT-TCXO. As is apparent from Fig.7, the variation is too large to neglect for the target stability. To reduce the variation, the temperature coefficients of the trimmer capacitor and the ceramic capacitors must be equal. Therefore, two kinds of trimmer capacitors having different capacitance temperature coefficients are used in accordance with the extent of temperature compensation. A trimmer capacitor having a capacitance temperature coefficient of 0ppm/°C is used for group A and B resonators (Fig. 2). In addition, a trimmer capacitor having a capacitance temperature coefficient of -750ppm/°C is used for group C and D resonators. By means of this technique, the temperature coefficients of the trimmer capacitor and the ceramic capacitors can be balanced and the varia-

tion has been reduced less than 0.2ppm for a temperature range of -20°C to +70°C.

Phase-Noise Characteristics

Crystal oscillators employing NS-GT cut resonators are designed with extremely low drive level in order to suppress spurious oscillations. This NS-GT-TCXO has been designed with a drive level of under 5uW. This low drive level may degrade the phase-noise characteristics of the oscillators. Moreover, the phase-noise performance of a synthesizer at offset frequencies below a loop bandwidth is dominated by a reference oscillator. Precise modulation accuracy is required in digital cellular telephones, which means that the phase-noise performance in the immediate vicinity of the carrier is of great concern. Relevant to that, the phase-noise characteristics of the NS-GT-TCXO have been evaluated.

Figure 8 shows a block diagram of the phase-locked loop circuit used for the phase-noise evaluation. Because the loop bandwidth is designed with 1.5kHz, phase-noise values at offset frequencies below 1.5kHz are dominated by the performance of the NS-GT-TCXO as a reference oscillator. Figure 9 shows the results of the phase-noise measurement. The phase-noise requirement for a cellular telephone system are also shown in Fig. 9. The measured data shows the phase-noise characteristics of the NS-GT-TCXO are suitable for use in cellular telephones.

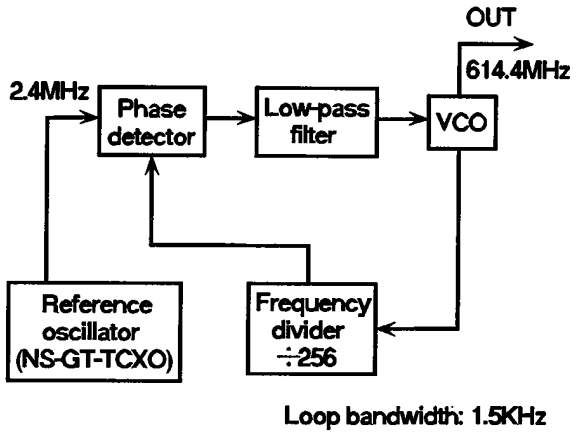


Figure 8. Block diagram of the phase-locked loop circuit used for the phase-noise evaluation.

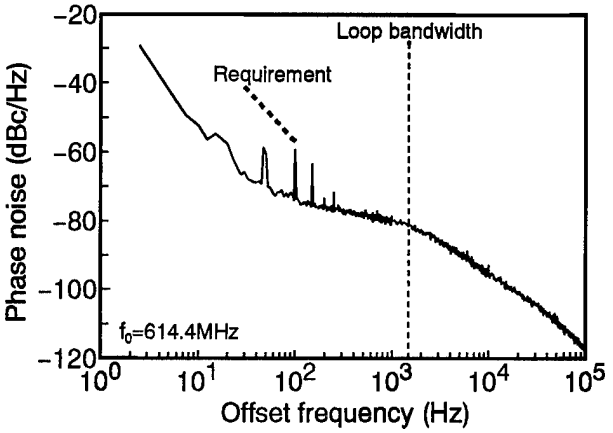


Figure 9. Measured phase-noise of the phase-locked circuit using the NS-GT-TCXO as a reference oscillator.

Configuration

Figure 10 shows the configuration of the NS-GT-TCXO. The NS-GT-TCXO is a surface-mount technology (SMT) module. The components are mounted on both sides of a glass-epoxy substrate. The NS-GT cut resonator is mounted in a ϕ3mm x 8mm tubular package. This NS-GT-TCXO is a mere 11.5mm x 8mm x 4mm, in spite of using ordinary and cost effective components. Further miniaturization could be feasible by means of circuit integration and the use of smaller components.

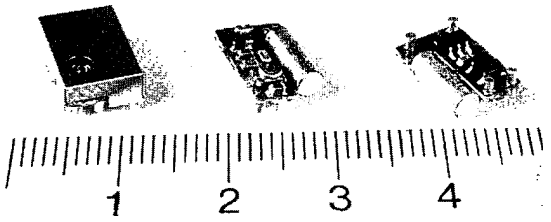


Figure 10. Configuration of the NS-GT-TCXO

Current consumption

Under conditions wherein the source voltage is 4V and an output load-impedance of parallel connection of a 10pF capacitor and a 10kΩ resistor, the current consumption of the NS-GT-TCXO is approximately 0.8mA. This low current consumption results from the dual facts that the NS-GT cut resonator has a relatively low resonance frequency of 2.4MHz and that the CMOS inverters are used.

Conclusion

This paper has demonstrated TCXO's employing NS-GT cut resonators for use in cellular telephones are easily achieved by compensating only the first order frequency temperature coefficient of NS-GT cut resonators and this method has advantages for miniaturization, power consumption, and cost effectiveness.

In addition, we have successfully mass-produced the NS-GT-TCXO's. The developed NS-GT-TCXO has an oscillation frequency of 2.4MHz, a frequency stability of ±1.5ppm within a temperature range of -20°C to +70°C, low current consumption of approximately 0.8mA and a small size of 11.5mm x 8mm x 4mm. In addition, the phase-noise characteristics of the NS-GT-TCXO satisfy the requirements for cellular telephones. Further cost effective, miniaturization of the NS-GT-TCXO is possible.

Acknowledgements

The authors wish to thank Dr. H. Kawashima, Mr. M. Matsuyama, and Mr. K. Sunaga of Seiko Electronic Components Co.,Ltd. for helpful discussions and for providing samples of the NS-GT cut resonators.

References

[1] H. Kawashima and K. Sunaga: "Temperature Compensated Crystal Oscillator Employing New Shape GT Cut Quartz Crystal Resonator",Proc. the 45th Annual Symp. on Frequency Control, pp.410-417(1991)

[2] H. Kawashima and K. Sunaga: "Voltage Controlled S-TCXO's Employing NS-GT Cut Quartz Crystal Resonators", Proc. 1993 IEEE International Frequency Control Symposium, pp.670-678(1993)

[3] H. Kawashima: "An Analysis of Oscillation Frequency Characteristics in a CMOS Oscillating Circuit Using Coupling Quartz Crystal Resonator", Proc. the 44th Annual Symp. on Frequency Control, pp.585-592(1990)

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

OPERATION OF A NEURAL NETWORK CONTROLLED CRYSTAL OSCILLATOR

D.B. Opie, C.T. Butler, W.M. Golding†, Fred Danzy*, John Yates, Mary C. Sharp

Physical Sciences Inc., Alexandria, VA, *Naval Research Laboratory, Washington, DC
†SFA Inc., Landover, MD

Introduction

The neural network control of a quartz oscillator has been demonstrated. We have shown that a single neural network can correct an oscillator's output frequency while several environmental sources of frequency shift act on the oscillator. The advantage that a neural network offers over microprocessor controlled oscillators or Kalman filter control techniques is that a neural network does not need an a priori model of the physical system to arrive at the correct control algorithm. The results of this demonstration indicate that a neural network can have beneficial applications in a variety of frequency standard devices. We believe that the neural controller will work best as a control system supervisor, rather than as the main controller of the frequency standard system. Specifically, the neural network would learn the non-linear effects associated with the several servos that control the main physical parameters of a frequency standard (temperature, magnetic field, etc.), and correct the oscillator's frequency based on the state of these servos and any additional sensed environmental parameters. The dominant remaining technical issue is the training of the neural network. During this demonstration program, maintaining a stable, reproducible environment that could be varied quickly and randomly over the whole training parameter space proved to be a significant technological challenge. Future efforts will focus on methods to more efficiently train the neural networks and the identification of specific devices.

Oscillator Development and Construction

Approach

The purpose of this investigation was to evaluate the neural network's potential for substantially improving the stability of precision time standards by reducing their sensitivity to environmental changes. All precision frequency standards show some degree of sensitivity to environmental changes [1]. Using a more stable frequency standard as a training source, a single neural network can be taught to act on a less-stable oscillator to sense and compensate for many of these environmental changes. When the training source is removed, the network can continue to control the oscillator as if the more stable standard were still present. This approach has two main advantages over other control methods. First, a neural controller will adaptively evolve the correct transfer function that will transform the input vector to the desired output vector, without prior knowledge of the transfer function's form. Second, errors arising from any number or variety of environmental sources can be compensated by a single control system. These features may simplify the correction of non-linear, difficult to model, or low signal-to-noise processes. Since all frequency standards are subject to such systematic processes, we expect that neural controllers could find application in all classes of frequency standards.

To demonstrate the operation of this device, a test system was designed that would allow for easy training and evaluation of a neural controller. The general architecture for such a network is schematically shown in Figure 1. We used the well understood and highly successful backpropagation neural network design, with one modification. Because the output of the network is unknown, a graded-learning modification to the backpropagation design was used [2]. The neural network was given three temperature inputs, while also being given a training signal consisting of the frequency error between the test and reference oscillators. During the training, the network builds a non-linear internal model of the relationship between the temperature inputs and the output frequency of the test oscillator. Once the network had evolved a sufficiently accurate model, the reference oscillator was disconnected and the network controlled the test oscillator's frequency using the nonlinear model developed during training.

0-7803-1945-1/94 $3.00 © 1994 IEEE

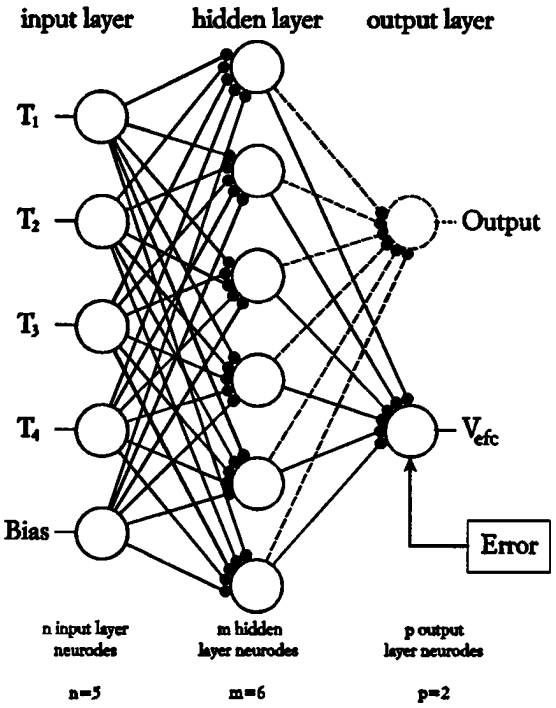


Figure 1. Schematic of a neural network, showing the input, hidden and output layers. For this demonstration, one output neurode was used, but for generality, two are shown in this diagram.

Design and Construction

Figure 2 shows a schematic of this test system. The complete system includes the test oscillator, the reference oscillator, a computer with the neural network simulation program, and a means of providing a randomly chosen training temperature for each of the temperature controlled elements. To aid in the comparison of the test and reference oscillators, the construction of the test system began with two similar oven controlled crystal oscillators (OCXO). One oscillator was used as the reference, while the other was reconfigured as a neural network controlled crystal oscillator (NNCXO). For this reconfiguration, there were several modifications to the oscillator circuit. For example, the varactor diode and associated DC blocking capacitor were replaced with elements that allowed for a broader tuning range. Additionally, the quartz crystal was thermally isolated from the oscillator circuit, whereas the OCXO has the quartz and circuitry in the same ovenized temperature zone. Thermally isolating these components allows the temperature of these two components to be individually measured and separately controlled during the training.

The temperatures of the crystal and oscillator circuit were measured by thermistors and bridge circuits. The voltage across the bridge was used as the input to the neural network. To ensure that the temperature of the crystal and the oscillator circuit were precisely measured, the temperature of the bridge circuit components was also measured and input into the neural network. Although the network is capable of many more inputs, we limited this demonstration to only these three inputs. The selection of these three components was based on the relative strength of each component's contribution to the oscillator's frequency temperature coefficient. The neural controller read these inputs and produced an output signal according to the current transfer function defined by the learned nonlinear model. The network output adjusted the electronic frequency control voltage, V_{efc}, in such a way as to minimize the error function by compensating for the frequency shifts caused by the temperature fluctuations. This error function is defined as the fractional frequency difference between the test oscillator and the reference oscillator.

It is important that the training presentations to a neural network randomly cover the complete parameter space. With the hardware used for this demonstration, randomly covering the parameter space meant producing random temperatures at the sensed components. To achieve this, the computer generated a set of random numbers for each training presentation, and these numbers were sent to the D/A board to produce voltages proportional to these numbers. These voltages were used by the temperature control circuit as set points for each of the control circuits. Each control

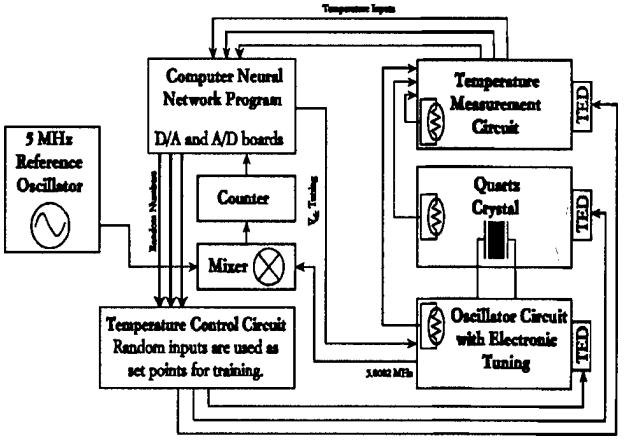


Figure 2. Schematic of the NNCXO test system.

circuit drove a thermal electric device (TED) to provide heating and cooling of the temperature monitored components. The voltages from the D/A board were scaled so that the range of random numbers spanned the desired temperature-training range for the individual components. In this example, a temperature range of six degrees Celsius centered on room temperature was used.

After a time period which allowed the temperatures to settle, the network read the temperatures as inputs to the neural network. A V_{efc} signal was sent to the oscillator, and the frequency of the test oscillator was compared with the reference oscillator to generate an error signal. During training, this error was backpropogated to update the neurode interconnect weights. After the weights were updated, a new input pattern of random numbers was generated and applied to the temperature control circuit, completing the cycle. This cycle was repeated until a sufficiently accurate nonlinear model was developed, that is, until the network was adequately trained.

The period of the training cycles was determined by the component that requires the longest time to equilibrate to the each new temperature set point. In this case, we found it most difficult to measure and control the temperature of the oscillator circuit, which is comprised of two separate circuit boards. Since it is impractical to heat sink each item in this circuit, we chose to regulate the temperature of the air around this component. The air in the enclosure built for the circuit boards takes 6 minutes to reach a steady temperature. This temperature equilibration time limited the period of the training cycles to more than 6 minutes per presentation. At this rate, 5000 presentations required 500 hours, or about 21 days. This may be too long for practical applications, however, this issue can be addressed in future development efforts.

Results

We have found that the NNCXO operates as a frequency standard that is correctly compensated to reduce the influence of changing environmental parameters. The learning process is illustrated in Figure 3. This graph shows the fractional frequency error data recorded during the training, and plotted as a function of the number of training cycles. The error is found by dividing the frequency difference between the reference and test oscillators, Δf, by the total frequency range the uncontrolled oscillator will have between the minimum and maximum temperatures. At the beginning of the training, the error amplitude is large and corresponds closely to the random temperature set points. After nearly 5000 presentations, the magnitude of the error signal has been decreased by more than a factor of ten. It is expected that nearly 10000 presentations are necessary to adequately train the oscillator. Therefore, we expect improved performance as training continues.

After the NNCXO has trained to some sufficient level, no more modifications are made to the nonlinear model and the oscillator is operated in the control mode. Figure 4 shows the low-pass filtered beat frequency of the mixed test and reference oscillator signals as the temperature of the crystal was swept through the range over which the system was trained. During this measurement, the temperatures of the other components were held fixed. The same quantity for the uncontrolled oscillator is shown as a reference; it reflects the temperature dependence of the quartz oscillator. Although the NNCXO was not fully trained when these data were taken, the oscillator's temperature dependence was significantly reduced when the neural controller was active.

Conclusion

We find that the results obtained to date are encouraging, and further development of neural networks for frequency standards is warranted. Future efforts will focus on the remaining technical issues: network designs and training protocols that will expedite training; cascading the neural controller with other control methods; and identification of specific frequency standard applications.

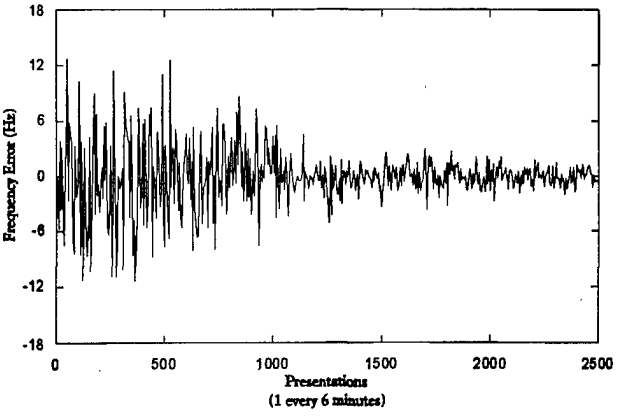


Figure 3. Fractional frequency error as a function of presentations.

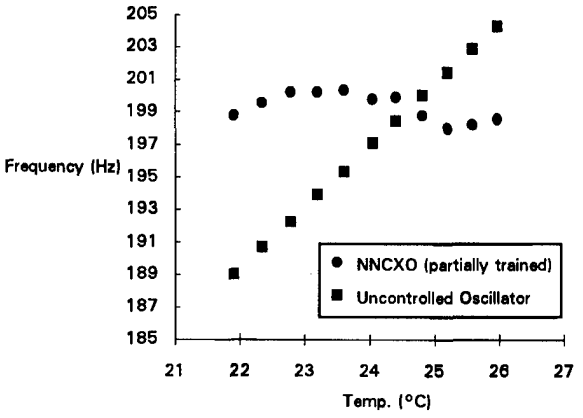


Figure 4. Comparison of the partially trained NNCXO with the uncontrolled oscillator as the temperature is swept from 21 to 27 °C.

The length of time required for training is a critical issue. The cost increase for training an oscillator for 30 days will be too large for many inexpensive commercial oscillators. There are neural designs that learn faster than the one used in this demonstration, and such designs could be employed in future applications. Another means of reducing the overall training time would be to first train the neural controller on a simplified numerical model of the physical system, then finish the training on the physical system to trim the model. This pre-training would require significantly less time, because the training cycles would occur at the speed which the computer processes information, rather than the temperature equilibration time of the physical system. However, the training may still require more time than is acceptable for inexpensive clocks. Therefore, the application of neural control methods is expected to be most appropriate for high priced standards or ensembles.

The optimium application of a neural controller may be when the neural network is used in conjunction with other less sophisticated controllers. In this configuration, the majority of the control is handled by an analog circuit or another type of controller that is based on a model of the physical system (MCXO, Kalman filters, etc.). The neural controller focuses on the smaller but not less significant features of the oscillator's response to environmental perturbations. These more subtle responses of the oscillator are where other controllers fail and the neural network is best employed. For example, the large temperature ranges an oscillator may encounter in an operational environment are often much larger than the designed tuning range of the oscillator. Furthermore, the bit resolution of most high quality, 16 bit, D/A and A/D interface boards divides voltage ranges into 64,000 discrete steps. For temperature compensation of an oscillator over a 60 °C range, bit resolution of a digital controller limits temperature resolution to about 1 mK. This would not be adequate for many frequency standard applications. A better solution is to temperature control the device to within 1 K, which is relatively simple, then have the neural controller handle the balance of the control needs. In this role, the neural controller can monitor the states of the other controllers and environmental parameters, and learn to compensate for additional frequency shifts that can be correlated with the measured events. These additional shifts are often subtle and their sources are not easily identified. Thus, we expect that the neural controller will be best suited as a supervisor of other simple controllers. The neural controller is the best candidate for learning the pertinent correlations and trimming a frequency standard based on the learned model.

References

[1]. Vanier, J., Audoin, C., The Quantum Physics of Atomic Frequency Standards, Bristol and Philadelphia, Adam Hilger, 1989.

[2]. Caudill, Maureen, and Butler, C. Naturally Intelligent Systems, The MIT Press, Cambridge, MA, 1990. *and* Caudill, Maureen, and Butler, C. Understanding Neural Networks Volume I: Basic Networks , The MIT Press, Cambridge, MA, 1992.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

The Functions of the Precision Frequency Measurement Technique in MCXO

Wei Zhou, Ming Wang

Dept. of Measurement and Instrumentation, Xidian University
Xi'an, 710071, P.R.China

Abstract

The MCXO is not only a combination of a crystal oscillator with a microcomputer, but also based on some frequency measurement techniques and an automatic MCXO test system. A highly precision frequency measurement technique with a simple structure and a fast response time can be used in MCXO and its test system. It can be used to enhance the accuracy and stability of MCXO, simplify the structure of MCXO. This paper describes several very useful high-precision, fast response time frequency measurement techniques. These techniques are not only highly precision, of fast response time in a very wide frequency range, but also simple and all digital in circuit. By means of the techniques, with short gate times the measuring error is less than 10^{-9}. In many MCXO's and relative instruments, the highly preciaion measuring technique and circuits can be used in place of complicated frequency transformation and control circuits, and the devices can thus be simplified, with their frequency stability enhanced and other advantages obtained. This paper also describes several different MCXO crystal oscillators, and proves the outstanding contributions of the highly precision measurement technique in the MCXO circuits for their performance.

1. Introduction

In MCXO's, especially in those using SC-cut crystal, frequency measurement technique and circuits are used. They are used or to measure temperature, or in frequency control circuits. In MCXO test system, frequency measurement equipment must be used. With precision frequency measurement technique the frequency stability of MCXO and its response time can be improved, and its structure can be simplified.

In some frequency and time measurement instruments, MCXO can be combined with the measurement circuits of the instruments, and the control and compensation circuits can be omitted and the accuracy of the instruments can be enhanced. The accuracy of the instruments depends on the accuracy of the measurement circuits.

2. Several high precision and fast response time measurement techniques in MCXO's

Some precision, fast response time, wide frequency range and simple structure frequency measurement techniques have been developed. They are very useful in MCXO's and their test system. References [1], [2] and [3] describe frequency measurement techniques based on phase coincidence detection method. By means of them many frequency measurement instruments have been developed. They are not only digital, but also highly accurate in a very wide frequency range (from several KHz to several hundred MHz). For example with 0.01 s to 0.1 s gate time 10^{-9} measuring accuracy can be obtained. Figure 1 is a typical diagram of the instrument.

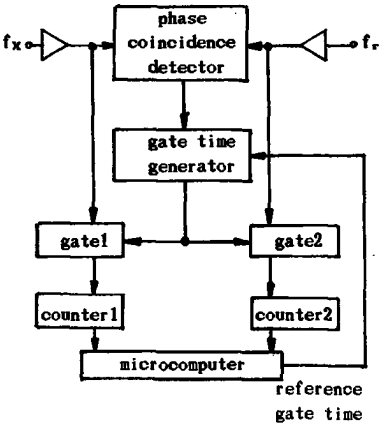


Figure 1

In the instrument all circuits are digital and integrated. When they are used in MCXO's the microcomputer can also have other funtions.

Since in MCXO, the range of frequency measurement is not very wide, the measuring devices can be more simple. A new frequency measurement method using a common oscillator is remarkable. It is similar to the method in reference [3], except without two phase coincidence detectors. With this method, $10^{-11}/\tau$ measurement accuracy can be obtained when the difference of frequencies of two signals or their

0-7803-1945-1/94 $3.00 © 1994 IEEE

multiples or fractions is not too large. Figure 2 is its principle diagram.

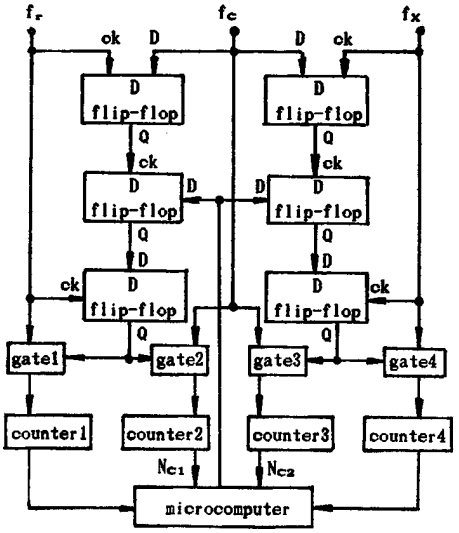


Figure 2

In the measuring device, two gate time are synchronized with the two compared signals respectively. Therefore, there are not their count errors in the measurement result. The two digital mixers composed of D flip-flops can be used to detect the periodic closest phase difference cases between two frequency signals to assist to form the gate time. Therefore, with the simple circuit the measuring accuracy can be enchanced obviously. In certain frequency range, the count errors of the common oscillator which has suitable frequency in the two gate time can increase or decrease synchronously. In the equation [3]

$$f_x = f_r N_x N_{c1} / N_r N_{c2} \qquad (1)$$

the count errors of N_{c1} and N_{c2} can be eliminated with each other. For SC-cut dual mode crystal oscillator, the difference frequency between C mode and B mode signals is about 9 - 10% of C mode frequency, the measuring technique is suitable. Because its high accuracy, simple structure, fast response time and other advantages, it can have wide application in MCXO.

3. A design combined with measuring technique, MCXO and instrument

In reference [4] the method combined with MCXO and measuring technique of the instrument was given. By means of high accuracy measurement technique better result can be obtained. It also is the base of some MCXO' s.

If an accomplished MCXO is used as the reference signal of time and frequency measurement instruments, its compensation for temperature variation has an error. It must have an influence on the accuracy of the instruments. Therefore, if the error of the crystal oscillator produced by temperature can be known accurately and considered in calculation of measuring result, higher measuring accuracy can be obtained. Using this method, the crystal oscillator has better short-term frequency stability.

The principle diagram of an instrument is shown in Figure 3. For simpler circuit, the frequencies of measured signal and B mode signal are measured alternatively with same count circuit.

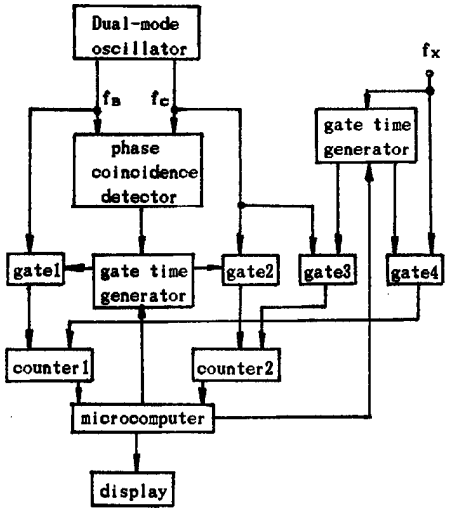


Fig. 3

When the B mode signal f_B is measured with its C mode signal f_C as a reference signal, the measurement method shown in Figure 1 is used. The gate time is less than 0.1 second, fast response time and high measuring accuracy can be obtained. The synchronization measurement method of multiple periods is used in measuring the measured signal f_x. The f_B frequency is equal to measured B mode average frequency before and after measuring f_x. According to the temperature duplication of the SC-cut crystal oscillator, the increment of the temperature compensation can be less than 0.1℃.

The instrument has two measurement procedures. One is that C mode signal f_C of the dual mode crystal oscillator is used as the reference signal to measure the the frequency f_B of its B mode signal. Using the measured values, from PROM, is obtained a correction factor Δf_C which is only dependent on the temperature error of f_C. Through a calibration run the correction factor has been gotten and stored in the PROM. Because of the uncertainty of f_C frequency, the obtained f_B frequency is not exactly equal to the B mode frequency, but is relative to the frequency difference of them. But the accuracy of Δf_C is not influenced. Another measurement procedure is that the signal f_C which has a known error Δf_C is used as the reference signal to measure the measured signal f_x.

When the frequency of the measured signal is

calculated, in the equation the accurate frequency $f_C=f_O+\triangle f_C$ is used as the reference frequency.

$$f_X=f_C N_X/N_C = (f_O+\triangle f_C)N_X/N_C \qquad (2)$$

where the gate time is synchronized by the measured signal f_X, N_X is the cycle number count of f_X in the gate time, and N_C is the cycle number count of f_C in the gate time.

The flow diagram of the instrument is shown in Figure 4.

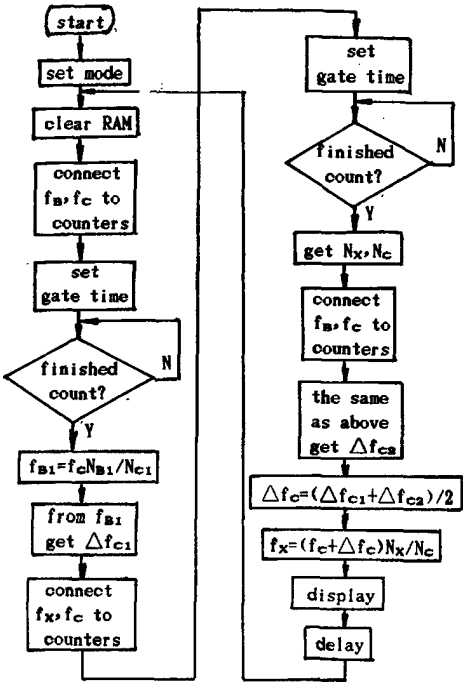


Fig. 4

The error of the instrument depend mainly on the temperature duplication of the crystal oscillator, the storage volume of PROM and the accuracy of the test system used in the calibration run. When f_B is measured the error produced by the measurement accuracy is

$$A=A_1 S_C/S_B \qquad (3)$$

where A_1 is the measurement error when C mode signal is used to measure B mode signal; S_C is the frequency-temperature coefficient of C mode signal, because of its nonlinearity S_C is the maximum of the coefficient; and S_B is the frequency-temperature coefficient of B mode signal. Therefore, when the fast response time is demanded the high accuracy measurement technique must be used then the MCXO accuracy can be obtained.

4. MCXO's using highly accurate measurement technique

We have used the highly accurate measurement technique and circuits in the design of MCXO's. Because in short time higher measurement accuracy can be obtained, the MCXO's can have a fast compensation response to a temperature variation, and their structures are simple.

Figure 5 shows a MCXO diagram combining a temprature sensor oscillator with a VCXO.

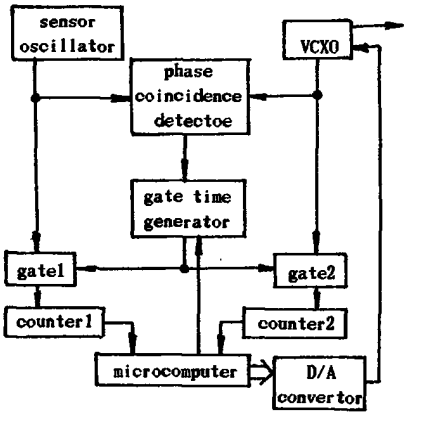


Fig. 5

The temperature sensor oscillator is of Y+5° cut, and VCXO is of AT cut. They are in well thermal coupling. The operation procedure is similar to that of SC- cut MCXO. The frequency - temperature property of the temperature sensor oscillator is the same as that of B mode signal of a SC-cut dual mode oscillator, it has a large and linear temperature coefficient (about -95ppm/℃). VCXO is used as a reference frequency signal to measure the frequency of temperature sensor f_B. According to the measurement result the control voltage of the VCXO is changed and the accurate frequency is obtained. In the PROM, according to frequency f_B of different temperature the digital signals fed into D/A converter have been stored. The storage process is done in the temperature test of the crystal. Because in the measurement the frequency of f_C is uncertain, the process from measurement to control of VCXO is a feedback process. When temperature changes, f_C and f_B frequencies change. Frequency of f_C deviates from its nominal frequency, there is certain deviation in the measured f_B frequency. Since the temperature coefficient of f_B is much larger than that of f_C, if the demand for the accuracy of MCXO is not very high, this deviation can be omitted. From -30℃ to +70℃, 5×10^{-7} frequency stability can be obtained.

Because the temperature coefficient of f_B is very large, the frequency difference between it and f_C changes along with temperature monotonously. The frequency - temperature compensation for MCXO is of negative feedback. When the measurement circuit in MCXO is not only of high precision but also of fast response time, then the temperature compensation can respond to temperature change closely, especially in temperature shock case. For any temperature change, several measuring - control periods are re-

quired to enable MCXO stable at a new balance case. Therefore, the gate time of measurement circuit in Figure 5 ought to be from 0.01 second to 0.1 second (to different frequency relationships, the gate times are different) . In the gate time the frequency measurement accuracy of the measuring circuit can be 10^{-7}. The relative minimum temperature increment is

$$\Delta t = A_1 / S_1 \qquad (4)$$

where A_1 is the frequency measuring accuracy when measuring the sensor oscillator; S_1 is the slope of frequency - temperature change of the sensor oscillator. To a simple MCXO, using a 8 digits D/A converter, the temperature increment is about 0.5 ℃. $2 \sim 5 \times 10^{-7}$ frequency stability can be obtained.

Figure 6 is a MCXO composed of a high accuracy, simple frequency measurement device combined with a SC cut dual mode crystal oscillator.

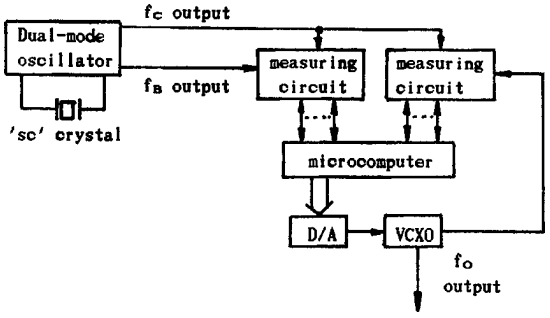


Fig. 6

In the MCXO, the measuring circuit and a VCXO are used instead of the DDS synthesizer of reference [6]. The C mode signal of the dual mode oscillator is used as reference signal to measure the frequency of its B mode signal, and according to its frequency the modification data of the C mode signal can be obtained. Using the C mode signal with known accurate frequency the simple VCXO frequency is measured accurately. Comparing the measured result with the demanded VCXO frequency by software, with their difference the signal fed to D/A converter is modified and the frequency of VCXO is regulated to demanded frequency. The crystal of VCXO is a simple AT cut quartz crystal. The 2×10^{-8} frequency stability has been obtained.

In Figure 6, when the frequency of B mode signal is measured, the response time ought to be fast and the measuring accuracy can not be too high. When the frequency of VCXO is measured, the demand for measuring accuracy and the response time must be high.

5. Conclusion

MCXO is a kind of very useful crystal oscillator. A lot of factors can dicide its performance. Some highly accurate, simple frequency measurement techniques can show an excellent effect in enchancing MCXO accuracy, simplifying its structure. The techniques can be used in MCXO and its test system. The uses of these techniques are very important for higher accuracy of MCXO. Now in a very wide temperature range, with the highly precision frequency measuement technique, the MCXO's can have 10^{-8} frequency stability. The accuracy limitation depends mainly on the frequency - temperature duplication of the crystal in a very wide temperature range. By means of combining frequency measurement technique with improvement of the crystals, 10^{-9} or higher accuracy MCXO can be accomplished.

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, No4 (1993), pp. 358-362.

[3] Wei Zhou and Zongqiang Xuan "A Simple Precision Frequency Standard Comparator", Proceedings of the 1993 IEEE International Frequency Control Symposium, pp. 335-338, 1993

[4] Jone A. Kusters, et. al., "Dual Mode Operation of Temperature and Stress Compensated Crystals", Proceedings of the 32nd Annual Frequency Control Symposium, pp. 389-397, 1978

[5] Wei Zhou, "A New Technique of Frequency Synthesis", Proceedings of the 1993 International Frequency Control Symposium, pp. 251-254, 1993

[6] John R. Vig, "Quartz Crystal Resonators and Oscillators for Frequency Control and Timing Applications", A Tutorial, January 1993

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

APPLICATION OF LFE SC-CUT RESONATORS AT HIGH STABILITY OSCILLATORS

A.Masiukiewicz,B.Gniewińska,W.Szulc

Piezoelectric Department
Tele and Radio Research Institute, Ratuszowa 11, 03-450 Warsaw

ABSTRACT

The LFE SC-cut resonators have some quite well known advantages . The main are: lack of electrodes in the active part of the crystal plate and very efficient suppresion of B mode. Because of this application of LFE resonators can improve both long and short term frequency stability , critical for high stability oscillators parameters. A few models of different construction single side resonators were made and their parameters were measured. These resonators are very difficult to adjust (because of very low C1) so the special procedure of calibration using chemical treatment was developed. The especially optimized oscillation circuit was applied /ref.11/. Circuits and oven parameters were measured and discussed.

1.INTRODUCTION

Lateral field excitation resonators have a few advantages compared with typical high stability resonators /ref. 1,2,3,4,5,6 /. The most important features are as follow:
- better long term stability due to removing of the electrodes from the active part of the blanks,
- higher quality factor of unloaded resonator as a result of surface free from the metal layer,
- more efficient suppresion of unwanted modes (e.g. B mode of SC-cut resonators).

But on the other side lateral excitation change basic parameters of the resonators. Very high as for this frequency range motional resistance of the resonator, complicates the construction of the oscillation circuit which should assure high short-term frequency stability and small motional capacitance considerably reduces the frequency adjustment range.

Design of the oscillation circuit for the 5 MHz SC-cut lateral field excitation resonator, which guarantees the electronic circuit influence on short-term frequency stability lower then E-13 for averaging time $\tau \geq$ 1sec is presented in this paper. The construction and parameters of three chamber oven of very small temperature fluctuations at the level of E-7 are also presented.

2.PARAMETERS OF LFE SC-CUT RESONATORS

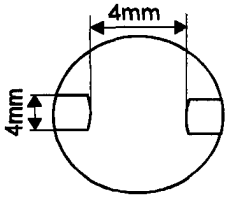
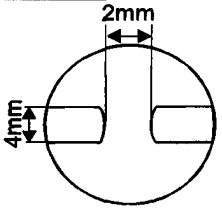
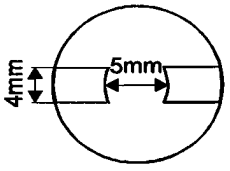
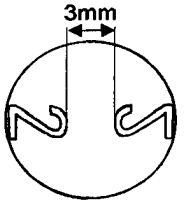
A few types of 5 MHz SC-cut resonators operating at 3rd overtone mode have been realized. Parameters achieved for three types of electrodes configuration are presented in Table 1./ref. 7/

Beside basic parameters, dynamic temperature behaviour and nonlinearity of resonators were measured. Typical measured, dynamic temperature coefficient of the LFE resonator is within the range of ±2E-7 (the measurement was carried out according to the method presented in reference 8, basing on estimation of resonator response to step temperature change).

The characteristic of nonlinear behaviour of the resonator is presented in figure 1. According to expectation this characteristic is very similar to the characteristic obtained for typical SC-cut resonators.

0-7803-1945-1/94 $3.00 © 1994 IEEE

Table 1 LFE resonator parameters

Elektrodes configuration	Motional resistance [Ω]	Motional capacistance C1 [E - 18]	Uniloaded resonator Q factor [E6]	B mode suppresion [dB]
	1100 - 1300	average 9	2.4 -2.6	> 30
	400 - 650	average 18	2.8 - 3	> 40
	1500 - 1800	average 8	2.2 - 2.6	-
	600 - 800	average 15	2 - 2.5	-

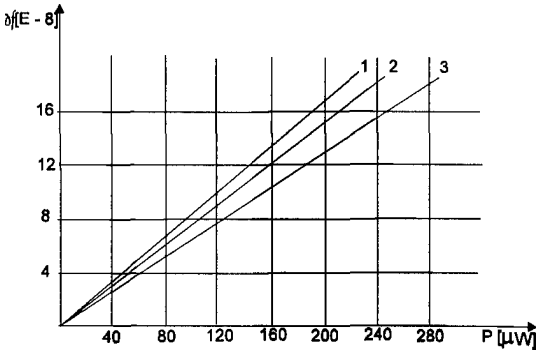


Fig.1. Frequency versus drive power characteristic

1 SC-cut resonator R1=70Ω (nr 258)
2 LFE SC-cut resonator R1=500Ω (nr 15)
3 LFE SC-cut resonator R1=600Ω (nr 13)

3.OSCILLATION CIRCUIT

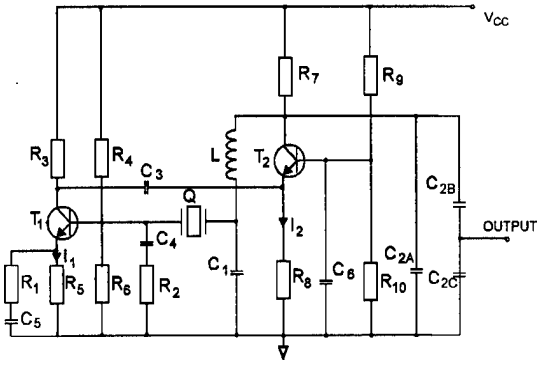
Considering resonator parameters, application of following oscillation circuits were taken into account:
- bridge circuit with differential amplifier,
- Butler with the resonator connected through the transformer to the emitter circuit,
- Butler with the resonator in the base circuit.

Essential features of studied circuits are presented in Table 2.

Table 2. Features of oscillation circuits

Type of Circuit	Features
Bridge	No reactance elements in the feedback circuit. Improved efficiency of phase noise reduction, because of bridge circuit influence. Very effective mechanism of drive level limitation.
Butler with the resonator in the emitter	Neccesity of matched impedance circuit application. Phase noise reduction due to negative feedback. For achieving proper drive level limitation, the additional circuit is required.
Butler with the resonator in the base	The resonator and the input impedance of the amplifier can be matched without any transformation circuits. Negative feedback for parametric noise reduction can be applied. For achieving proper drive level limitation, the additional circuit is required.

Considering the phase noise measurement results in the open loop (which are similar for these three types of circuits) and simplicity, the Butler configuration with the resonator in the base was chosen for further experiments. The diagram of this circuit is shown in figure 2.



I1≈ 1mA
I2≈ 0.2mA
R1 negative feedback
R2= 400÷1200Ω

Fig. 2. Diagram of the oscillation circuit

Two stage amplifier with second stage built around T2 transistor (operating as limiting circuit) and negative feedback in the emitter circuit of T1 transistor, guarantee very good phase noise characteristic in the open loop for a broad range of equivalent resonator resistance from 400 to 1200 Ω.

Basing on methods describe in ref. 9 the open loop measurements were made and coefficients A and B were calculated. The A coefficient which characterize $\frac{1}{f}$ noise is about 3.7E-14 and the B coefficient defining f^0 noise is 3.5E-16. These values were measured under conditions, which guarantee drive level lower than 5 μW and efficiency of the amplitude limiting system as shown in figure 3.

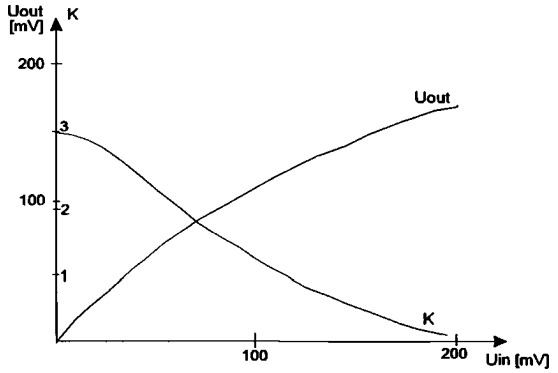


Fig.3. Amplitude limiting efficiency (for R1=30Ω)

Basing on following equation /ref.9/:

$$S_\phi(f) = (\frac{A}{f} + B)[1 + (\frac{f_0}{2Qf})^2]$$

the spectral density of phase noise fluctuations, for resonator of Q factor close to 3E6 was calculated. The results are presented in figure 4.

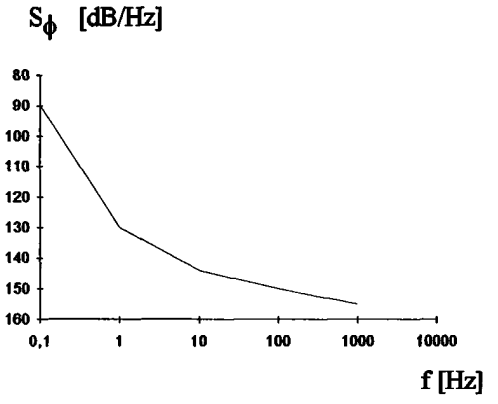


Fig.4. Spectral density of phase fluctuations

4. OVEN CONSTRUCTION

Basic principles of the oven construction are presented in figure 5. The oven consists of three chambers of stable temperature, where the resonator and electronic circuits of higher influence on frequency are placed.

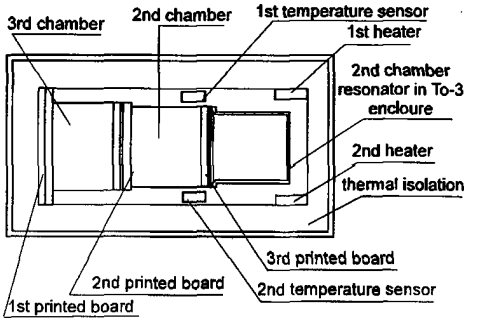


Fig.5. Oven construction

The results of the oven temperature stability are shown in Table 3.

Table 3. Oven temperature stability

Temperature range		-10÷50 C
Quality factor :	I chamber	>2000
	II chamber	>500
	III chamber	>100
Short-term temperature stability of I chamber at 25 C		
τ=1s		5E-7
τ=10s		4E-7
τ=100s		1E-6

Temperature stability n the first chamber was measured by the means of B mode of the Sc-cut resonator. It was thus assured that the temperature sensor was placed exactly at the same place as the LFE resonator.

To obtain the high resolution of measurements, the high stability oscillator of the same frequency as B mode was built. Due to such solution, typical equipment for short term frequency stability measurement could be used. Measurements in second and third chambers were made using precise thermistors.

Analysis of oven temperature stability depending on power loss in the resonator itself was made according to reference 10. The change of loss power from 5 to 40 μW decrease two times the temperature fluctuations. The results for loss power equal 5 μW are sufficient for obtaining short term frequency stability (caused by temperature fluctuations) better than E-13 for τ=1s and E-14 for longer averaging time.

5.CALIBRATION PROCEDURE

The very small value of C1 as a result of poor coupling, makes the lateral field excitated resonators difficult for calibration. The influence of electrodes (mass) on frequency was reduced so significantly that calibration by typical procedure and equipment was impossible. The only way of achieving proper frequency is by changing the thickness of the plate.
A speciall chemical procedure was developed. With this method the final calibration of 100Hz was obtained.

6.OSCILLATOR CONSTRUCTION

The block diagram of the oscillator is shown at figure 6. The oscillator consists of following circuits:
- oscillation circuit (Butler),
- tunning circuit,
- LFE resonator,
- voltage regulator 2 (low noise LT1021),
- resonance output amplifier,
- temperature control circuit (proportional),
- voltage regulator 1.

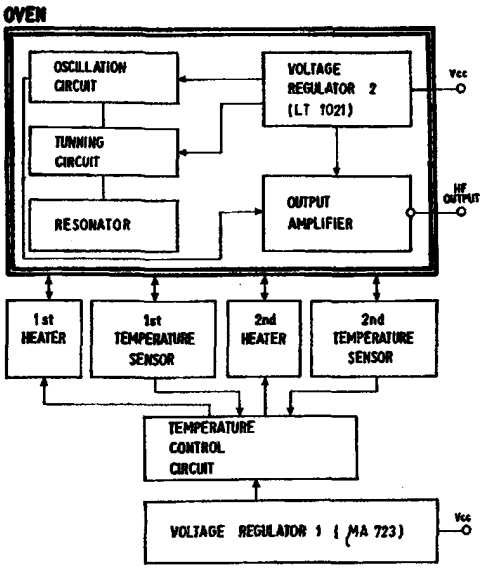


Fig. 6. Block diagram of the oscillator.

All high frequency circuit were placed in the three chamber oven what is shown at figure 7.

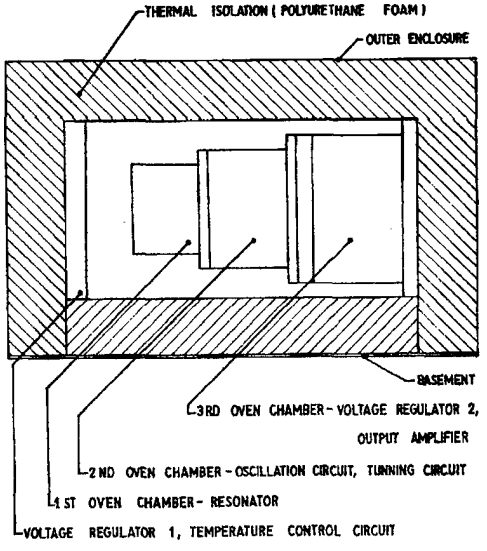


Fig. 7. Circuits location in the oven.

7.MEASUREMENTS RESULTS

The frequency versus temperature characteristics of three oscillators were measured. The frequency change in the temperature range $0 \div 50\,^{\circ}C$ was below E-9 and the slope was better than 2E-11/K.

8.CONCLUSION

Applying the proposed oven construction and carrying out further optimization of the LFE resonators operating conditions it is possible to obtain short-term frequency stability better than E-13 for averaging time longer than 1s.

9. REFERENCES

/1./ A.Ballato, E.R.Hatch: Lateral Field Excitation of Quartz Plates. 34th AFCS, 1983

/2./ R.C.Smythe: An Approximate Expression for The Motional Capacitance of A Lateral Field Resonator. 41st AFCS, 1987

/3./ A.W.Warner: Lateral Field Resonators. 39th AFCS, 1985

/4./ A.W.Warner: Measurement of Plano Convex SC Quartz Blanks Using Lateral Field Excitation. 42nd AFCS, 1985

/5./ M.M.Driscoll: Fabrication and Methods for Evaluation And Circuit Utilization of Prototype Lateral Field Resonator. 41st AFCS, 1987

/6./ B.Capelle: Synchrotron Radiation X Ray Topography Study of Lateral Field Resonators. 6th EFTF, 1992

/7./ W.Szulc, K.Weiss: High Stability Oscillator Using LFE Resonator. Part I. Model of LFE Resonator. Contract B3-2, 1993

/8./ B.Gniewińska:Effets Thermigues Dans les Oscillateurs a Quartz. Seminaire "Etalons de Frequence". Besancon, 1981

/9./ D.B.Leeson: A Simple Model of Feedback Oscillator Noise Spectrum. Proceedings of the IEEE, 1966

/10./ J.P.Valentin: Analyse de Quelqes Aspects du Comportement Thermique des Resonateures a Quartz a Ondes de Volume. These a L'Universite de Franch Comte, 1983

/11./ B.Gniewińska, A.Masiukiewicz: Optimization of LFE SC-cut Resonators Oscillation Circuits. 8th EFTF , 1994

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SPECTRAL PROPERTIES OF DDFS: COMPUTER SIMULATIONS AND EXPERIMENTAL VERIFICATIONS

Věnceslav F. Kroupa
The Institute of Radio Engineering and Electronics
Academy of Sciences of the Czech Republic
182 51 Praha, Czech Republic

Abstract: In the present paper we shall discuss generation of spurious signals in direct digital frequency synthesizers (**DDFS**) with different output waves with the assistance of computer simulations.

First, we shall deal with simple square wave and triangular wave outputs.

However, our main attention will be paid to the systems generating sine signals since they exhibit the lowest level below one cannot get.

The major sources of spurious signals are phase truncation, the limited lenght in sine look-up tables, the finite resolution of digital to analog convertors **DAC**, and intermodulation problems. The latter are not easy to appreciate.

I. Introduction.

The task of **DDFS** is to generate a desired output frequency $\mathbf{f_x}$ from the input standard or clock frequency $\mathbf{f_i}$, generally, with the assistance of **Modulo-N** ari-thmetics. The governing equation of this system is

$$f_x = \frac{k}{N} * f_i \qquad (1)$$

where the denominator **N** is generally a power of two, i.e.

$$N = 2^R \qquad (2)$$

For the examination of the **DDFS** properties we must start from the normalized frequency

$$\xi_x = \frac{f_x}{f_i} = \frac{k}{2^R} = \frac{X}{Y} \qquad (3)$$

where **R** is number of bits stored in the accumulator and **X** and **Y** are mutually prime integers, as a consequence **X** is always odd and **Y** a power of two as high as **R** (e.g. R=24, 32, etc.) or often even lower, some $\mathbf{R_{eff}}$, in instances where **k** is even or has a factor formed by power of 2. We shall see later that **k** should be always an odd number.

In recent years we have studied behaviour of **DDFS** in respect to spectral purity [1 to 5]. First we have paid attention to the simple system with the rectangular output waves. Then we have investigated devices providing triangular output waves. However, most of our work was dedicated to sinusoidal frequency synthesizers.

Since the published data about spectral purity are often incomplete on one hand and contain to much spurious spectral lines, the origin of which is difficult to trace on the other hand, we have started with computer simulations.

II.Square wave outputs.

Pulse and rectangular output waves are plagued with very large spurious signals [1,2,3] and are used only in some instances where we need to synthesize an exact value of the output frequency, e.g. the **PAL** colour subcarrier in [1]. However, there are also some special communication applications which require square wave signals [6,7,8].

A/ Quasiperiodic omission of pulses.

In order to enhance the fundamental harmonic in output signal the rectangular output wave should have the space-mark ratio as close to 1:1 as possible. This requirement can be met by different means, the simplest of which is to put a flip-flop or binary divider at the end of the synthesizing chain [1,2].

For one single output frequency the author suggested as mathematical models:

a/ continued fraction expansion
b/ modified Engel series expansion
c/ combinations of both in some instances
d/ gear box approach provides output frequencies with very small errors [9].

In a more general case we can use
e/ Modulo-N arithmetics.
The advantage is the possibility of application of commercial **IC** chips and to refer simply to the **lows** and **highs** of the **MSB** [7,8], the difficulty is a slow convergence of the binary systematic fractions in respect to the desired normalized frequency ξ_x.

0-7803-1945-1/94 $3.00 © 1994 IEEE

B/ Spurious signals.

In the useful range of **DDFS**, extending ideally from zero to the Nyquist frequency $f_c/2$ we encounter a lot of discrete spurious signals which can be divided in several groups.

1/ Harmonics of the output signal.

Particularly in instances where the normalized frequency from ξ_x is small compared with the Nyquist "frequency" $\mathbf{Y/2}$ we must be careful with rectangular wave harmonics which are ideally only odd ones, however,large since they decay inversely with their order - see Fig. 1.

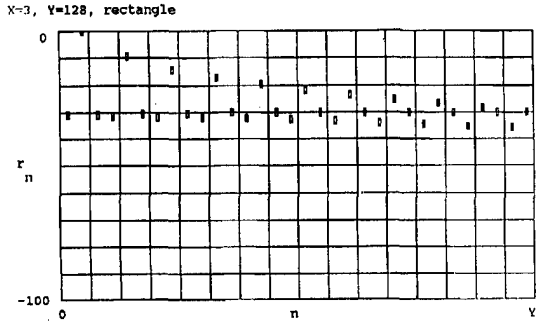


Fig. 1 Spurious signals in the rectangular wave with the normalized frequency ξ_x = 3/128.

2/ Harmonics of the device fundamental frequency.

Most of the spurious signals in **DDFS** are harmonics of the device fundamental frequency

$$f_{fund} = \frac{1}{YT_c} \tag{4}$$

However, it is generally easy to estimate power level and the frequency of the larger ones. The problem was discussed in the depth by the author [1,2]. Here we shall recall the results. The spurious phase-time modulation is give by

$$s(t_r) = T_c[r\frac{Y}{2X} - integer(r\frac{Y}{2X})] = T_c[r(A_o + \frac{a_1}{B_1} - \frac{a_1a_2}{B_1B_2} + \ldots - integer(r\frac{Y}{2X})] \tag{5}$$

At the 1st order approximation of $\mathbf{Y/2X}$ by $\mathbf{A_1/B_1}$, **r** in eq.(5) changes from 1 to B_1 ($r=1,2,..B_1$) and we arrive at a sawtooth wave with the period $\mathbf{T_cA_1}$ and the amplitude **1.** The 2nd order approximation reveals an additional saw-tooth wave with amplitude $\mathbf{1/B_1}$ ($r=1,2, \ldots B_2$) and the new modulation period $\mathbf{T_cA_2}$ which is also imposed on the first sawtooth wave, etc.

An example of such a modulation function is shown in Fig.2.

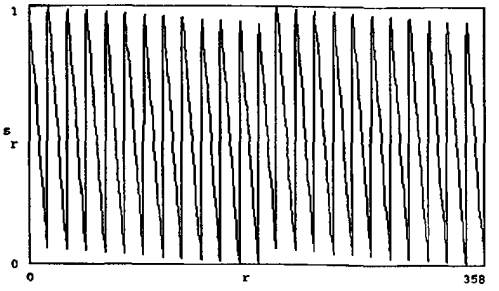


Fig. 2 Spurious phase-time modulation $s(t_r)$ for the normalized frequency ξ_x= 179/10000.

To illustrate the problem we shall investigate a sine wave phase modulated by $\omega_x s(t_r)$ i.e.

$$z(t) = \sin[\omega_x t + \omega_x s(t_r)] \tag{6}$$

which simplifies for $\mathbf{s(t_r)}$ approximated with the largest sawtooth wave only to

$$z(t) = \sin[\omega_x t + 2\pi\frac{X}{YT_i}T_i\sum_{m=1,2,\ldots}\frac{(-1)}{m}\sin(m\nu t)] \tag{7}$$

where the fundamental modulation frequency is

$$\nu = \frac{2\pi}{A_1T_c} \tag{8}$$

and amplitudes of individual saw tooth wave harmonics are

$$a_m = \frac{2X}{mY} \tag{9}$$

With the assistance of the phase or frequency modulated signals we get for the level of the first spurious components

$$a_{1,1} \approx \frac{J_1(a_1)[J_o(a_2 \pm J_1(a_2)}{J_o(a_1)J_o(a_2)} \tag{10}$$

where the minus sign is for the lower and the plus sign for the upper side bands.

For Bessel functions with small indexes the above equation simplifies to

$$a_{1,m} \approx J_1(a_m) \approx \frac{a_m}{2} \approx \frac{X}{Y} \tag{11}$$

For the amplitudes of the next pair of spurious signals we find

$$a_{2,m} \approx \frac{a_{1,m}}{4} \tag{12}$$

etc.

After introducing relation **(9)** into the eq.**(11)** we get for the spurious signal levels "near" the carrier (in

dB measure) for the **1st** order approximation, cf. **(5)**

$$a_{1,m_1} \approx 20\log(\frac{X}{m_1 Y}); \quad (m_1=1,2,...) \qquad (13)$$

with the modulation frequency

$$f_{m_1} \approx \frac{f_c}{A_1} \qquad (14)$$

Relation **(12)** is often cited in the literature, however, without warning that it gives useful results for very few $\mathbf{m}_1$'s.

For the 2nd order approximations we have

$$a_{2,m_2} \approx 20\log(\frac{X}{m_2 B_1 Y}); \quad (m_2=1,2,...) \qquad (15)$$

with the modulation frequency

$$f_{m_2} \approx \frac{f_c}{A_2} \qquad (16)$$

We can proceed in the same way till we arrive to the spurious signals closest to the carrier, and with the smallest modulation frequency

$$f_{m_{n-1}} = \frac{f_c}{A_n} = \frac{f_c}{Y} \qquad (17)$$

and amplitudes

$$a_{n-1,m_{n-1}} \approx 20\log(\frac{X}{Y}\frac{1}{m_{n-1}B_{n-1}}) \qquad (18)$$

Note, that there are some difficulties with the above approximations: relations for the level of the spurious signals are valid only for small values of $\mathbf{m}_1$, $\mathbf{m}_2$, etc. Further, that they supply the better results the smaller is the ratio **X/Y**. However, in these instances we must be careful to remove harmonics of the output frequency $\mathbf{f}_x$ from the useful pass band (cf. Fig. 1) Another problem with the precision of the estimated amplitudes may be caused by alising.

Example 1.

Estimate the major spurious signals in the square wave output of the normalized frequency

$$\xi_x = 179/10000$$

With the assistance of the modified continued fraction expansion of **Y/2X** we get

k	b_k	a_k	A_k	B_k
0	28	1	28	1
1	15	-1	419	15
2	12	-1	5000	179

Tab. 1

From which we find for the spurious phase-time modulation

$$\frac{s(t_r)}{T_c} = [r(28-\frac{1}{15}-\frac{1}{15+179})] \qquad (19)$$

Evidently it is formed by two superimposed sawtooth waves as shown in Fig. 2.

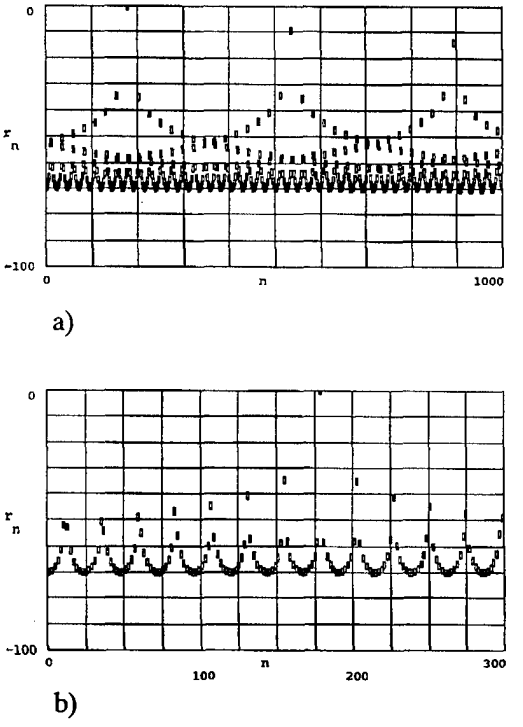


Fig 3. Spurious signals in the rectangular wave with the normalized frequency ξ_x = 179/10000: a) till n =1000; b) reveal more details.

The first saw tooth wave has the amplitude **≈ 1** and the modulation "frequency" **10000/28 = 23.87.** Estimation of the first order spurious signals according to **(13)** reveals

-34.94 dB

Exact computation gives

$\mathbf{r}_{155}$ **= -34.81** and $\mathbf{r}_{203}$ **= -35.12 dB**

For the 2nd order approximation we find for the first spurious side bands the modulation "frequency"

10000/5000 = 2

and for their level in accordance with **(15)**

-58.47 dB

Whereas the exact computation gives

$\mathbf{r}_{177}$ **= -58.35** and $\mathbf{r}_{181}$ **= -58.38 dB**

For actually computed spectra see Fig. 3.

Example 2.

In Fig. 4, reproduced from [1], we compare the simulated and actually measured spectra of the **DDFS** designed for the PAL colour subcarrier.

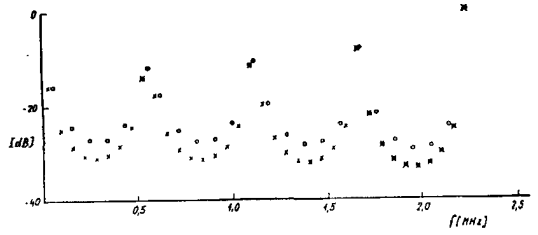
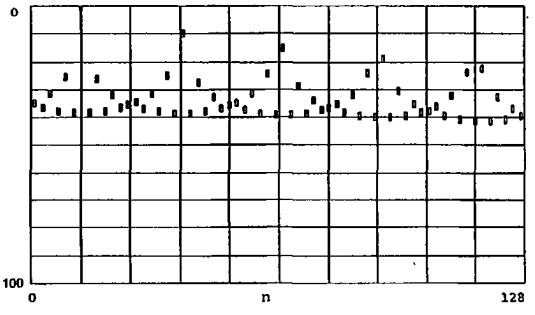


Fig 4. DDFS for the PAL colour subcarrier; 1st (.) and 2nd (o) order approximation and theoretical (x) and experimental (.) results [1].

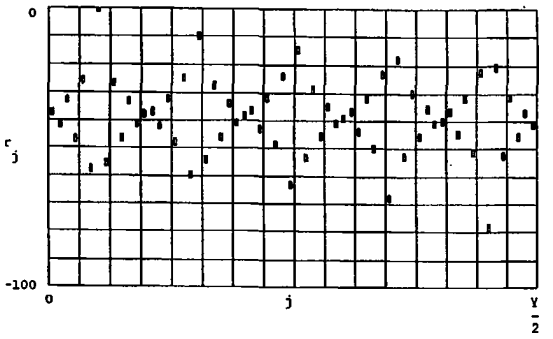
C/ Limitation of the sine waves.

Another way to generate the rectangular output wave is by limitation of the output sine wave.Giebel Lutz and O'Leary [6] suggested "to convert to a square wave, after lowpass filtering, the 6 bit sine **DAC** output". In addition they state that "the synthesized output clock becomes asynchonous with respect to the reference clock". However, the computer simulation did not proved the latter statement.

As to the spectral properties the larger spectral lines are the same as those found by investigation of square waves generated by quasiperiodic omission of pulses whereas the background level is much more dispersed - see Fig. 5.



a)

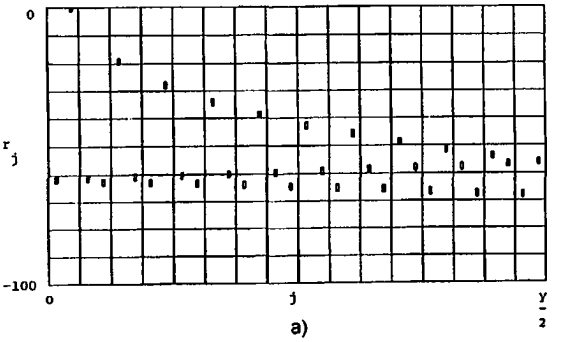


b)

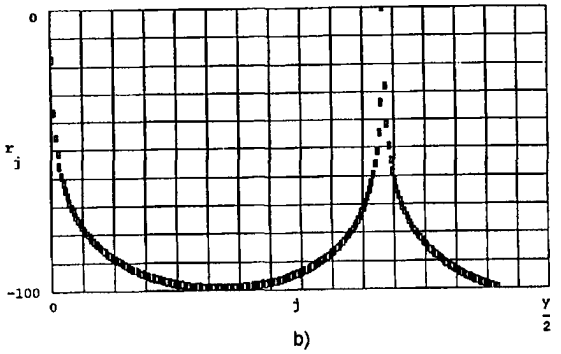
Fig. 5 Spectral properties of the square wave output for for ξ_x = 13/256;
a) "quasiperiodic" pulse omission
b) sine wave limitation.

III. Triangular wave outputs.

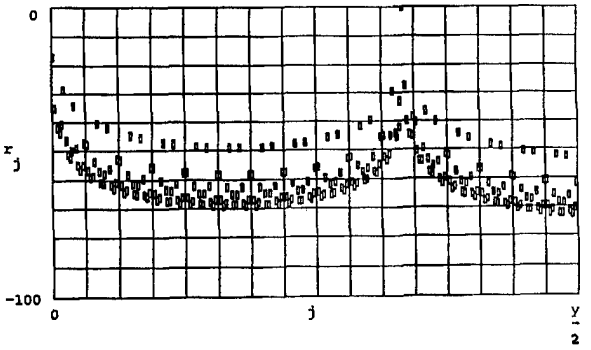
Another, rather simple, **DDFS** can be built with the triangular output wave. The advantage is that no sine look-up table is needed since the **DAC** is fed directly from the phase accumulator. Two **MSB** are used for changing the saw-tooth output from the accumulator into the triangular output.



a)



b)



c)

Fig. 6 Spectral properties of the triangular wave output signal for the normalized frequencies:
a) ξ_x = 3/128 and DAC A = 10;
b) ξ_x = 341/1024 with A = 10;
c) ξ_x = 341/1024 with A = 5.

Another advantage, in contradistinction to the rectangular output wave, the spectral properties are better since harmonic signals (ideally only odd ones) of the f_x decay withe square of their order. Some examples are shown in Fig. 6.

IV. Sine wave outputs.

The sine wave output is the most desired form of the output signal since it exhibits the lowest level of inevitable spurious signals.

There are two major spurious signal generators: the phase truncation and the amplitude truncation.

Their level and locations have been investigated in the past, particularly, by the author in [2-5].

A/ Phase truncation.

In accordance with the simplified Fig. 7 we use from the **R** phase bits stored in the accumulator only **W** bits for steering the sine look up table.

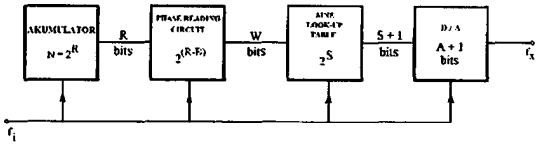


Fig. 7 the simplified block diagram of sources of spurious signals in DDFS due to the loss of information in different block diagrams.

The problem was first discussed by Mehrgardt [10] and later solved by Nicholas and Samueli [11]. However, it was the author [2,3] who presented a very simple closed form relation for the level of the spurious signals of this origin. We shall repeat here the results in short. In instances where we neglect ***B*** phase bits, i.e.

$$B = R - W \tag{20}$$

phase readings are still performed at each $\mathbf{mT_c}$ (**m** = **0,1,...,Y-1**) . As a consequence amplitudes of individual samples (rectangles) in the output sine wave change into

$$\sin[2\pi\frac{2^B}{Y}integer(\frac{mX}{2^B})] = \sin\{2\pi\frac{2^B}{Y}[\frac{mX}{2^B} - s(m)]\} \tag{21}$$

where

$$s(m) = \frac{mX}{2^B} - integer(\frac{mX}{2^B}) \le 1 \tag{22}$$

The above equation resembles the earlier relation **(5)** and we again face a superposition of saw-tooth waves.

Example 3.

Find the phase modulation function **s(m)** for the normalized frequency

$$\xi_x = 207/1024$$

with 5 **LSB** neglected. In accordance with **(22)** we get

$$s(m) = m(207/32) - integer(m \times 207/32)$$

The modified continued fraction expansion of **207/32** leads to

k	b_k	a_k	A_k	B_k
0	6	1	6	1
1	2	1	13	2
2	7	1	97	15
3	2	1	207	32

Tab. 2

and the respective series expansion to

$$s(m) = m[6+\frac{1}{2}-\frac{1}{2*15}+\frac{1}{15*32}] - integer(\frac{m207}{32}) \tag{23}$$

see also plot in Fig. 8.

Series expansion **(23)** makes it possible to compute or estimate the levels and frequencies of spurious signals quite easy in comparison (to the author knowledge) with earlier approaches [10,11].

By considering that all modulation periods in the phase function **s(m)** must be coherent we can sum all harmonics of individual saw-tooth waves with the same frequency and arrive at

$$s(m) \doteq \frac{1}{\pi}\sum_{r=1,2,..}^{2^{(B-1)}} \frac{1}{r}\sin(2\pi r m R_b) \qquad (24)$$

where we have introduced the remainder $\mathbf{R_b}$

$$s(m=1) = \frac{X}{2^B} - integer(\frac{X}{2^B}) = R_B \qquad (25)$$

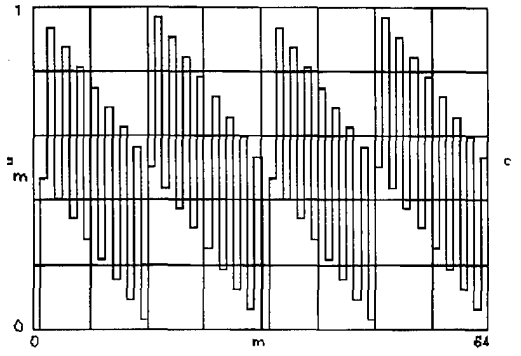


Fig. 8 The plot of the phase truncation function s(m) for the normalized frequency ξ_x = 207/1024 for 5 bits neglected, i.e. B=5.

In instances where $R_b > 0.5$ we may refer without any loss of generality to $1-R_b$, i.e. to refer only to the range

$$0 \le |R_b| = \frac{P}{Q} \le \frac{1}{2} \quad (Q = 2^B) \qquad (26)$$

After some computation we arrive at [4]

$$\frac{2^B}{rY}\{\sin[2\pi m(\frac{rP}{Q} + \frac{X}{Y})] + \sin[2\pi m(\frac{rP}{Q} - \frac{X}{Y})]\} \qquad (27)$$

Evidently both components in the above expression are staircase values of spurious sine waves around the carrier. Consequently we can compute their amplitudes with the assistance of FFT and we arrive at

$$c_{sr} = \frac{2^B}{Y}\frac{\sin[\pi(\frac{X}{Y} \pm \frac{rP}{Q})]}{\pi(\frac{X}{Y} \pm \frac{rP}{Q})} \qquad (28)$$

After dividing the above relation by that for c_x we get for the **r-th** order spurious signal level

$$\frac{c_{sr}}{c_X} = \frac{2^B}{Y}\frac{X/Y}{X/Y \pm rP/Q}\frac{\sin(\pi(X/Y \pm rP/Q))}{\sin(\pi\frac{X}{Y})} \doteq \frac{2^B}{rY} = \frac{2^{-R+B}}{r} = \frac{2^{-W}}{r} \qquad (29)$$

or in dB measure

$$\frac{c_{sr}}{c_X} \doteq -6W - 20\log(r)\ [dB] \qquad (30)$$

Example 4.

With the assistance of FFT applied on eq. **(21)** we have computed the level of spurious side bands for the situation investigated in Example 3. The result is shown in Fig. 9 and reveals that the largest spurious signals are about -30 dB as expected and extend down to about -50 dB - from eq. **(30)** we get -54 dB, which is a fairly good result.

However there are several remarks to be made:
First we would like to know the number of the spurious signals, caused by phase truncation, in the DDFS pass band which in normalized notation is ideally equal to *Y/2*.
The problem was already solved in [4]. We repeat it , here, in short. From eq.**(28)** we can deduce the spurious line numbers (the carrier line number is **X**)

$$n_r = X + (r\frac{PY}{2^B} + sY) < Y/2 \qquad (s = ...,-2,-1,0,1,2,...) \qquad (31)$$

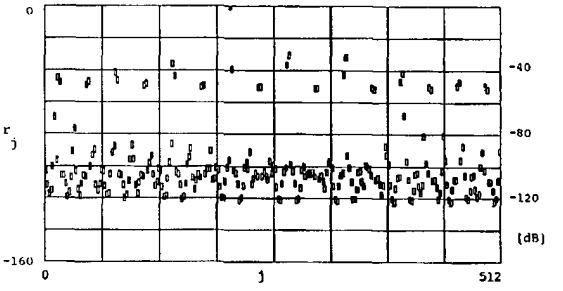


Fig. 9 An example of spurious side bands of ξ_x = $207/2^{10}$ with 5 LSB neglected [4].

Investigation of the above inequality reveals different $\mathbf{n_r}$ only for $\mathbf{r < 2^B/2}$ and a double one for $\mathbf{r=2^{B-1}}$. Consequently the lowest level of spurious caused by phase truncation should not exceed

$$-6(W + B - 1) = -6(R - 1)\ [dB] \qquad (32)$$

In addition the number of side bands is give by (leaving out the carrier)

$$2^B - 1 \qquad (33)$$

Note the 32 large spurious in Fig. 9, and 7 in Fig. 10.
Second: with the assistance of the relation **(31)** we easily arrive at actual spurious frequencies

$$f_{sr} = \frac{n_r}{Y}f_c = f_x + (r\frac{P}{2^B} + s)f_c \qquad (34)$$

Third: investigation of Fig. 9 and 10 reveals a lot of "background" spurious signals which are much smaller than those due to the phase truncation. The problem will be discussed in the next section.

B/ Sine word truncation.

Ideal sine wave with stair case amplitudes

$$\sin(2\pi m\frac{X}{Y}) \qquad (35)$$

would reveal in the output pass band **Y/2** no spurious components. However, the economy of actual **DDFS** requires that the information about the sine values in look-up tables is generally stored only as **S**-bits long words. Consequently, truncation of stored sine values results in changing actual amplitudes into

$$\frac{1}{2^S}integer[2^S\sin(2\pi m\frac{X}{Y})] \qquad (36)$$

or even into

$$\frac{1}{2^S}integer\{2^S\sin[2\pi\frac{2^B}{Y}integer(\frac{mX}{2^B})]\} \qquad (37)$$

Example 5.

With the assistance of the FFT algorithm and relation (37) we shall investigate the **DDFS** spectra for different **W** and **S**. The results are shown in Figs. 10 and 11. Note that for smaller **S** (in Fig. 11) the "background" level is higher (in respect to that in Fig. 10) which one would expect.

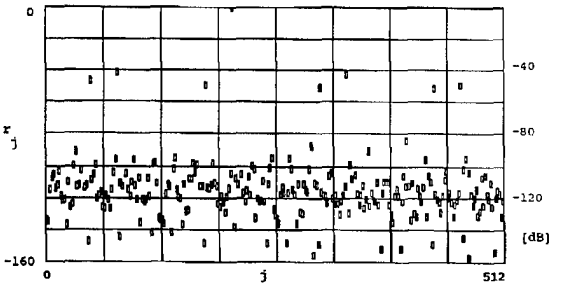


Fig. 10 An example of spurious side bands in DDFS with the normalized frequency $\xi_x = 207/2^{10}$ with 3 LSB neglected, i.e. B=3 and W=7; S=12 [4].

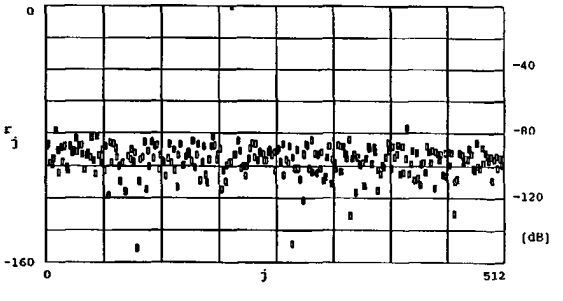


Fig. 11 Background spurious signal level for $\xi_x = 207/2^{10}$ with B=0, W=10 and S=10 [4].

C/ Estimation of the background spurious signal level.

To get more insight into the problem we shall compute the error signal

$$e(m) = \sin(2\pi m\frac{X}{Y}) - \frac{1}{2^S}integer\{2^S\sin[2\pi\frac{2^B}{Y}inetger(\frac{mX}{Y})]\} \qquad (38)$$

From the above relation we easily find out (for **B=0**) that

$$|e(m)_{max}| < 1/2^S \qquad (39)$$

By plotting the cumulative [12] distribution of **e(m)** we find for larger **m** and **S** a straight line with two different slopes - see Fig. 12.

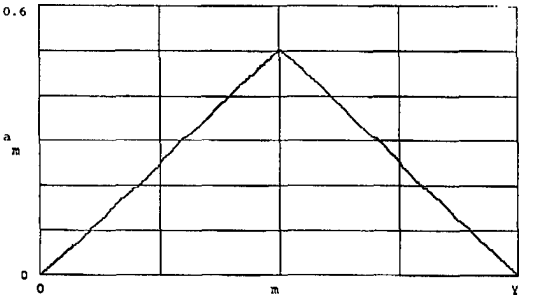


Fig. 12 Cumulative distribution of sine word errors: m = Y = 1024

Evidently the amplitude distributions of **e(m)'s** are constant in the ranges between **$-1/2^S$ to 0** and from **0 to $+1/2^S$** - i.e.

$$\begin{array}{ll} p(e) = 1/(2*2^{-S}) & \textit{inside the interval} \\ p(e) = 0 & \textit{outside the interval} \end{array} \qquad (40)$$

As a consequence we find for the variance

$$\sigma^2(e) \doteq (1/3)*2^{-2S} \qquad (41)$$

which agrees quite well with actually computed values as it is shown in the following example.

Example 6.

Compare the computed variances of $\sigma^2(e)$ with the theoretical estimations by **(41)**.

S	$\sigma^2(e)$	var(e) Y=32	var(e) Y=128	var(e) Y=1024
8	5.09 e^{-6}	5.33 e^{-6}	5.18 e^{-6}	5.48 e^{-6}
10	3.18 e^{-7}	2.85 e^{-7}	3.44 e^{-7}	3.15 e^{-7}
12	1.99 e^{-8}	0.96 e^{-8}	1.90 e^{-8}	2.00 e^{-8}
14	1.24 e^{-9}	1.16 e^{-9}	1.40 e^{-9}	1.27 e^{-9}

Tab. 3

Investigation of the Tab. 3 in the above example reveals rather small differences between estimated and actually computed variances. As a consequence we can conclude that we are justified to use relation **(41)** in our pursuit of the estimation of the "background" spurious signal levels.

Further Figs. 8, 9, and 10 reveal that there might exist a mean level $\langle r_j \rangle$ of the spectral lines r_j with some dispersion. To this end we have again plotted the cumulative distribution of positive values of $|r_j|$ and got straight lines one of which is shown in Fig.13. The evident conclusion is that the background spectral lines have also a uniform distribution.

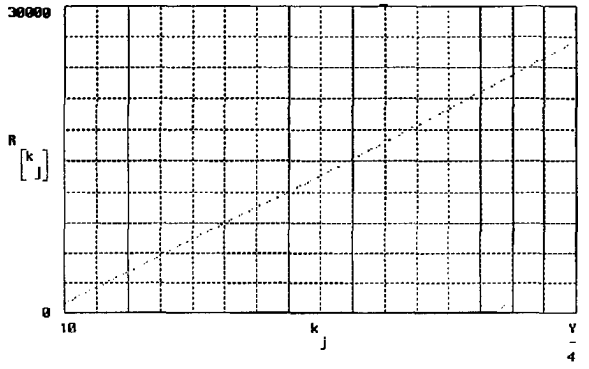


Fig. 13 Cumulative distribution of the positive values r_j .

Since the number of them in the pass band, extending from 0 to **Y/2**, is

$$Y/4 - 1 \qquad (42)$$

(after leaving out the carrier and all spurious signals with even line numbers, which are negligibly small) we may assume for the mean power of the background spurious signals

$$\langle S(n) \rangle \sim \frac{\sigma^2(e)}{Y/4-1} \approx \frac{1}{3} * 2^{-2S} * \frac{4}{Y-4} \qquad (43)$$

For large **Y** we can approximate **<S(n)>** in the dB measure as

$$\langle S(n) \rangle \approx 1.25 - 6S - 10 * \log Y \quad [dB] \qquad (44)$$

However, computer simulations summarized in Example 7. reveal values about **-6 dB** lower.

To remove these discrepances we have tried a new approach. In a recent paper [5] we have computed the expected level of the background spurious signals with the assistance of the "Quantization noise spectra" theory [13]. This time we have got a rather good agreement (see the **Appendix**), i.e.

$$\langle S(n) \rangle \approx \frac{1}{3Y} * 2^{-2S} \qquad (45)$$

Example 7.

Computation of means **<S(n)>** and their dispersions for different **Y** and sine look-up resolutions **S**.

Y	S	<S(n)>	σ
32	8	-57.77	1.28
	10	-77.88	3.61
	12	-91.99	4.58
128	8	-75.27	13.04
	10	-87.58	8.08
	12	-101.57	11.14
512	8	-81.81	9.51
	10	-93.21	9.19
	12	-106.09	9.55
1024	8	-85.99	10.95
	10	-96.80	9.68
	12	-109.91	10.18

Tab. 4

Finally in Example 7 we have also computed dispersions of the actual values of the background spurious signals around the mean.

D/ Estimation of the spurious harmonics.

A closer investigation of many simulations of background spurious signals in **DDFS** reveals that there are few of them exceeding the dispersion level of $+2\sigma$. An example is shown in see Fig. 14. We have looked for their source and found it in higher harmonics of the output signal f_x.

Let us revert to eq. **(38)** and compute the mean value of **e(m)** both for the positive and negative half period of the output sine wave. We get a rectangular wave with the amplitude $2^{-S}/2$. Its Fourier series expansion reveals only odd harmonics with amplitudes

$$a_h = \frac{2*2^{-S}}{\pi} * \frac{1}{h} \quad (h = 3,5,\ldots) \qquad (46)$$

For the level of the 3rd spurious harmonic we get in dB measure

$$20 * \log\left(\frac{c_{s,3x}}{c_x}\right) \approx -6S - 14 \quad [dB] \qquad (47)$$

Note that we have neglected the effect of sampling on amplitudes. The above relation is in a good agreement with many computer simulations and experimental measurements. However, the 5th and higher harmonics are often masked by the dispersion of the spurious

signals - see Fig. 15.

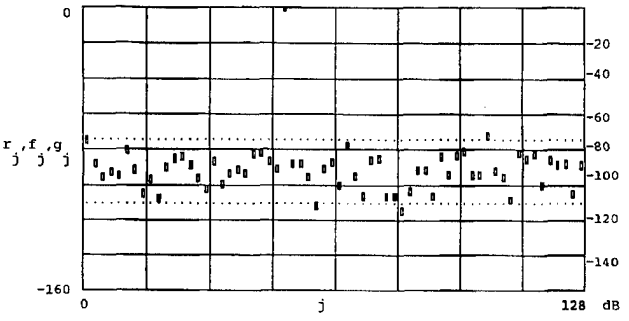


Fig. 14 FFT simulation of background spurious signals for S=10 and ξ_x = 51/256; points indicate 2σ dispersion lines around the mean.

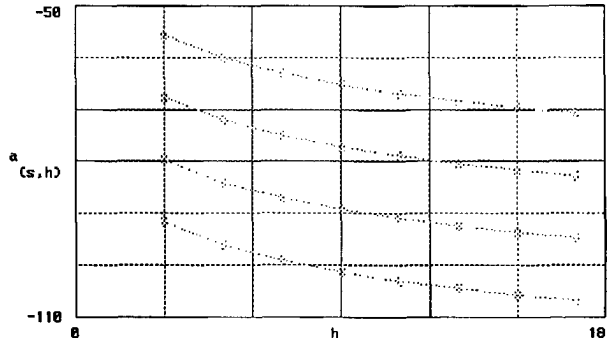


Fig. 15 Expected amplitudes of spurious harmonics of the output signal (S=7,9,11,13).

Generally only the 3rd harmonic exceeds the background level. The remedy provides rounding off sine values in look-up tables to the nearest integers. The problem may reappear in instances where the length **S** exceeds the resolution of the used **DAC**.

E/ Relation between **W** and **S**.

In instances where we would like that spurious signals due to the truncation do not exceed the most prominent signal caused by amplitude truncation we must compare relations **(47)** with **(30)** and we arrive to the following inequality

$$W - 3 \le S \le W - 2 \qquad \textbf{(48)}$$

IV. Digital to analog converters.

There are three major sources causing degradation of the output spectral purity by **DAC**, namely, finite bit resolution, nonlinearity, and transient effects or glitches.

A/ Finite bit resolution.

From the point of view of present technology we are not able to build **DAC** with any arbitrary resolution. On the other hand the higher the resolution the more expensive **DAC**s are. Consequently we encounter in **DDFS** practice **DAC** with resolutions

$$D = 8 \textit{ till } 12 \qquad \textbf{(49)}$$

For discussion of the **DAC** bit resolution we have to take into account that from the **S** bit long sine word we use only **A** bits for converting the digital information about the amplitude into its analog version. Note that in instances where a **D** bit **DAC** is used in **2's** complement version

$$A = D - 1 \qquad \textbf{(50)}$$

The consequence is that in instances where **S > A** we must replace in all relations for computation of background spurious signals the sine word length **S** with the bit resolution **A** (cf. paper by Garvey et al.[14]). The conclusion is that the extent of the sine look-up table and the bit resolution of **DAC** are closely related in properly designed **DDFS**.

B/ DAC nonlinearity.

Investigation of the real output spectra of **DDFS** reveal also even harmonics of the output signal and odd harmonics generally higher then values in accordance with **(46)**. The reason is the nonlinearity of the **DAC** transfer characteristics which is difficult to predict since these differ between individual **DAC**.

C/ Leakage of the clock signal.

We can hardly prevent the leakage of the clock signal into **DAC**. Its nonlinearity and its behaviour as a switching modulator [cf.14] provides source of intermodulation signals, in the output pass band, of the type

$$f_{u,v} = f_x u + f_c v \quad (u,v = 0,\pm1,\pm2,\ldots) \qquad \textbf{(51)}$$

A particular difficulty is encountered in instances where the normalized frequency is close to **1/4**. In that case both the intermodulation signal and the aliased 3rd harmonic, in accordance with **(46)**, are superimposed and very close to the desired output frequency.

V. Output phase noise.

In the last year paper [4] we have arrived at the conclusion that the background spurious signals often form a "continuum". In short we have, eventually, in the **1 Hz** frequency bandwidth

$$1/\Delta f_{s,s} = Y/2f_c \tag{52}$$

and for

$$Y > 2f_c \tag{53}$$

we get with the assistance of **(45)** for the "white mean background noise"

$$S(f) \approx \frac{<S(r)>}{\Delta f_{s,s}} \doteq \frac{1}{3}\cdot\frac{2^{-2A}}{f_c} \tag{54}$$

or in **dB** measure approximately

$$-5 \quad -6*A \; - \; 10*\log(f_c) \qquad [dB] \tag{55}$$

In instances where **Y** is a small integer the above relations are not met, the background spurious signals form a comb with a real phase noise between them. However, it source is not yet fully understood. It seems that it is higher than the clock oscillator noise reduced by the division factor ξ_x or the phase noise of "dividers" and amplifiers in the respective **DAC**.

VI. Conclusions.

In the present paper we have derived a simple closed form solution for frequencies and power level of spurious signals in **DDFS** with square wave outputs, with triangular outputs, and sine wave outputs.

Since **DDFS** are generally plagued with spurious signals, sometimes even large, and often of unknown or dubious origin we have presented a lot of computer simulations to provide the reader with information about the lowest level of spurious signals below one cannot get.

Appendix.

Computation of the background power spectral density.

After replacing the variable **m** with **t** in relation **(38)** we can expand the random variable $\epsilon(t)$ as

$$\epsilon(t) = \sum_{-\infty}^{\infty} a_n e^{jn\omega_o T_i} \qquad \omega_o = 2\pi/YT_i \tag{56}$$

Evidently the n-th spectral line of the k-th rectangle $\epsilon_k(t)$ is

$$a_n = \frac{1}{YT_i}\int_{kT_i-T_i/2}^{kT_i+T_i/2} e(t)e^{-jn\omega_o t}dt$$
$$= \frac{1}{YT_i} * e(kT_i) * e^{-jn\omega_o kT_i}\frac{\sin(n\omega_o T_i/2)}{n\omega_o/2} \tag{57}$$

from which the variance is

$$a_n a^*_n = \frac{1}{Y^2}\sum e^2(kT_i)Sin^2(n\omega_o T_i/2)$$
$$= \frac{1}{Y}\frac{\sigma_e^2}{2}Sin^2(n\omega_o T_i/2) = \alpha_n \tag{58}$$

where we have introduced

$$\frac{\sin(x)}{x} = Sin(x) \tag{59}$$

Now we proceed with computation of the autocorrelation function $\mathbf{R(\tau)}$

$$R(\tau) = \sum_{-\infty}^{\infty} \alpha_n e^{jn\omega_o \tau} \tag{60}$$

From it we easily arrive at the power spectral density

$$S(\omega) = \int_{-\infty}^{\infty} R(\tau)e^{-j\omega\tau}d\tau$$
$$= \int_{-\infty}^{\infty}\sum \alpha_n e^{jn\omega_o\tau}e^{-j\omega\tau}d\tau \tag{61}$$
$$= \sum \alpha_n \delta(\omega - n\omega_o)$$

and finally for the mean value of one sided power spectral density we get with the assistance of **(58)** and **(41)**

$$S(n\omega_o) = \frac{1}{3Y} * 2^{-2S} Sin(n\omega_o T_i/2) \tag{62}$$

Since the transfer function **Sin(x)** is rather flat in the pass band from zero to **Y/2** (the maximum error is -4 dB) we feel justified with the simplification used in relation **(45)**.

References:

[1] V.F. Kroupa, "Spectra of Pulse Rate Frequency Synthesizers," Proc. IEEE, Vol.67, pp. 1680-82, December 1979.

[2] V.F. Kroupa, "Spectral purity of direct digital frequency synthesizers," Proc. 44th Annual Symposium on Frequency Control, pp. 498-510, May 1990.

[3] V.F. Kroupa, "Principles of direct digital frequency synthesizers," Proceedings: Kleinheubacher Berichte, Band 36, pp. 663-678, 1993.

[4] V.F. Kroupa, "Discrete spurious signals and background noise in direct digital frequency synthesizers," Proc. of the 1993 IEEE International Frequency Control Symposium, pp. 242-249, June 1993.

[5] V.F. Kroupa, "Intermodulation signals in direct digital frequency synthesizers," to be published in the Proceedings of the 8th European Frequency and Time Forum, held March 9-11, 1994.

[6] B. Giebel, J. Lutz, and P.L. O'Leary, "Digitally Controlled Oscillator," IEEE J. of Solid-State Circuits, vol. 24, pp. 640-5, June 1989.

[7] P.H. Saul and D.G. Taylor, "A High-Speed Direct Frequency Synthesizer," IEEE J. of Solid-State Circuits, vol. 25, pp.215-9, 1990.

[8] Fang Lu, H. Samueli, J.Yuan, and Ch. Svensson, "A 700-MHz 24-b Pipelined Accumulator in 1.2 μm CMOS for Application a Oscillator,"IEEE J. of Solids a Numerically Controlled-State Circuits, vol. 28, pp. 878-86, August 1993

[9] G.W. Small," A Frequency Synthesizer for $10/2\pi$ kHz," IEEE Tr. on Instr. and Measurement, Vol. IM-22, pp. 34-7, March, 1973.

[10] S. Mehrgardt, "Noise Spectra of Digital Sine-Generators Using the Table-Lookup Method", IEEE Trans. ASSP-31, pp. 1037-39, August 1983.

[11] H.T. Nicholas, H. Samueli and B. Kim, "The Optimization of Direct Digital Frequency Synthesizer Performance in the Presence of Finite Word Length Effects". Proc. 42nd AFCS, pp. 357-363, 1988.

[12] Gray, M. Robert, "Quantization Noise Spectra," IEEE Trans. on Information Theory, vol. IT-36, (1990), No. 6, pp. 1220/1244.

[13] Gernot M. R. Winkler, "Introduction to robust statistics and data filtering," presented at the 1994 IEEE Annual Symposium on Frequency Control (Tutorials), 31 May 1994.

[14] J.F. Garvey and D. Babitch, "An exact spectral analysis of a number controlled oscillator based synthesizer", Proc. 44th AFCS, pp. 511-521, May 23-25, 1990.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DESIGN AND ANALYSIS METHODS OF A DDS-BASED SYNTHESIZER FOR MILITARY SPACEBORNE APPLICATIONS

Thomas J. Endres, Robert B. Hall, and Alan M. Lopez
Hughes Space and Communications Company
P.O. Box 92919
S04, M/S S330
Los Angeles, CA 90009

Abstract

Modern wideband, multi-user antijam communication systems typically employ frequency hopping techniques to maximize system processing gains. Such frequency plans place strict constraints on the hopping LO in the form of spurious, phase noise, and transient response criteria. Historically, synthesizer designers have relied on phase lock techniques to meet these demands. However, the demand for compact satellite communications payloads, with reduced size, weight, power, and cost, is revealing the direct-digital synthesizer (DDS) as the next-generation synthesizer workhorse.

To date, application specific integrated circuit (ASIC) development has not progressed sufficiently to accommodate the clock rates and edge speeds necessary for DDS compliance with military satellite communication (MILSATCOM) specifications. Original design and analysis methods are therefore required to employ the DDS, at the expense of hardware complexity.

This paper discusses the translation of a low frequency, narrowband DDS spectrum to a Ku-band, 1 GHz bandwidth (BW) output that meets the strict spurious, phase noise, and transient specifications required of MILSATCOM spacecraft payloads. Requirements are listed in Table 1 below.

Introduction

Low spurious, low phase noise frequency synthesizer architectures historically have used some type of phase lock technique and upconversion circuitry to meet the requirements for MILSATCOM payloads. Typically, an RF signal from a voltage-controlled oscillator (VCO) is divided and phase-locked to a known reference (Figure 1).

A narrow loop bandwidth for phase noise and spurious performance usually requires acquisition or help circuitry for fast phase settling. In Figure 1, a digital pretune coarse tunes the VCO, allowing the loop to track out any additional error. Such settling characteristics may also require high loop gains, creating susceptibility to undesired disturbances and leading to PLL performance that severely deviates from ideal [1], [2].

In addition to circuitry surrounding the loop, a dual or ping/pong approach is often required; that is, one loop tunes or acquires while the other loop is on the air. This ping/pong architecture, however, comes at the expense of precious size, weight, power, and cost.

Such PLL deficiencies have led synthesizer designers to consider alternative architecture approaches to meet the dynamic MILSATCOM requirements [3]. Hughes' evolution of MILSATCOM synthesizers is shown in Figure 2.

Switch and Mix Synthesizer

An alternative to phase lock techniques is the DDS approach, which utilizes high-speed digital ASIC technology. Historically, poor spurious performance from the DDS has limited its use in high performance satellite payload applications, but recent advances in high speed digital technology have made the DDS highly viable and efficient source for fine frequency generation. However, even with these technology advances, the DDS is still limited in its

Table 1. Synthesizer Requirements

Hop Bandwidth	1 GHz (Ku Band)
Discrete Spurious	
≥ 15 MHz Offset	≤ −65 dBc
Integrated Spurious/Phase Noise (SSB)	
100 Hz to 15 MHz	≤ −38 dBc
Noise Floor (SSB)	
≥ 15 MHz Offset	≤ −127 dBc/Hz
Phase Settling	≤ 8 deg in 600 nsec

0-7803-1945-1/94 $3.00 © 1994 IEEE

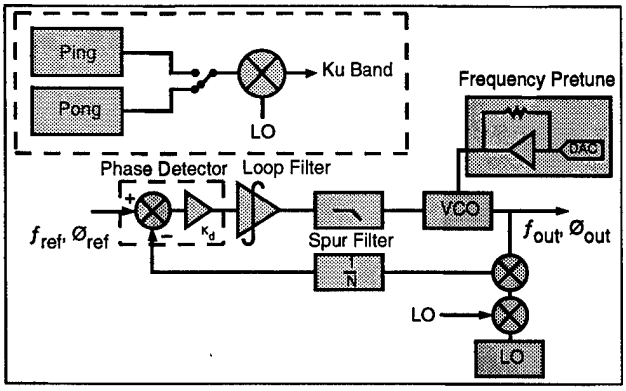


Figure 1. PLL Used as a Frequency Synthesizer

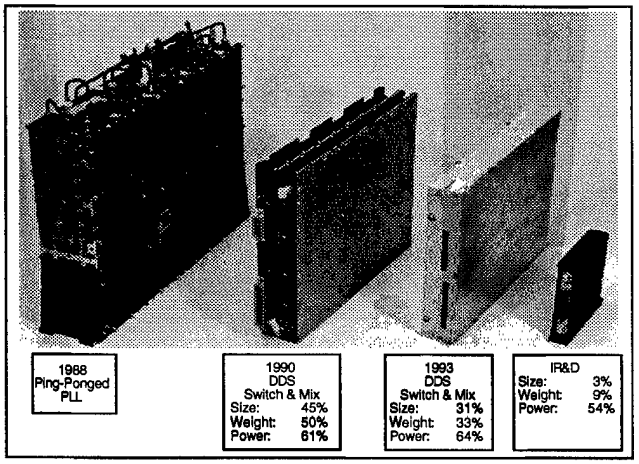


Figure 2. Synthesizer Evolution

spurious performance and usable bandwidth and therefore requires compensation elsewhere in the synthesizer.

One switch and mix approach provides the necessary bandwidth expansion and upconversion via comb generators, switched LOs, and high-performance mixers (Figure 3).

Four LOs are generated from a single RF input by multiplication via comb generation. Each pair of LOs is then switched into one of two upconversion mixers to expand the fine frequency DDS output to the desired Ku-band output with a 1 GHz BW. The first mix quadruples the DDS BW by independently selecting upper and lower sidebands for each of the two LOs. Switched filtering is required to remove undesired LOs and low-order mixer intermodulation products. The second mix doubles the BW again by mixing with the second pair of LOs using only one sideband, resulting in an overall bandwidth expansion of eight times the DDS bandwidth.

DDS

At the heart of the switch and mix architecture is a 500 MHz-clocked DDS used for VHF fine frequency generation (Figure 4). The phase accumulator is simply a linear modulo counter which increments with each clock cycle at a rate dependent on the frequency control word (FCW). This linear phase ramp is converted to a sinusoidal amplitude via a ROM algorithm. The digital to analog converter (DAC) converts the digital word to an analog output, which is typically filtered to remove undesired frequency components. Spur generation within the DDS may be attributed to various mechanisms.

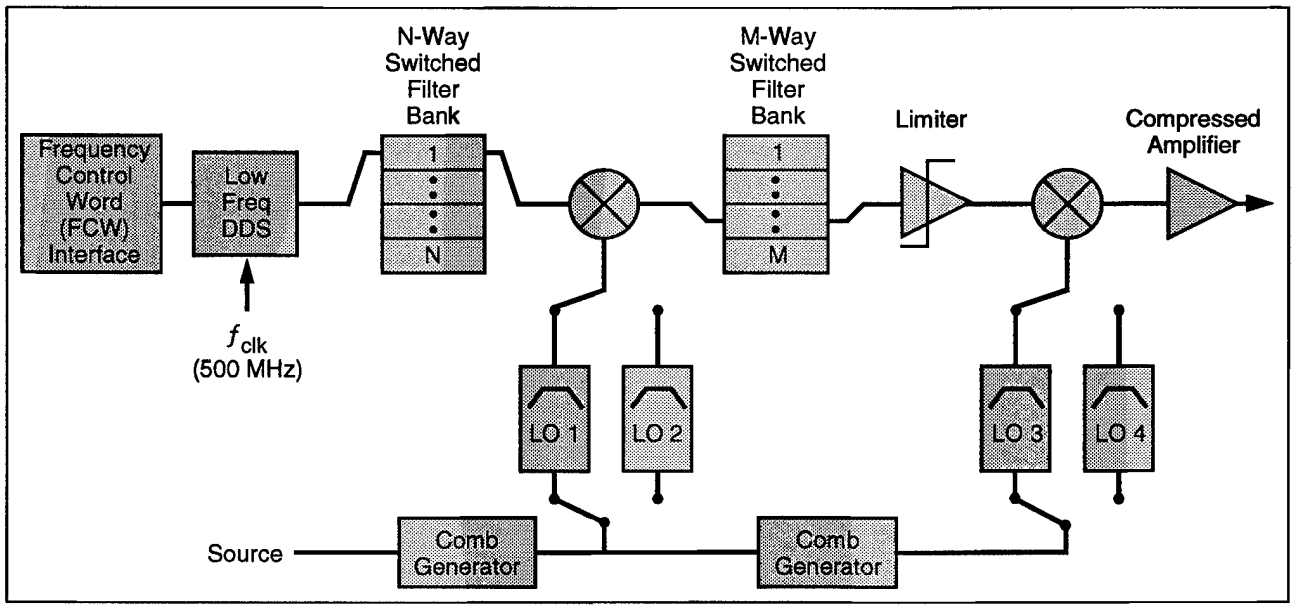


Figure 3. DDS-Based Switch and Mix Synthesizer Architecture

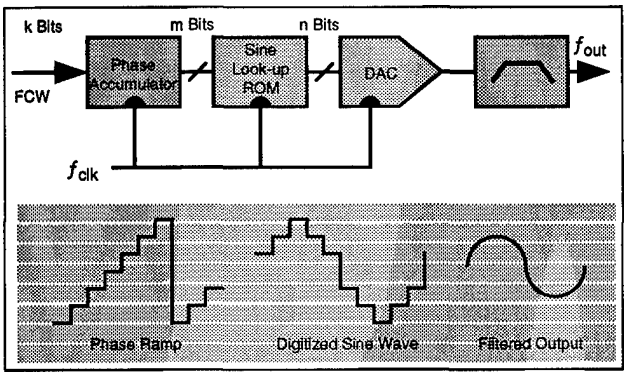


Figure 4. DDS for Fine Frequency Generation

Bit Truncation

Ideally, all bits of accumulator phase information are used by the ROM in the transformation to amplitude. However, ROM size limitations and complexity typically constrain the usable number of phase bits, requiring truncation between the accumulator and ROM. The effect of this truncation is to cause deviation from the ideal sawtooth ramp (Figure 5).

In this simple example, two of four phase bits are truncated when 5/16 of the clock frequency is synthesized. The resulting jitter is observed as phase modulated spurious at the DDS output. A similar truncation may also be seen between the ROM and DAC, resulting in amplitude modulated spurious. These truncation effects, representing the theoretical spurious limit of the DDS, are well understood and accurately modeled [4], [5], [6].

DAC Nonlinearities

Perhaps the most important but least understood DDS spur mechanism is the dynamic response of the DAC. Such nonlinearities are manifested in the time domain as glitch energy, slew rate limiting, and waveform asymmetries (Figure 6). In this example, the glitch energy is seen as overshoot as from an underdamped condition when the DAC changes state. Such ringing may be considered as errors from ideal values of discrete samples. These deviations therefore transform to significant DDS spurious. Fourier analysis also shows that asymmetric rise and fall times (possibly due to slew rate limiting in the DAC) generate spurious harmonically related to the synthesized tone.

Sampling Theory

In addition to these harmonics, sampling theory predicts aliased harmonics that may be described in frequency by

$$mf_{clk} \pm nf_o$$

for integers m and n, synthesized tone f_o, and clock frequency f_{clk}. This results in a DDS spectrum as shown in Figure 7 [7].

Post-DDS filter design, when generating more than an octave of BW (represented by the shaded rectangle in Figure 7), must address the following conditions. First consider the case when a low frequency tone is synthesized, resulting in low-order, in-band harmonics which require

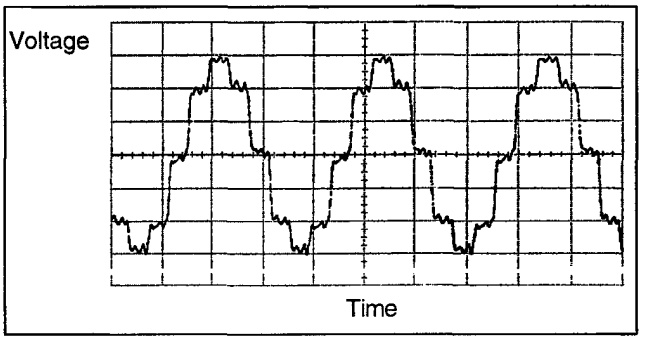


Figure 6. Twelve-bit DAC Output for 1/8 of the Clock Synthesized with Zero Initial Phase

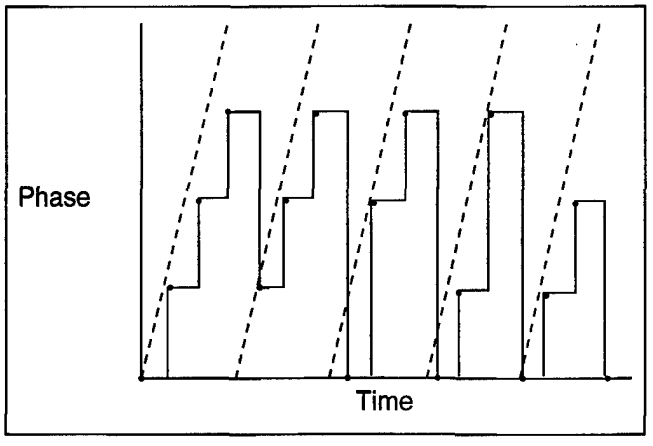


Figure 5. Phase Bit Truncation Example. Two of four bits truncated. $F_{out} = (5/16)(F_{clk})$

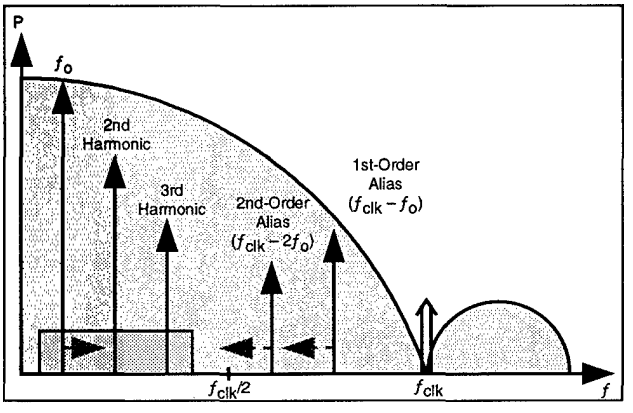


Figure 7. DDS Spectrum with Harmonics and Aliased Tones

filtering. Next, consider the case when a higher frequency tone is synthesized. In this case, the DDS spectrum replicated about the clock frequency results in aliased harmonics that walk in-band as the synthesized frequency increases. Indeed, these "walk-through" spurs will approach and eventually pass through the carrier as the DDS tunes higher in frequency. These spurs may therefore drive both close-in integrated spurious/phase noise requirements as well as filter rejection specifications.

Hence, the low spurious, wideband synthesizer requires wide DDS usable bandwidth to minimize the amount of upconversion/bandwidth expansion circuitry. The expense paid for this DDS bandwidth is undesirable in-band spurious, which necessitates the use of complex filtering design techniques.

Software Interface/Spur Mapping

Since wide variances among DDS chipsets in these spur mechanisms are common, and frequency changes corresponding to as little as an accumulator LSB typically result in seemingly uncorrelated spectra with a myriad of spectra available, design of the switched filter bank following the DDS requires original methods based on known DDS performance.

The switched filter bank design approach discussed here processes empirical DDS data with custom software (Figure 8).

The spectrum analyzer measures empirical DDS data over the synthesized bandwidth or bandwidth of interest for a DDS step size that gives a reasonable number of spectra while remaining computationally manageable. One thousand and one points per spectrum are translated to an ASCII format and dumped to a PC.

Custom MATLAB code derives N Chebyshev BPFs of arbitrary order, center frequency, ripple, and bandwidth. These filter responses start with the n pole locations of a Butterworth LPF. The pole locations are then mapped to an ellipse in the s-plane, representing the poles of a Chebyshev LPF. These poles are next mapped to a Chebyshev BPF with n zeros at the origin [8], [9]. The exact filter responses are then slightly modified in bandwidth, order, and ripple to closely match real-world filter behavior, accounting for finite isolation, coupling, insertion loss, Q, etc.

The software next partitions the vector of DDS data for combination with corresponding BPF response. Two sets of spectra are then output by the software, one unfiltered with the corresponding filter response and the other filtered, as shown in Figures 9 and 10.

In Figure 9, exactly 1/8 of the clock frequency is synthesized, causing aliases and harmonics to lie on top of each other, so that spurs are harmonically related to the fundamental. Next consider the case when 33/256 of the clock frequency is synthesized, corresponding to $1/8f_{clk} + \Delta$. In this case, not all spurs are harmonically related to the carrier, resulting in many low-level, aliased spurs that require filtering. In both cases, the filtered spectra are sufficient for bandwidth expansion and upconversion to meet MILSATCOM requirements.

In Figure 10, approximately 1/3 of the clock frequency is synthesized. In the first case ($f_{DDS}/f_{clk} = 21/64$), the aliased spur resulting from the clock frequency minus twice the fundamental is slightly attenuated by the filter. However, as the DDS tunes up to 85/256 of the clock frequency, this aliased spur walks down towards the carrier so that it is unattenuated by the filter and within 15 MHz of the carrier. As previously mentioned these walk-through spurs typically drive close-in integrated spurious and phase noise as well as filter rejection requirements.

In all cases, this custom MATLAB code affords near real-time filter design specifications based on empirical DDS data.

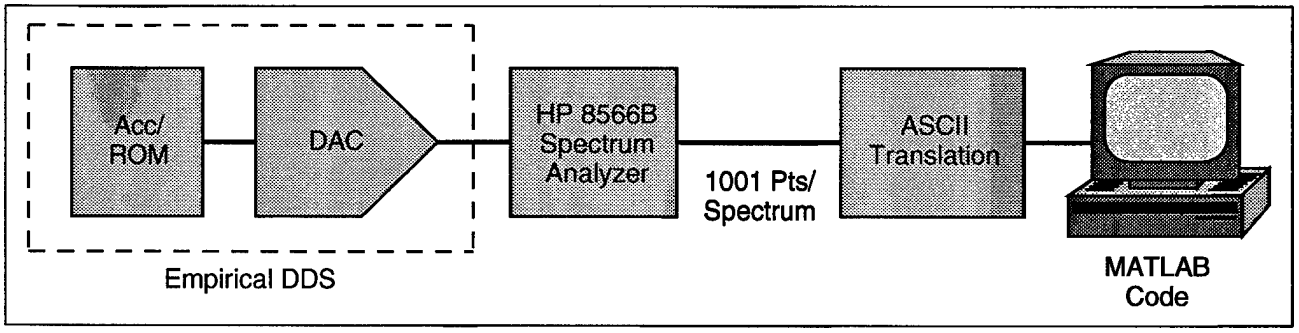


Figure 8. Empirical DDS Data Processed Via Custom MATLAB Software

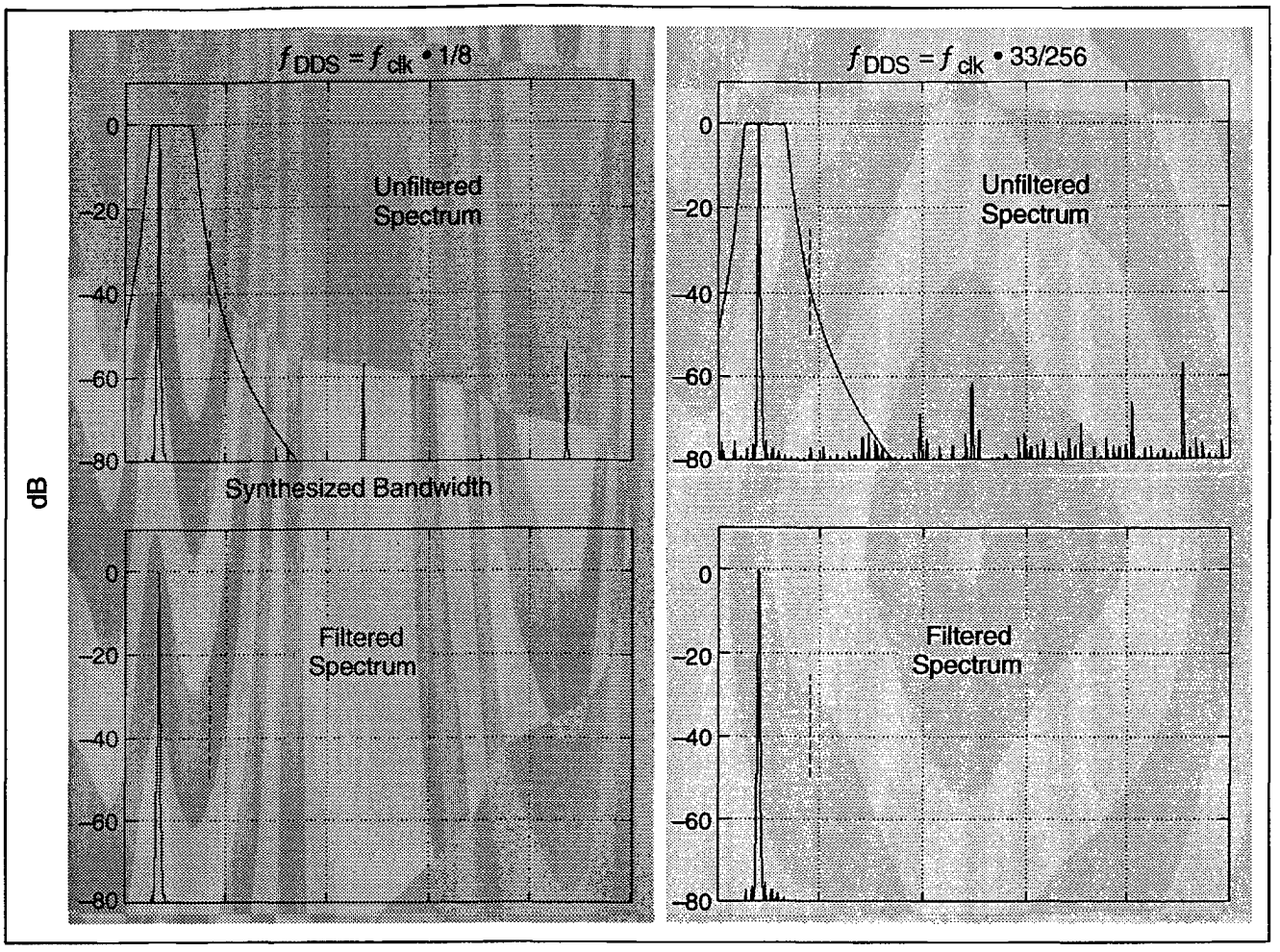


Figure 9. High Resolution MATLAB Reproduction for f_{DDS}/f_{clk} = 1/8 and 33/256. Dashed line represents 15 MHz carrier offset requirement.

Settling Time

The expense paid for the narrowband filtered DDS spectra is significant group delay, which directly impacts phase settling performance. Although settling time performance for PLL-based synthesizers has been well analyzed and published, less documentation exists for the DDS-based architecture [1], [2].

Switch and mix synthesizer settling time may be bounded by determining the maximum values of each contributor, in terms of group delay for the filters, propagation delay for the digital components, and settling time for switches and compressed RF amplifiers. The worst case settling time path for the synthesizer is the RF path depicted in Figure 11. The switched LO paths are negligible contributors since the large delay in the main path is such that the LO paths have settled while the RF signal still propagates in the main path.

Filters/Switches

A significant contributor to overall synthesizer settling performance is group delay due to both the N-way DDS switched filter bank before the first mix and the M-way switched filter bank following the first upconversion mix. These narrowband, high order Chebyshev filters are necessary to meet the stringent synthesizer spurious requirements. The magnitude of each filter's contribution to delay/settling time is analyzed and bounded by modeling from both FILSYN (filter simulation program) and MATLAB simulations and verified by empirical testing.

Because the transient response of these filters due to a switched RF signal is ideally negligible compared to group delay, it is not a contributor. In order to select the required filter and to provide for proper isolation between filter inputs and outputs, both switched filter banks contains two FET

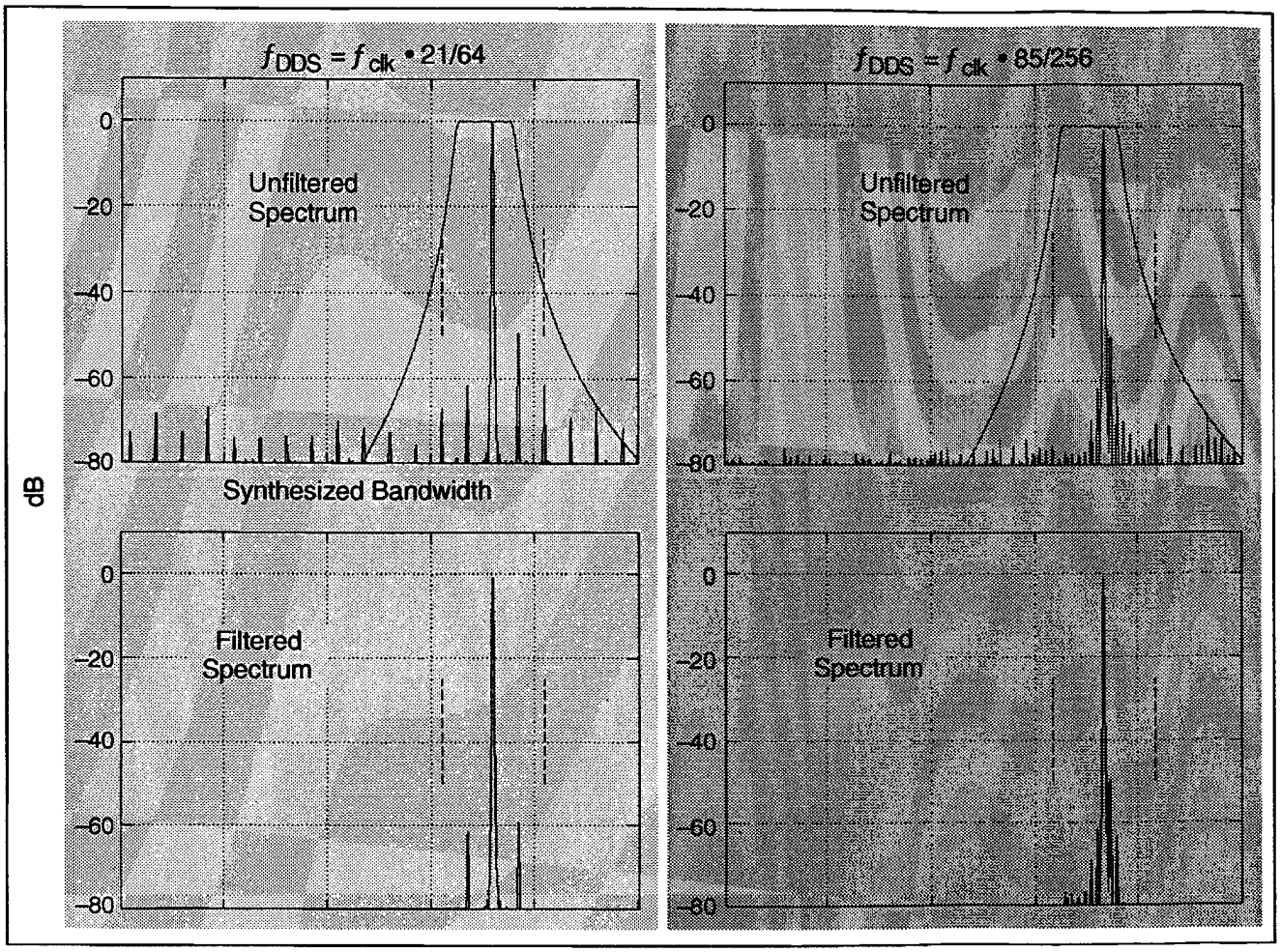


Figure 10. High Resolution MATLAB Reproduction for f_{DDS}/f_{clk} = 21/64 and 85/256. Dashed line represents 15 MHz carrier offset requirement.

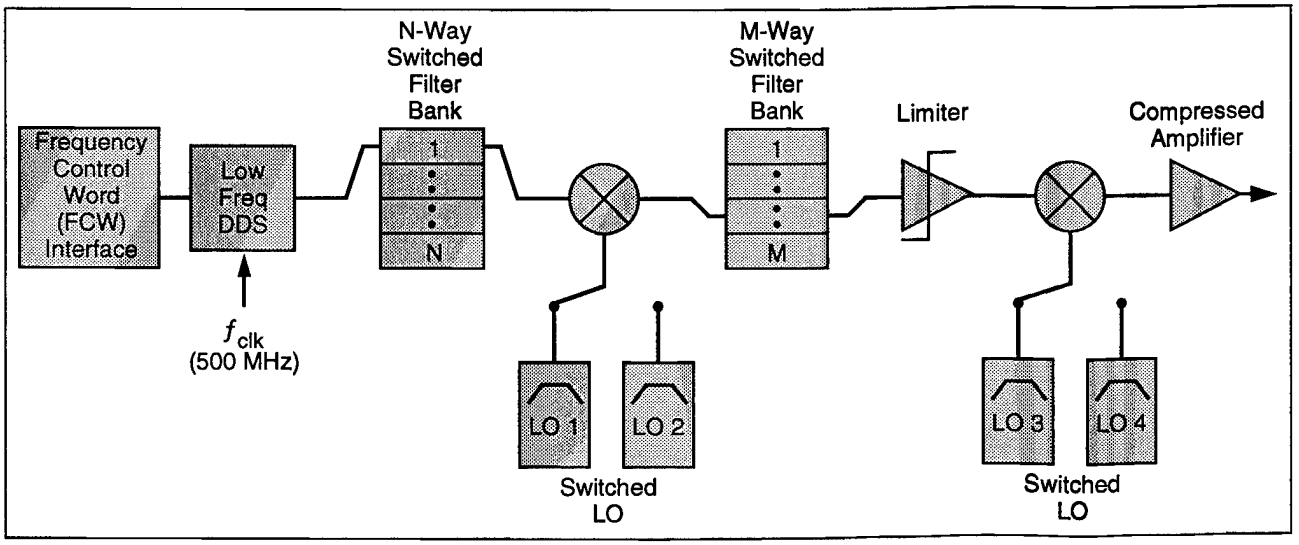


Figure 11. Phase Settling Contributors

switches on both the input and output that also contribute to the synthesizer's settling time. These switches have inherent propagation delays due to the internal CMOS decoders for the control lines and the switch drivers, in addition to the amplitude and phase settling characteristics of the switches themselves. Moreover, interaction between the switches and the filters may cause settling characteristics worse than the predicted sum of the switches and filters due to VSWR effects.

Digital Circuits

Another significant contributor to the overall synthesizer settling is the propagation delay of the digital circuits. The digital circuits consists of the Command Logic (CL), which includes TTL differential line receivers used for the reception of the frequency control word (FCW) for the DDS ACC/ ROM. The worst case propagation delay for the CL circuits is based on circuit simulations, including wafer process variation (CL implemented via CMOS SOS ASIC) and the calculated effects of temperature, voltage, and radiation. An additional contributor to the digital delay is the DDS ACC/ ROM chip. Its settling time is actually a latent delay due to the number of clock cycles required to load in the new FCW. This delay is dependent on the details of the DDS architecture and the allowable single-event-upset error rates caused by the space environment [10].

RF Amplifiers

Compressed discrete and MMIC amplifiers used to minimize power variations in the high frequency LO chain and RF output chain may significantly impact settling performance and are possibly the most difficult contributors to model for transient behavior. Key parameters affecting compressed amplifier settling characteristics are input drive level, input amplitude step size, input frequency step size, level of compression, and bias networks. The phase settling waveform of the amplifiers may generally be considered as a decaying exponential, the time constant of which is related to the device physics and is therefore inalterable. Hence, the input parameters are optimized so that the amplitude of this phase error is minimized and phase settling performance is desirable.

Settling Time Bound

Synthesizer settling time is bounded by a linear addition of propagation and group delays with the root sum squared (RSS) of RF settling components. This RSS approach for the RF components is valid since these waveforms are uncorrelated. Allocations and results are summarized in Table 2.

An alternative method to bound settling time is a sum of delayed decaying exponentials with various time constants, representing the waveforms of the different contributors. This approach yields results similar to those in Table 2.

Figure 12 shows the synthesizer settling waveform when both LOs are switched and group delay of both filter banks is maximal. This worst case scenario yields settling to 8 deg. in 470 ns, showing compliance.

Phase Noise

The overall phase noise of the switch and mix synthesizer is a function of the phase noise of the reference LO used to derive the switched LOs and DDS clock. Since the DDS is

Table 2. Summary of Synthesizer Phase Settling Contributors and Unit Performance

Contributors	*Time to 8 deg or Prop/Group Delay (nsec)*	
	Allocations	Measurement
FCW Interface	86	86[1]
DDS ACC/ROM	60	60[1]
Switched N-Way BPF	175 + 75	237[2]
Switched M-Way BPF	65 + 75	130[2]
Limiter	131	50
Compressed Output Amp	131	50
Switched LOs 1 and 2		25
Switched LOs 3 and 4		25
Total (Linear Sum of Delays plus RSS of RF Circuits)	600	470

Notes: [1] Simulation of circuit models
[2] Switched BPF includes group delay and settling time of switches

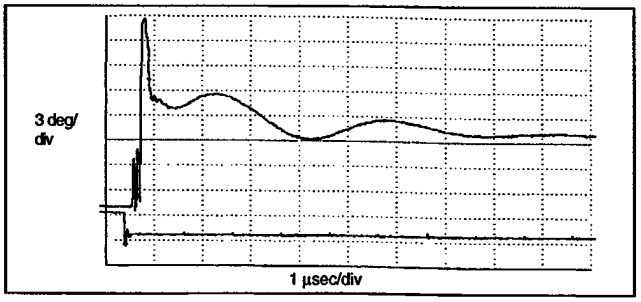


Figure 12. Synthesizer Settling Time

essentially a frequency divider, the DDS phase noise is ideally a 20 log (N) reduction of the clock phase noise, where $N = f_{out}/f_{clk}$ is the fraction of the clock frequency that is synthesized. However, since the switched LOs are derived via large multiplication of the reference LO, the DDS phase noise is dominated by the consequent enhancement of the reference LO's noise. Typically, narrowband filtering of the switched LOs is required to minimize these effects. Moreover, the noise figure of low drive level RF amplifiers will further degrade the noise density. A typical SSB phase noise power spectral density is shown in Figure 13 for approximately 13 GHz carrier frequency.

Summary/Conclusions

This paper has demonstrated that the tradeoff between settling time and spurious requirements necessary for MILSATCOM payloads can be met with existing DDS technology, at the expense of hardware complexity. Such architectures require original design and analysis methods. Hughes Space and Communications Company is dedicated to be the world leader in compact, lightweight, low-power spacecraft payloads. Future synthesizers to meet this goal will be achieved utilizing high-speed, low-spurious DDS architectures which minimize the need for filter banks and LO generation.

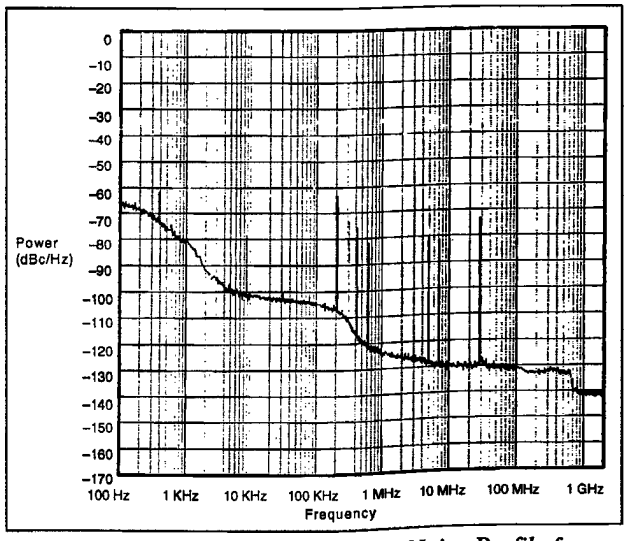


Figure 13. Synthesizer SSB Phase Noise Profile for Approximately 13 GHz

Acknowledgments

The authors wish to acknowledge the design and test capabilities of Gar Calvetti, without whom this synthesizer work would not be available. Also, software efforts by Don Lafferty are greatly appreciated. Input from Mario Bustamante and Troy Schaffer has been invaluable. Thanks also to the many others who have made this work possible.

References

[1] T.J. Endres and J.B. Kirkpatrick, "Sensitivity of Fast Settling PLLs to Differential Loop Filter Component Variations," Proceedings of the 1992 IEEE Frequency Control Symposium, May 1992

[2] T.J. Endres and G.T. Calvetti, "Induced End-of-Life Errors in a Fast Settling PLL," Proceedings of the 1993 IEEE Frequency Control Symposium, June 1993

[3] V. Reinhardt, K. Gould, K. McNab, and M. Bustamante, "A Short Survey of Frequency Synthesizer Techniques," Proceedings of the 40th Annual Frequency Control Symposium, May 1986

[4] H.T. Nicholas III and H. Samueli, "An Analysis of the Output Spectrum of Direct Digital Frequency Synthesizers in the Presence of Phase-Accumulator Truncation," Proceedings of the 41st Annual Frequency Control Symposium, 1987

[5] V.F. Kroupa, "Discrete Spurious Signals and Background Noise in Direct Digital Frequency Synthesizers," Proceedings of the 1993 IEEE Frequency Control Symposium, June 1993

[6] V.F. Kroupa, "Spectral Purity of Direct Digital Frequency Synthesizers," Proceedings of the 44th Annual Symposium on Frequency Control, May 1990

[7] V.S. Reinhardt, "Spur Reduction Techniques in Direct Digital Synthesizers," Proceedings of the 1993 IEEE Frequency Control Symposium, June 1993

[8] N.S. Wagner, "Computation of Frequency Response for Standard Filters," Hughes Aircraft Company Technical Internal Correspondence No. 2292/379, September 1967

[9] N.S. Wagner, "Some Notes on the Computation of Parameters of Transformed Filters," Hughes Aircraft Company Technical Internal Correspondence No. 2292/609, June 1968

[10] M. Shoga, K, Jobe, M. Glasgow, M. Bustamante, et al. "Single Event Upset at Gigahertz Frequencies," to be published at 1994 IEEE Nuclear and Space Radiation Effects Conference, July 1994

[11] W.H. Beyer, CRC Standard Mathematical Tables, 26th edition, Boca Rotan, FL: CRC Press, 1981

[12] V. Manassewitch, Frequency Synthesizers, New York: John Wiley and Sons, Inc., 1987

[13] F.M. Gardner, Phaselock Techniques, 2nd edition, New York: John Wiley and Sons, Inc., 1979

[14] R.M. Gagliardi, Introduction to Communications Engineering, New York: John Wiley and Sons, Inc., 1988

[15] W.C. Lindsey and M.K. Simon, Telecommunication Systems Engineering, Englewood Cliffs, N.J.: Prentice Hall, 1973

[16] M.K. Simon, S.M. Hinedi, and W.C. Lindsey, Digital Communication Techniques, U.S.C. EE564 and EE664 text, 1993 and 1994

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

FREQUENCY CONTROL REQUIREMENTS OF RADAR

Nathan Slawsby
ANRO Engineering, Inc.
450 Bedford Street
Lexington, MA 02173-1520

Abstract: The demands placed upon signal sources and other system components by today's radar systems often exceed the state of the art for these devices. This paper explores the operational conditions which give rise to these demands and illustrates some of the methods used to derive component specifications from them. The key role of radar clutter in generating these requirements is emphasized and illustrated by examples.

Introduction

Originally, radar was a simple device with the limited functions of "RAdio Detection And Ranging", from which the acronym "RADAR" is derived. In the early days, only rudimentary frequency control was needed, and this aspect of the design was, therefore, given little consideration by either the systems or the components engineers.

This is no longer the case. Radar has developed into an elaborate and sophisticated technology with many applications, some of which give rise to very stringent frequency control requirements indeed. Of necessity, frequency control technology, particularly in the area of stable, low-noise signal sources, has kept pace with these demands. At every stage of this development, the components industry has responded to the challenge, only to be faced with increasingly severe requirements for the next generation of systems. This is, perhaps, the inevitable result of success, since there are always incentives to improve upon existing capabilities. In military applications, present successes actually generate the requirement for improved performance in the future, since an adversary will always attempt to counteract an existing capability by increasing the sophistication of his equipment and tactics.

The purpose of this paper is to explore some of the phenomena which give rise to the requirements for frequency control in radars; to demonstrate the techniques used by systems engineers to generate the specifications for components affected by these requirements; and to derive a few typical specifications, based upon realistic system designs, which are near or beyond the current state of the art of operationally-useful components. In passing, some of the limitations imposed on radar performance by components other than those of concern here will be described.

The discussion will begin with a description of radar types and their general requirements for frequency control. A few of the basic relationships involved in radar calculations will then be presented, including, particularly, those defining received signal power and receiver thermal noise power density. From these, the effects of frequency fluctuations on performance will be derived, and the primary importance of clutter, as the mechanism which converts frequency fluctuations into signals which interfere with target detection, will be demonstrated. Various sources of frequency fluctuations, or, equivalently, signal phase and amplitude modulation, will be pointed out, and the equation for the cancellation for common-mode FM noise will be presented. Finally, numerical examples will be given for several radar configurations.

Definitions And Relationships

The definitions of several terms which appear throughout the discussion are presented here:

- Monostatic Radar: A radar in which the transmitter and the receiver are collocated, or in such proximity as to permit the use of common signal sources.
- Bistatic Radar: A radar system in which a transmitter acts an illumination source for one or more receivers which are quite distant from it. In general, no direct connection exists between the transmitter and the receivers.
- CW Radar: A radar which emits continuously. Echoes from all ranges arrive at the receiver simultaneously with such a system.

0-7803-1945-1/94 $3.00 © 1994 IEEE

- Pulse Radar: A radar which emits energy in pulses. In general, the receiver is active only in the intervals between the pulses. Echoes from some ranges may be received simultaneously.
- Doppler Frequency: The frequency shift between the transmitted and the received signals caused by the rate of change of the range to the target.
- Clutter: Received echo signals from any objects other than targets of interest.

The number of monostatic radars in use throughout the world far exceeds the number of bistatic radars, and pulse radars are much more common than CW radars.

The collocation of the transmitter and the receiver effectively eliminates long-term and even medium-term frequency fluctuations as problems in monostatic radars. Only variations which occur during the round-trip propagation delay time–that is, during the time required for the signal to propagate from the transmitter to the target and back to the receive –can be sensed by the system, and this time delay is seldom greater that a few milliseconds. Fluctuations occurring within such short time intervals are usually specified as power-density spectrum functions.

Any such function can be decomposed into an amplitude modulation (AM) and a frequency modulation (FM) component, and, since AM and FM produce different effects in the radar, it is customary to present component specifications in this form. The results of the examples given subsequently will, therefore, be given as descriptions or plots of the maximum permissible AM and FM noise spectra.

All bistatic radars are susceptible to medium-term and long-term frequency variations to a degree determined by the nature of their mission and the techniques which they employ. In some deployments, the transmitter and the receiver are so completely isolated that no direct communication between them is possible for periods of hours, days, or, in extreme cases, months. The frequency sources used in such an application must be sufficiently stable as to ensure that the receiver will cover the band occupied by radar echoes from the desired types of target. In some cases, this could require the use of sources employing atomic frequency references.

By far the most extensive application of bistatic radar to date has been in semi-active seekers for guided missiles. A radar transmitter at the missile launch site illuminates the target, reflected energy from which impinges on the seeker antenna. The receiver processes this to develop guidance commands for the missile. In this application, the receiver is, or can be, in direct contact with the transmitter until the time of launch. The time interval over which frequency drift can take place, therefore, is limited to the missile time of flight, which is usual on the order of a minute. Also, a direct line-of-sight path usually exists between the transmitter and the missile during the entire flight, permitting continuous synchronization of the two signal sources. Systems which make use of this option are said to be rear-referenced, while those which do not are said to use an on-board reference. Clearly, the latter method imposes the more severe long-term stability requirements.

The Continuous Wave (CW) radar, provides a clear illustration of the effect of range resolution on the clutter problem, since this type of radar effectively has no range resolution against clutter. Echoes from reflecting objects at all ranges return to the receiver continuously, and signals from targets at the maximum range of the radar and those generated by clutter objects at very short ranges are, therefore, received simultaneously. Since, as will be seen, signal strength for discrete reflectors varies inversely with the fourth power of range, this can produce clutter signals which are coincident in time with, and many orders of magnitude larger than the target signal. Clearly, it would be impossible to detect the target under these conditions unless some parameter other than time of arrival could be used to distinguish between the two signals.

Most, but not all, of the reflecting objects which produce the clutter echo are stationary. Since a stationary target is, from the radar point of view, in no way different from any of these objects, it is effectively just another clutter reflector, and its echo is just one of many which merge to form the clutter return. If the target has a radial velocity component, however, its echo signal undergoes a doppler shift. The difference in the carrier frequencies of the target and the clutter, then, provides a basis for resolving the two signals.

Filters are use in the receiver to select the target signal and reject the clutter. Because of the high input clutter-to-signal ratio, these filters must have a very high out-of-band rejection ratio. Since we are concerned with only those issues which affect frequency control requirements, it is assumed that the radar meets this and all other operational requirements, and that only those design elements relating to frequency control remain to be specified . Under this assumption, the clutter level at the output of all filters in the target frequency band would be below thermal noise if the AM and FM spectra of all frequency sources were ideal. Any real, non-ideal source will impress its power-density spectrum on the clutter signal, and some of the components of this spectrum will extend into the band occupied by the target filters. The specifications are derived by requiring that the level of these components be below receiver noise in all of these filters.

The key role of clutter in this process is evident. The power-density spectrum, as usually presented, specifies FM or AM noise power density per Hertz relative to carrier power (dBc/Hz), as a function of frequency offset from the carrier. When these spectra are impressed upon the clutter, the clutter signal itself becomes the carrier–that is, the power reference for the spectrum. If there is no clutter, this mechanism will produce no interference signal in the receiver no matter how noisy the frequency sources may be. A strong clutter environment, on the other hand, imposes severe AM and FM requirements. In general, to satisfy the criterion stated above, the spectra, at an offset from the carrier equal to the difference between the target and clutter frequencies, must be down from the carrier by the difference, in dB, between the peak clutter power and the receiver noise power density .

Pulse Doppler (PD) radars employ techniques similar to those used in CW radars to resolve targets from clutter. Since these are pulse radars, some elements of the problem associated with CW are eliminated. Others, unfortunately, are introduced. Examples of both types of radar are included in this paper.

Clutter returns can be produced by many types of reflectors; discrete and distributed, stationary and moving. Distributed clutter is produced by clusters of objects distributed over a region much larger than the spatial resolution of the radar. When the distributed-clutter region is two-dimensional, such as the surface of the earth, it is said to produce area clutter. A 3-dimensional distribution of reflectors, such as rain, fog, or chaff, produces volumetric clutter. Discrete clutter consists of returns from large, isolated man-made objects and from such natural objects as birds. As previously mentioned, most clutter sources are essentially stationary, although most terrain and ocean surfaces exhibit some velocity spreading due mainly to wind-induced motion. Some discrete clutter reflectors, however, such as moving vehicles and flocks of birds, have velocities high enough to pose significant difficulties for radars which use doppler frequency as a discriminant. The examples given in this paper consider only the relatively simple cases of discrete and volumetric clutter, which, besides illustrating the principles with minimum computational complexity, actually establish the frequency control requirements of many operational systems.

A number of equations used in the examples are given in Table 1. In most cases, the nomenclature and symbols used here follow the conventions of reference [1], which contains a full development of these and other radar relationships too complex to be discussed here.

Summarizing briefly, (1) is just the time delay experienced by electromagnetic energy in traversing a path of length $2R$ at the velocity of light, while (2) simply expresses the proposition that the phase of the RF signal returning to the receiver shifts by an additional 2π radians, or one cycle, whenever the path length changes by a wavelength.

Equation (3), which can be derived rather easily from first principles [1], is an expression for the power returned from a reflector of given cross section and range by a radar of given characteristics. The equation holds for clutter as well as targets of interest, since the distinction between the two is essentially subjective: in fact, it is often the case that on radar's target is another radar's clutter. Of particular interest in this expression is the range term in the denominator, which, as mentioned previously, magnifies the effects of clutter which originates at ranges much less than that of the target. It should also be noted that none of the terms are related to the frequency stability in any obvious way. Barring a catastrophic failure, the received signal power from both targets and clutter is essentially unaffected by frequency control considerations.

Equation (4) is the expression for the thermal noise power density produced by a resistor at temperature T_0, multiplied by a factor which quantifies the added noise power introduced by the receiver circuitry. Detection of the target echo in the presence of this interference is the basic problem of all radars, and it can be assumed that any system for which the effects of frequency fluctuations are being considered has been designed to solve this problem. In particular, the values of power and antenna gain have been set by this requirement.

The ratio of the received clutter power to N_0 is, as has already been mentioned, quite significant in establishing the FM and AM specifications for frequency sources and other components. As can be seen by dividing the units of signal power by those of noise power density, the ratio has the units of Hertz. When, as described above, this is multiplied by the FM and AM spectra of the frequency sources, which have units of inverse Hertz, the result is a dimensionless ratio of clutter-reflected noise to receiver thermal noise. If the fluctuations are random, this ratio can be used to establish specifications for the spectra regardless of the signal processing methods employed in the receiver, since the processing will have the same effect on both types of signal interference. Again, the assumption is that the tolerable level of increased interference is one which does not degrade the ideal clutter-free performance by more than some preestablished amount.

Systematic rather than random fluctuations produce narrow spikes in the spectrum which the signal processor may treat more like the signals than the noise. To specify allowable levels for these, it is usually necessary to consider the details of the processing. The examples given here consider only the random effects.

Table I Radar Equations

Time Delay	$t_d = \frac{2R}{c}$	(1)
Doppler Frequency*	$f_d = -\frac{2\dot{R}}{\lambda}$	(2)
Received Power**	$S_r = \frac{P_t G_t G_r \lambda^2 \sigma}{(4\pi)^3 R^4 L_t L_r L_\alpha}$	(3)
Receiver Noise	$N_0 = kT_0F$	(4)
Rain Backscatter Coefficient	$\eta_r = \frac{5.7\times10^{-14} r^{1.6}}{\lambda^4}$	(5)
Rain-Clutter Radar Cross Section	$\sigma_r = \eta_r \left(\frac{R_c \theta_b}{L_p}\right)^2 \Delta R$	(6)
Received Power From One Rain Cell	$C_r = \frac{11.3 P_t G_t \lambda^2 \eta_r \Delta R}{(4\pi)^3 R_c^2 L_p^2 L_t L_r L_\alpha}$	(7)
Common-Mode FM Noise Cancellation Factor	$A(f_d, R_c) = (2\sin 2\pi f_d R_c/c)^2$	(8)

*For bistatic radar, replace $2\dot{R}$ with $\dot{R}_t + \dot{R}_r$

**For bistatic radar, replace R^4 with $R_t^2 R_r^2$

t_d = 2-way propagation delay (s)

R = range to reflecting object (m)

c = velocity of light = $3\text{x}10^8$ m/s

f_d = doppler frequency (Hz)

$\dot{R}$ = range rate of change (m/s)

λ = wavelength of radar (m)

R_t = range from transmitter (m)

R_r = range from receiver (m)

S_r = received signal power (W)

P_t = peak transmitted power (W)

G_t = transmit antenna gain

G_r = receive antenna gain

σ = radar cross section of reflector (m^2)

L_t = transmitter RF loss

L_r = receiver RF loss

L_α = 2-way propagation path loss

N_0 = receiver noise power density (W/Hz)

k = Boltzmann's constant = $1.38\text{x}10^{-23}$

T_0 = ambient temperature =290°K

F = receiver noise figure

η_r = rain backscatter coefficient (m^2/m^3)

r = rain rate (mm/hr)

θ_b = radar antenna beamwidth (rad)

R_c = range to clutter cell (m)

L_p = beamshape factor = 1.33

ΔR = radar range resolution (m)

Equation (5) is taken from [1], where it is said to represent the backscatter properties of rain very accurately for frequencies in the normal radar bands. The wavelength term in the denominator indicates that rain clutter backscatter increases rapidly with radar frequency. The coefficient is used in (6) to compute the radar cross section of rain in the volume included within a range and angle resolution cell of a pencil-beam pulse radar. The other terms in the equation define the volume of this cell, as depicted in Figure 1. This shows the radar antenna main beam and two range resolution cells of width ΔR. The first of these cells is at a range R_{c1} from the radar and the second, at an incremental distance R_a farther away. This increment is the range ambiguity spacing of the waveform–that is, the range which, when inserted into (1), will give a 2-way propagation delay equal to the time between pulses. It can be seen that, with this spacing, the echo of the pulse most recently transmitted will return to the radar from the first

zone at the same time that the echo of the preceding pulse returns from the second. Since all transmitted pulses are identical, there is no way to distinguish between the two echoes, and the two range cells are truly ambiguous.

The figure also shows why the equation for the clutter volume of each region is of the form given by (6). Each resolution cell is essentially a cylinder of height ΔR and diameter $R_c\theta_b$. By inserting these values into the equation for the volume of a cylinder, the expression found in (6) can be derived, except for a constant factor which, along with a correction term to account for the gain variation across the beam, is included in L_p.

Equation (7) is obtained by substituting (6) and a relationship between antenna gain and beamwidth from [1] into equation (3). The significant effects of this are that one of the gain terms disappears and the square rather than the fourth power of the range appears in the denominator. Although the range dependence of received power is much less for rain than for discrete clutter, it is apparent that it still increases without limit as the range goes to zero. This is obviously a physically unrealizable condition, which indicates that the equations for gain and beamwidth no longer apply close to the antenna. Practically, this means that some minimum range must be set for all rain clutter computations.

The final equation in Table I embodies an effect with profound influence on the specification of frequency sources which provide signals to both the transmitter and the receiver, which is the meaning of the term "common-mode" as it is used here. The correlation factor multiplies the spectra of common-mode sources as they appear in the reflections from the clutter. It is periodic in the product of doppler frequency and clutter range and, significantly, varies as the square of the product as either of these goes to zero. Since the spectra of most sources of frequency fluctuation tend to rise sharply at low frequencies, and, as has been shown, the strength of the clutter echo varies inversely with some power of range, this factor is most effective where it is most needed.

The block diagram presented in Figure 2 illustrates the configuration of a monostatic radar employing the Master Oscillator Power Amplifier (MOPA) technique which most modern radars use. It also shows the frequencies of signals at various locations in the system, which can be used to demonstrate the common-mode cancellation process.

The Master Oscillator is the source of the RF signals supplied to both the transmitter and the receiver. As shown, it supplies a local oscillator input to the receiver mixer and to the single-sideband generator in the transmitter. As also shown, the frequency which it produces is not a pure sinewave, as it would be ideally, but contains a component, δf, which varies with time. This represents an undesired frequency fluctuation for which it is necessary to derive a specification.

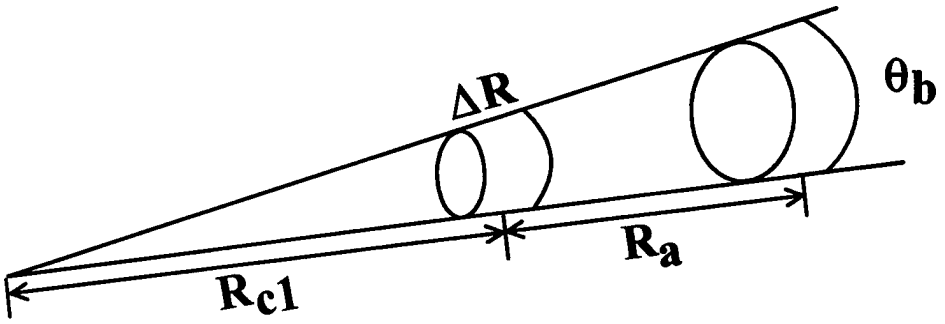


Figure 1 Rain Clutter Geometry

The Master Oscillator input to the single-sideband generator is mixed with a signal at IF, which could be a pure sinewave or might be a phase-encoded pulse-compression waveform. The signal which results from this mixing action is amplified and pulse-modulated, if required, to produce the transmitted signal. The frequency of this signal is the sum of the Master Oscillator and IF frequencies and, it should be noted, contains the time-varying frequency component originally generated by the Master Oscillator.

The radar depicted here uses the same antenna for transmission and reception, as do all pulse radars. The signal from the transmitter power amplifier passes through a circulator which directs it toward the antenna and away from the receiver, is radiated by the antenna, and impinges on clutter at range R_c. After a round-trip time delay of $2R_c/c$, the clutter echo returns to the antenna, where it is directed to the receiver mixer by the circulator. As shown, this signal retains all of the frequency components of the transmitted waveform, but the time delay of the propagation path is appears explicitly in the δf term. The local-

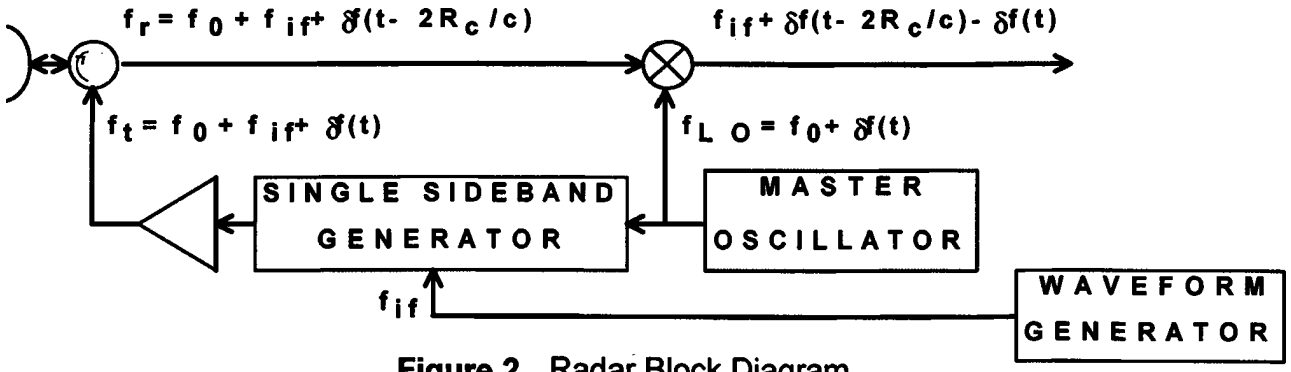


Figure 2 Radar Block Diagram

oscillator input to the mixer, therefore, contains the current value of δf, but the signal input contains the value of δf from some time in the past.

The frequency which appears at the mixer output port is the instantaneous difference between the two input frequencies. The non-fluctuating term of the Master Oscillator frequency vanishes in this operation, leaving only the IF which originated in the waveform generator and the difference between the delayed and the undelayed fluctuation term. It is apparent from the equation in the diagram that, when R_c is zero, the latter two terms are identical and the frequency fluctuation is cancelled. This is the effect noted previously in the discussion of equation (8). The equation itself can be developed rather easily by treating the phase rather than the frequency of a single FM component of the fluctuation.

In considering the diagram, it should be mentioned that there are a number of other possible sources of clutter-reflected noise, and not all of these are subject to common-mode cancellation. For example, it is apparent that any phase or amplitude modulation of the signal which takes place in the transmitter power amplifier will be present in the clutter echo, and there will be no corresponding term in the local oscillator signal to cancel it. Such amplifiers generate noise internally and also convert variations in power supply voltages and modulator waveforms into signal fluctuations. The waveform generator is another potential source of problems, as are any other local oscillator signals, especially if they are produced by voltage-controlled oscillators, which may be present in the system. Pulse timing jitter in the waveform modulator and in the sampling pulses of A-to-D converters produce similar effects. All of these must be included in an overall error budget for the system, and a permissible level for each must be specified. The examples that follow are somewhat unrealistic in that this procedure has not been followed in them. Rather, the entire allowable system degredation has been assigned each source as it is discussed.

Example 1: CW Radar

We consider first an X-Band CW radar with the following requirements:

- Elevation scan down to the horizon
- Minimum target velocity = 30 m/s
- Maximum target velocity = 1000 m/s
- Discrete clutter at 100 m range, 25 m^2 cross section

The problem is to specify allowable common-mode and independent FM and AM spectra such that the interference produced by the presence in the antenna beam of one of the clutter objects described above will not exceed receiver noise.

The parameters of this radar are as follows:

$P_t = 500$ W $\quad G_t = G_r = 35$ dB $= 3162$

$R = 100$ m $\quad L_t = L_r = 1.5$ dB $= 1.412$

$\lambda = 0.03$ m $\quad F = 6$ dB $= 3.981 \quad L_\alpha = 1$

$\sigma = 25\ m^2$ $\quad kT_0 = 4 \times 10^{-21}$

Inserting these parameters into the equations in Table I:

Received Clutter Power–Eq. (3)

$$C = S_r = \frac{(500)(3162)(3162)(0.03)^2(25)}{(4\pi)^3(100)^4(1.412)(1.412)(1)} = 2.843 \times 10^{-4}\ \text{W}$$

Receiver Noise Power Density–Eq. (4)

$$N_0 = (4 \times 10^{-21})(3.981) = 1.592 \times 10^{-20}\ \text{W/Hz}$$

Clutter-to-Noise-Power-Density Ratio

$$\frac{C}{N_0} = \frac{2.843 \times 10^{-4}}{1.592 \times 10^{-20}} = 1.786 \times 10^{16}\ \text{Hz} = 162.5\ \text{dB}$$

Target Doppler Frequencies–Eq. (2)

$$(f_d)_{min} = \frac{(2)(30)}{0.03} = 2000 \text{ Hz} = 2 \text{ kHz}$$

$$(f_d)_{max} = \frac{(2)(1000)}{0.03} = 66670 \text{ Hz} = 66.67 \text{ kHz}$$

Common-Mode Cancellation Factors–Eq. (8)

Minimum Doppler

$$A(2000, 100) = \left\{2 \sin\left[\frac{(2\pi)(2000)(100)}{3 \times 10^8}\right]\right\}^2 = 7.02 \times 10^{-5}$$

$$= -41.5 \text{ dB}$$

Maximum Doppler

$$A(66670, 100) = \left\{2 \sin\left[\frac{(2\pi)(66670)(100)}{3 \times 10^8}\right]\right\}^2 = 0.07748$$

$$= -11.1 \text{ dB}$$

The required specifications can be calculated from the results given above. For the common-mode FM noise, we have:

$W(f_d) = -[(C/N_0)_{dB} + A(f_d, R_c)]$ Specified Common-Mode Power Density Ratio in dBc/Hz

$W(2000) = -[162.5 - 41.5] = -121$ dBc/Hz

$W(66670) = -[162.5 - 11.1] = -151$ dBc/Hz

Since the argument of the sine term in the cancellation factor is small at every frequency of interest, the slope of the power density ratio is 6 dB per octave.

Since the independent noise sources do not get the benefit of the common-mode cancellation factor, their combined effect is just the negative, in dB, of the clutter-power to noise-power-density ratio. That is:

$U(f_d) = -162.5$ dBc/Hz Specified Power Density Ratio for Independent Sources at all Frequencies Between 2 kHz and 66.7 kHz

The specifications derived above are very challenging, but CW radars have been designed to meet such such requirements since the early 1960s. The Master Oscillators for these radars employed cavity-stabilized reflex klystrons, which are still among the least noisy frequency sources. Unfortunately, both they and the CW technique itself are unsuited to any application where mechanical vibration is a significant factor.

The obvious disadvantages of monostatic CW radar are that it must receive while it is transmitting, and this can only be done by using separate transmit and receive antennas and rather exotic feedthrough cancellation techniques. There are, however, a number of advantages for continuous rather than pulsed transmission. One important advantage is that the CW waveform is unambiguous in doppler, while the pulse waveform is ambiguous at integer multiples of the Pulse Repetition Frequency (PRF). The FM and AM noise spectra for CW, therefore, need be specified within the band occupied by the targets of interest, while pulse radar specitra must be controlled far beyond this region. Also, many of the noise sources, such as pulse timing jitter, which must be considered in pulse radars are absent in CW systems. It is also easier to degenerate the noise contributions of the CW transmitter amplifier.

In passing, it is appropriate to mention here the specifications for the doppler filters used in the receiver to reject the clutter signal. Although this is not strictly a frequency-control issue, the clutter power which feeds through these filters could easily be the dominant source of interference in the system. The clutter-to-noise ratio which determines this specification is found by dividing the clutter-to-noise-density by the filter bandwidth. In this example, the clutter-to-noise ratio is 162.5 dBHz. For a filter bandwidth of 1 kHz, we subtract 30 dB from this, to give 132.5 dB, which would be the clutter-to-noise ratio at the filter output if there were no isolation . To establish a unity clutter-to-noise ratio at the filter output, therefore, the rejection ratio must be -132.5 dB 2 kHz from the filter center frequency. This performance is obtained with mutipole crystal filters in existing radars.

Example 2: Pulse Doppler Missile Seeker

This is an example of a the type of requirements generated by current radar applications. It considers the problem posed by a monostatic, K_a-Band, High PRF pulse doppler radar used to generate guidance commands for a missile. The seeker must do this while flying through rain. The operational requirements are as follows:

- Minimum target velocity = 45 m/s
- Maximum target velocity = 1300 m/s
- Rainfall rate = 4 mm/hr at all ranges

The problem here is to specify the allowable FM noise power-density spectrum of the Master Oscillator such that the interfernce produced by rain in the first range resolution cell following the transmitted pulse will not exceed receiver noise at any target doppler frequency.

The parameters of this radar are as follows:

$P_t = 1000$ W	$G_t = 33$ dB $= 1995$	$L_\alpha = 1.7$ dB/km
$R_{c1} = 49$ m	$L_t = L_r = 3$ dB $= 2$	$L_p = 1.33$
$\lambda = 0.0086$ m	$F = 5.5$ dB $= 3.55$	$R_a = 135$ m
$\Delta R = 37$ m	$kT_0 = 4 \times 10^{-21}$	PRF = 1.11 MHz

It is not possible to approach this problem in the straightforward manner of Example 1, since there actually is no unique solution. Rather, it is necessary to assume a generic shape for the FM noise power-density spectrum, insert an arbitrary set of parameters into the generic form, and compute the interference signal at frequencies spanning the target doppler band. A few iterations of this procedure will usually be sufficient to realize the desired result.

The generic form of the spectrum shown in Figure 3 was used in this example. It is characterized by an inverse third power slope at low frequencies and a constant value at high frequencies. The numerical values associated with the curve in the figure were obtained by the iterated calculations described above.

The calculations themselves involve a rather laborious procedure. The clutter power from the first range ambiguity, and, indeed, from each of the other range ambiguities, must be computed by substituting the radar parameters into equation (7), and the common-mode cancellation factors for each of these at each doppler frequency must then be determined from (8). The effects of doppler ambiguities, spaced at integer multiples of 1.11 MHz must then be added

In working this example, this was done by writing a short program in the Mathcad language, which produced the results shown in Figure 4 for the Master Oscillator spectrum of figure 3. Since this does, indeed, provide the required performance in a 4 mm/hr rain, and it can be used to specify the device.

It can be seen that the specification is quite stringent, especially since it is imposed at K_a Band. Also, the unit must conform to the severe weight, volume, and vibration requirements of the missle. It should also be noted that, as in the first example, the entire fluctuation noise budget has been allocated to this one source, which almost certainly can not be done in this case. For example, it has been calculated that a random jitter of 20 ps rms in the leading and trailing edges of the transmitted pulse would also consume the entire allowance. The high-gain power amplifier, multivibrator power supply, and A-to-D converter used in this type of system are also likely sources of fluctuation noise.

Reference

D.K. Barton, *Modern Radar System Analysis*, Artech House, 1988

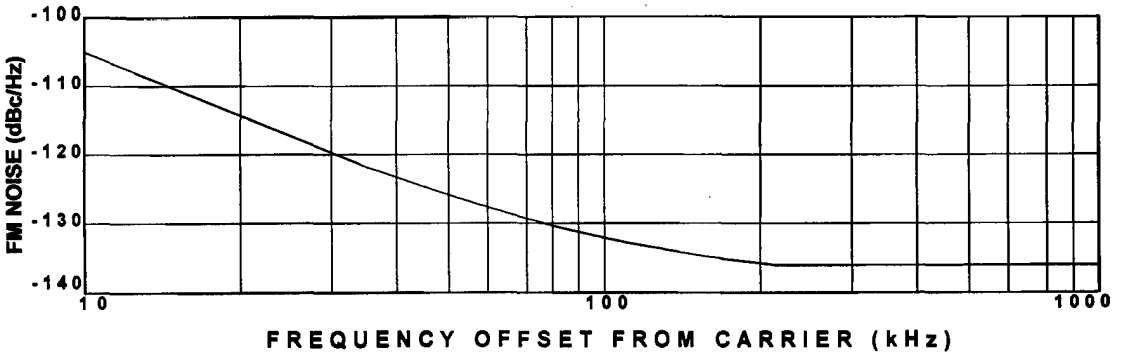


Figure 3 Assumed Master Oscillator FM Noise

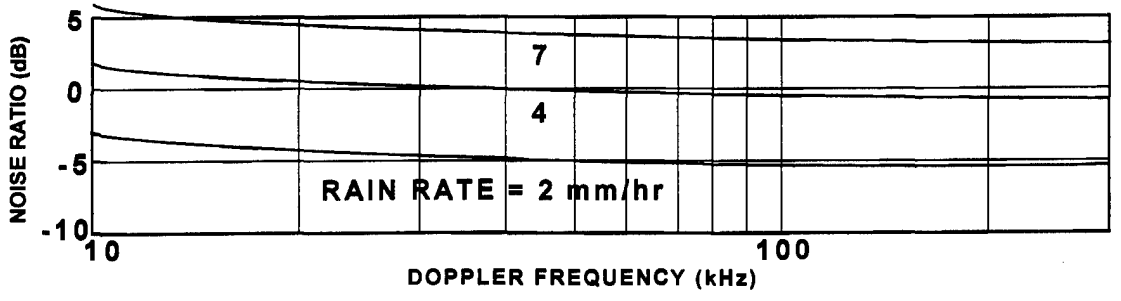


Figure 4 Ratio of Clutter-Reflected to Receiver Noise at IF

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

FREQUENCY CONTROL DEVICES APPLICABLE TO RADAR SYSTEMS

Stephen P. Caldwell

Westinghouse Electric Corporation
Electronic Systems Group
Baltimore, Md. 21203

ABSTRACT

An overview of frequency control devices used in radar systems is presented. These devices include frequency sources for master oscillators, frequency synthesizers for multiple frequency operation, and waveform generators for providing frequency or phase modulations on the radar transmit signal. The frequency control devices are presented from the perspective of their relation to each other, the application in the radar system, and their effect on radar system performance.

INTRODUCTION

Radar systems utilize frequency control devices of different types to achieve levels of performance in the basic radar measurements. These measurements are range, range rate (velocity), and angular position. Frequency control devices can affect accuracies and resolutions for any or all of these measurements. The effects and magnitude of the errors depend strongly on the frequency control components used and their application in the radar system. In addition, the radar system often is used in an electromagnetic environment where the signal of interest to the radar is much smaller than other signals entering the radar's antenna. These large signals, called clutter, impose difficult phase-noise and spurious (discretes) requirements on radar signal sources.

The applications of radar are quite varied[1] and so are the resulting implementations. As a result, a discussion of the effects of frequency control devices in a radar system will be somewhat subjective. This paper will use a block diagram of a coherent pulse doppler monostatic (one whose transmitter and receiver are collocated) radar for purposes of discussing relevant frequency control devices. A block diagram of analog hardware suitable for coherent radar operation is shown in Figure 1. The block diagram shows a master oscillator whose output is used to generate CW reference frequencies. These are then used to provide coherent references for system synchronization, frequency synthesis, and waveform generation functions. The frequency synthesizer and waveform generator outputs are combined to provide a transmit excitation to a power amplifier prior to application to an antenna. The synthesizer is also used as the first local oscillator (LO) for a triple down-conversion receiver. The second and third LOs are fixed frequency. The portions of the radar hardware which utilize frequency control devices are shown shaded in the figure. They will be combined into the following categories for the discussion to follow: master oscillator, frequency synthesis, and waveform generation.

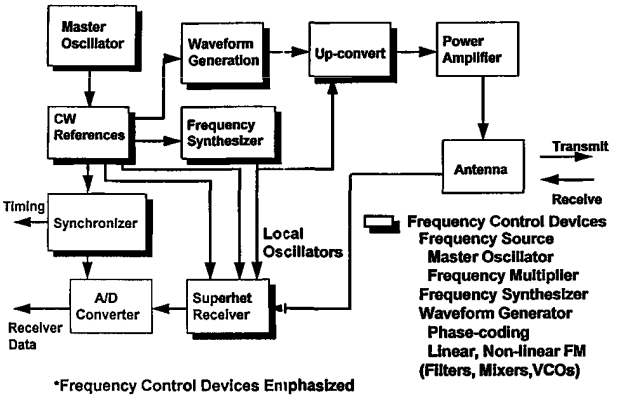


Figure 1
Coherent Radar: Analog Hardware Block Diagram
Showing Frequency Control Hardware

Master Oscillator

The master oscillator provides a stable signal from which all other RF signals in the radar are coherently derived. As Figure 1 shows, the output of the master oscillator is used to generate CW reference frequencies which are, in turn, used by waveform generation hardware for modulating the transmit signal and by frequency synthesis hardware for generating a frequency agile L01 and transmit signals. A L02 and L03 to the receiver and a clock to synchronizer for generating timing are also generated from the reference frequencies.

Master oscillator requirements are derived from radar requirements for clutter rejection, dynamic range, range and velocity accuracy. The radar operating frequency also

0-7803-1945-1/94 $3.00 © 1994 IEEE

affects the choice of master oscillator. The relation between radar requirements and corresponding master oscillator requirements are shown in Table 1 along with some applicable low phase-noise oscillator technologies.

The basis measurements of range and velocity by a radar are related to master oscillator frequency errors as follows:

$$\text{for range } \frac{dr}{R}=\frac{df}{f} \qquad (1)$$
$$\text{for velocity } \frac{dv}{v}=\frac{df}{f}$$

In (1), R is the range from the radar to a target, df/f is the fractional frequency accuracy of the master oscillator and v is the target doppler frequency as observed by the radar. These relations are derived from the basis relations for radar range, R=c*T/2, and velocity, v=c*fd/(2fo); where T is the round trip time to a target, c is the speed of light, fd is the target doppler frequency and fo is the radar operating frequency. Since most master oscillator technologies can easily exceed frequency accuracies of le-5 (.001%), their contribution to radar measurements errors are usually not significant. Some exceptions are bi-static radars and radars for radio astronomy.

Table 1
Radar Master Oscillators - Radar requirement, Oscillator Requirement, and Oscillator Technology

Radar requirement	Oscillator requirement	Oscillator technologies
Clutter rejection	Phase-noise	Bulk crystal
	Spurious	XO VCXO TCXO
Range accuracy	Frequency accuracy	SAW
Velocity accuracy	Frequency accuracy	Cooled resonator
Output frequency	Frequency	Atomic clocks

Some simplified radar, frequency, phase-noise and spurious relations:

dR/R=df/f
dv/v=df/f
L(f)=D*tau/2
Spurs=D/(2*Nint)

R=range v=velocity
df/f=fractional frequency
L(f)=SSB phase-noise
D=receiver dynamic range
tau=pulsewidth
Nint=# pulses integrated

Dynamic range requirements for a coherent doppler radar result in phase-noise and spurious requirements for the master oscillator (and other frequency control devices as well). The radar dynamic range requirement is derived by considering the largest (usually clutter) and smallest targets which the radar must simultaneously receive. A simple derivation of a radar phase-noise requirement can be obtained by defining receiver dynamic range as the signal to noise ratio at the receiver output for a signal at the top of the receiver's linear range. The resulting single-sideband CW phase noise L(f) is related to the dynamic range D by

$$L(f)<\frac{D\tau}{2} \qquad (2)$$

where τ is the radar pulsewidth. The relationship is derived as follows: Phase-noise power is 2*L(F)*B, where B is receiver noise bandwidth. Phase-noise power should be much less than receiver noise power so 2*L(F)*B<<D, or L(F) << D/(2*B). The noise bandwidth B of a matched filter for a rectangular pulse is $1/\tau$, which results in (2). Table 2 shows radar phase-noise as a function of receiver dynamic range and pulsewidth. For combinations of high dynamic range and short pulses, very low phase-noises are needed. An important assumption in (2) is that the phase-noise is white phase-noise over the bandwidth of interest.

Table 2
Radar phase-noise requirement as a function of Receiver Dynamic Range and Pulsewidth

Phase-noise

Pulse-width	Dynamic Range				
	60 dB	70 dB	80 dB	90 dB	100 dB
.1 usec	<133 *	<143	<153	<163	<173
1 usec	<123	<133	<143	<153	<163
10 usec	<113	<123	<133	<143	<153
100 usec	<103	<113	<123	<133	<143

L(f)=D*tau/2 * L(f) [dBc/Hz]

The master oscillator phase-noise is derived by considering its frequency in relation to the radar frequency and any allocations of noise to other system components such as waveform generation and frequency synthesis. In general, a master oscillator whose frequency is less than the radar frequency will have its phase-noise degraded at the radar output by

$$Lx(f)=N^2*Lmo(f)*|H(f)|^2 \qquad (3)$$

where Lx(f) is the transmit phase-noise, N is the ratio of the transmit frequency to the master oscillator frequency, Lmo(f) is the master oscillator phase-noise and H(f) is a filter function to account for any narrowband filtering (a phased-locked loop for example), if there is any. For example, a radar at 10 GHz and a master oscillator at 100 MHz, the degradation is 40 dB, assuming no narrowband filtering. The master oscillator noise is then

$$Lmo(F)<Lx(f)/(N^2*|H(f)|^2 \qquad (4)$$

A comparison of low phase-noise master oscillator technologies is shown in Figure 2. The master oscillator technologies are a bulk crystal oscillator[2,3] at 80 MHz, a Surface Acoustic Wave (SAW) oscillator at 500 MHz[4] and a cooled sapphire resonator oscillator at 13 GHz[5]. The figure shows they all have very low phase-noise floors (-165 to -178 dBc/Hz). Furthermore the low frequency noise of the 80 MHz source is much better (by about 25 dB) than the other two. Consider the application of these sources to a 13 GHz radar. Using (3), assuming perfect frequency multipliers, and no narrowband filtering, the curves shown in Figure 3 are calculated and plotted. Now the cooled sapphire oscillator has the best phase-noise at

all frequencies. The crystal has lower noise than the SAW for frequencies less than about 6 kHz.

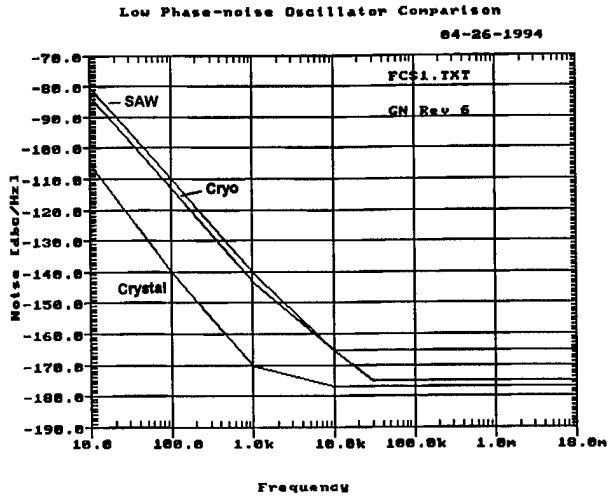


Figure 2
Phase-noise for 80 MHz crystal, 500 MHz SAW, and 13 GHz Cryogenic Sapphire resonators oscillators

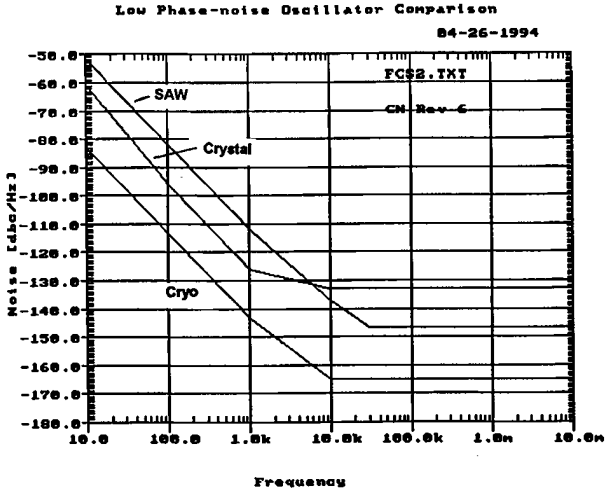


Figure 3
Oscillator Phase-noise from Figure 2 referred to 13 GHZ

The CW phase-noise spectrum can be altered by the radar as it processes the signal and noise. Phase-noise in the radar which is common to transmit and receive paths (the master oscillator and frequency synthesizer in Figure 1, for example) are correlated when they are observed at the receiver output. These noises are called common phase-noise. A phase-noise not common is called independent phase-noise. A source of common phase-noise is described by the function[6]

$$Lr(f) \equiv 4\sin^2(\pi * f * trt) * L(f) \quad (5)$$

Where Lr(f) is the noise as seen by the radar at the receiver output and trt is the round trip time to the largest clutter return. This function can accentuate noise by up to 6 dB or eliminate noise for frequencies greater than 1/trt. For frequencies < 1/(4*trt), noise is attenuated by 20 dB/decade. This is a significant factor to consider for synthetic aperture radar where noise requirements can extend to less than 1 Hz or for short range radars.

Since phase-noise can originate from multiple sources in the radar, a means for determining their contributions is necessary. This can be done by combing sources as common noise or independent noise and adding them as shown in the following equation

$$L_r(f) = \sum_n L_{i_n} + 4 \sum_m \sin^2(\pi * f * Trt) * Lc_m(f) \quad (6)$$

where the total radar noise Lr(f) is the sum of the independent noises Li(f) and common noises Lc(f) as modified by (5). The noises Li(f) and Lc(f) can then be adjusted depending on available technology to achieve the desired Lr(f).

A final consideration for phase-noise in the radar is the effect of the radars pulsed operation. The process of pulsing or gating a signal for transmission and then receiving those signals is analogous to a sampled system, where the sampling rate is the radar pulse repetition frequency (PRF). Since the bandwidth of the transmitted noise is much greater than the PRF, there can be a lot of aliasing of the noise components which can alter the spectrum of the noise appreciably from the CW noise spectrum. The resulting spectrum is sometimes called gated noise or folded noise or pulsed noise. Mathematically, the resulting spectrum is the convolution of the signal spectrum (sin(x)/x for a rectangular pulse) with the noise spectrum. The amount of distortion usually increases inversely with duty cycle (low duty cycle give more distortion).

As an example, the gated noise for the three sources in Figure 3 were computed for 3 different radar PRFs (low PRF, medium PRF, and high PRF). The pulsewidth is the same in each case so that the duty cycle is lowest in the low PRF case. The noise was assumed common for all the cases so an approximation to (5) was used; the round trip time was assumed to be 250 usec. Figures 4a, 4b, and 4c show the resulting spectrums. The plots show that the radar will alter the CW noise spectrum and the effects of non-white noise sources (real-world oscillators) in the radar depend on the source noise and its origin as well as the radar mode of operation. Although lower noise is always better, the improvements are not always dB for dB.

Vibration sensitivity of the master oscillator is an important consideration. Most low-phase noise technologies have a sensitivity to vibration, usually specified as a fractional frequency deviation per acceleration constant. The equation relating the vibration spectral density to the phase-noise spectral density is

$$S_g(f) = 2 * f^2 * L(f) / (\Gamma * fo)^2 \qquad (7)$$

where Sg(f) is the vibration spectral density in g^2/Hz, Γ is the vibration sensitivity constant, and fo is the radar transmit frequency. Using the crystal oscillator curve is the example in Figure 3, assuming a Γ of .5e-9/g and 13 GHz for fo, Sg(f) is computed and shown in figure 5. The figure shows the very low levels of vibrations that are needed for low phase noise oscillators, usually indicating the need for some type of mechanical vibration isolation.

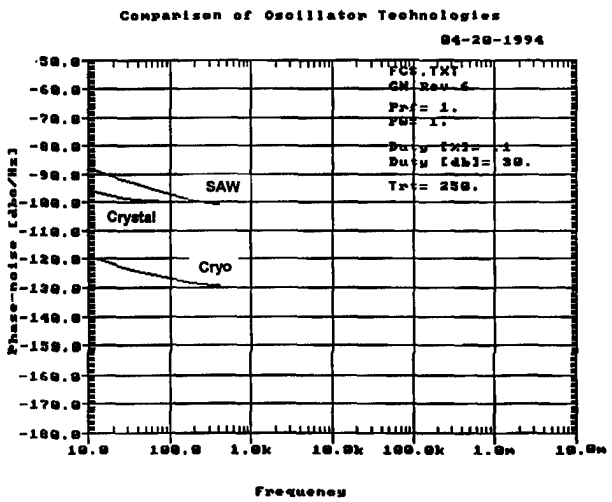


Figure 4a
Gated Noises for Low PRF (1 KHz)

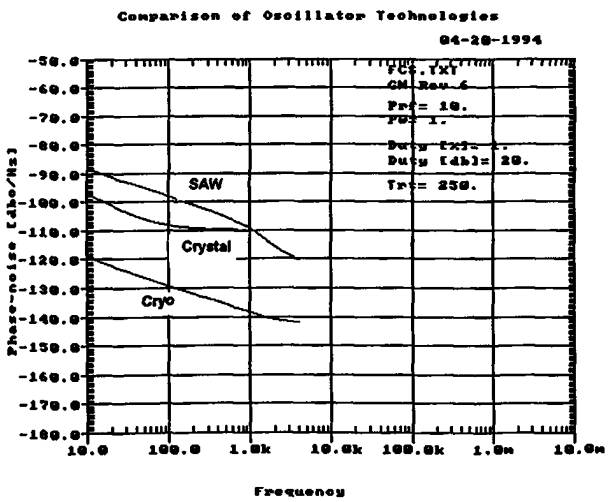


Figure 4b
Oscillator Gated Noises for medium PRF (10 kHz)

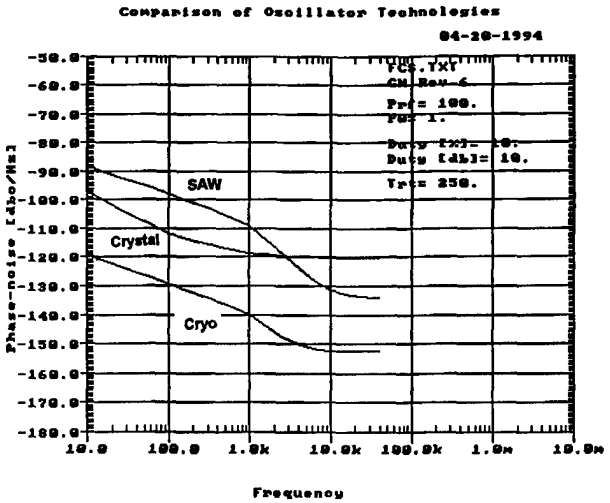


Figure 4c
Oscillator Gated Noises for high PRF (100 kHz)

The spurious or discrete modulations required by the radar can be derived from the phase-noise requirements in (2). A pulse doppler radar integrates a number of pulses to provide frequency resolution and increase dynamic range. The improvement in dynamic range increases in proportion to the number of pulses integrated. The spurious requirement is therefore approximately

$$Spurious < \frac{D}{Nint} \qquad (8)$$

where Nint is the number of pulses integrated. Table 3 for spurious is constructed from (8) similar to Table 2 for noise.

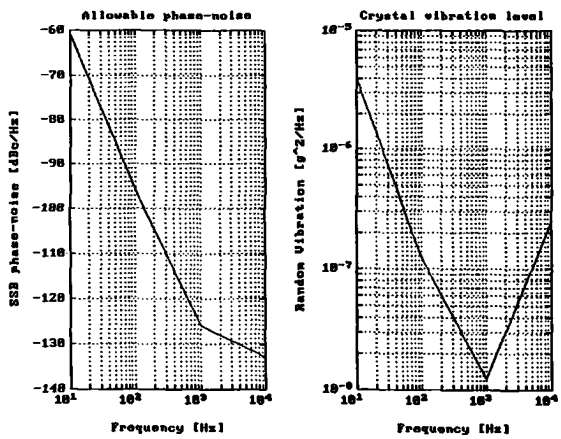


Figure 5
Crystal Vibration levels which produce phase-noise equal to non-vibrating oscillator. For 80 MHz oscillator multiplied to 13 GHz.

Table 3
Radar spurious requirement as a function of receiver dynamic range and pulses integrated

Spurious	Pulses Integrated	Dynamic Range				
		60 dB	70 dB	80 dB	90 dB	100 dB
	10	<70 dBc	<80	<90	<100	<110
	100	<80	<90	<100	<110	<120
	1000	<90	<100	<110	<120	<130

Spurs=D/Nint

Frequency Synthesis

The radar benefits from frequency agility in a number of ways[7]: interference from other emitters is reduced, undesirable propagation effects are reduced and target statistics are improved. These benefits result in improved target detection capability.

The requirements for radar frequency agility and the relation to synthesizers are shown in table 4 along with a list of frequency synthesizer techniques. The basic synthesis categories are indirect synthesis which use phase-locked loops[8], direct analog systhesis[9] which use mixers, filters, and frequency dividers, and direct digital synthesizers which generate digital samples of a signal and then convert to analog[10]. In practice, frequency synthesizer designs can combine elements of the different types. Because of this and since other radar signal generation requirements (waveforms for example) can influence the design of the frequency synthesizer, general comparisons of frequency synthesizers for radars can be difficult. Nevertheless, a subjective comparison by the author for the various synthesizer types is given in table 5.

Table 4
Radar Frequency Synthesizers - Radar Requirement, Synthesizer Requirement, and Synthesizer Technology

Radar requirement	Synthesizer requirement	Synthesizer technology
Clutter rejection	Phase-noise (additive or residual)	Indirect (PLL) Multi-loop Fractional-N Direct analog (mix and divide) Hybrid Direct digital (DDS)
Dynamic range	Spurious	
Mutual interference Multi-path Target statistics	Bandwidth Number of channels	
Timeline Waveforms	Switching time Phase memory	

Table 5
Comparison of Radar Frequency Synthesizer Technology

	Indirect	Direct	Hybrid	DDS
Phase-noise	limited by divide ratio	very good	limited by VCO loop bandwidth	technology GaAs/Silicon
Spurious	limited by divide ratios, loop bandwidth, filters	good close-in filters-far out	loop bandwidth, filters	DAC, sine quantization
Switching time	limited by loop bandwidth	fast, limited by filters	Can do better than indirect	fast, limited by clock
Bandwidth, No. of channels	VCO, divider	wideband comb generators, filters	VCO	clock speed
Size	small	big	in-between	small
Cost	inexpensive	expensive	moderately expensive	moderately inexpensive
Power	low	high	moderately high	moderately low

Waveform Generation

Waveforms transmitted by the radar can be phase or frequency modulated within a pulse to obtain better range resolution than would otherwise be possible. The basis relation for range resolution is

$$dr = c * dt/2 = c/(2 * B) \qquad (9)$$

where dr is the radar range resolution, c = speed of light, dt = phase code resolution, and B = waveform bandwidth. Equation 9 shows that range resolution can be increased by transmitting short phase codes or wide bandwidths within a pulse. As an example of (9), a 10 foot range resolution can be obtained by transmitting a phase-code with 20 nsec code transitions or a linear FM signal with 50 MHz bandwidth. Reference 11 gives much detail on some of the many possible waveforms. Phase and amplitude errors across the pulse must be kept low so that false targets or time sidelobes are not generated. In the case of linear FM, figure 6 shows that excellent control of amplitude and phase is required to keep these errors small. For example, a 35 dBc time sidelobe requires phase to be less than 2 degrees peak or amplitude ripple to be less than .30 db peak across the bandwidth of the FM waveform.

Another waveform, narrowband linear FM (less than 100 kHz in bandwidth), can be used to remove range ambiguities in high PRF radar. For this waveform, the duration of the sweep is over many transmitted pulses.

Table 6 describes the various waveform requirements and lists some of the waveform generation techniques employed. The methods for generating waveforms are even more diverse than frequency synthesis. A trend toward digital generation of waveforms[12] may reduce the types of waveform generators in the future, if the spurious in digital waveform generators are acceptable in the application.

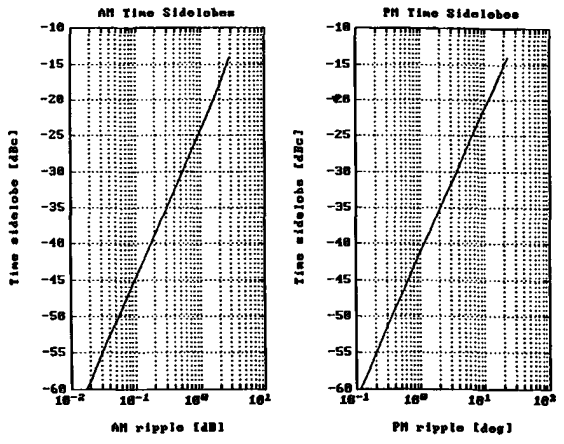


Figure 6
Amplitude or phase ripple required to achieve a given time sidelobe for a linear FM waveform

Table 6
Radar Waveform Generators - Radar Requirement, Waveform Generator Requirement, Waveform Generator Technology

Radar Requirement	Waveform Generator Requirement	Waveform Generator Technology
Range resolution dr=c*dt/2=c/(2*B) dr=range resolution dt=chip width B=bandwidth c=speed of light	Phase modulation Clock rate Number of code states Phase accuracy	Microwave PIN diode Phase shifters Microwave schottky diode biphase modulator Microwave I/Q modulators
	Frequency modulations Linear FM, non-linear FM Bandwidth Time Describing function	Waveform digital synthesis Accumulator Memory Linear VCXO Linearized (active feedback) VCO
	Phase and amplitude linearity	SAW dispersive delay line

SUMMARY

An overview of various frequency control devices utilized in modern radars has been presented. Parameters relating radar system performance and frequency control devices were emphasized.

REFERENCES

[1] Merrill Skolnik, Radar Applications, IEEE Press, 1987.

[2] 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", Proceedings of the 1992 IEEE Frequency Control Symposium, pp. 334-338.

[3] John R. Vig, "Quartz Crystal Resonators and Oscillators," Frequency Control Symposium, Tutorial notes, 1991.

[4] G. Montress, T. Parker, and M. Loboda, "Extremely Low Phase-noise SAW Resonator Oscillator design and Performance," Proceedings of the 1987 IEEE Ultrasonics Symposium, pp. 47-52.

[5] C. Flory and R. Taber, "Microwave Oscillators Incorporating Cryogenic Sapphire Dielectric Resonators," Proceedings of the 1993 IEEE Frequency Control Symposium, pp. 763-773.

[6] Merrill Skolnik, Radar Handbook, 2nd edition, McGraw Hill, 1990, p. 3.13.

[7] David K. Barton, Frequency Agility and Diversity, RADARS, vol. 6, Artech House, 1977.

[8] U. Rhode and T. Bucher, Communication Receivers, McGraw Hill, 1988, pp. 277-317.

[9] R. Stirling, Microwave Frequency Synthesizers, Prentice Hall, 1987.

[10] V. Reinhardt, K. H. McNab, W. Iwata, K. V. Gould, "Frequency Synthesizer Basics," 1992 Frequency Control Symposium, Tutorial Notes.

[11] Cooke and Bernfield, Radar Signals-An Introduction to Theory and Application, Academic Press, 1967.

[12] Caldwell, Driscoll, Foley, Haynes, Merrell. "Recent Advances in Exciter and Waveform Generation Technology at Westinghouse," Record of the 1993 IEEE National Radar Conference, pp. 40-45.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DESIGN AND PERFORMANCE OF AN ULTRA-LOW PHASE NOISE, RADAR EXCITER

M. M. Driscoll, A. C. Hazzard, and D. G. Opdycke

Westinghouse Electronics Systems Group
P.O. Box 746, Baltimore, MD 21203

Abstract

This paper reports on the design and performance of a radar exciter exhibiting state-of-the-art phase noise characteristics. The exciter provides fully coherent receiver local oscillator signals at HF, L-band, and X-band as well as requisite, auxiliary VHF and HF clock signals. Output signal spectral performance was achieved via use of a variety of recently-developed circuit technologies and frequency synthesis techniques. Included in these are use of multiple quartz crystals in the Master Oscillator, cascaded stages of narrowband, UHF SAW resonator filters, regenerative frequency division, and low noise, diode doubler and step recovery diode frequency multiplication. For the X-band outputs, the measured phase noise performance is characterized by £(100Hz)=-101dBc/Hz, £(1KHz)=-128dBc/Hz, and £(100KHz)=-145dBc/Hz. Exciter output signal spectral performance in the presence of modest levels of mechanical vibration is also presented. Included in the vibration data are results of SAW resonator vibration sensitivity measurements for several different SAWR die designs. The data indicates large variations in SAWR vibration sensitivity values as well as spectral peaking in the fm=7KHz carrier offset frequency region that is associated with SAWR mechanical mount resonances excited by environmental (sound level) stress and can be suppressed via acoustic shielding.

Introduction

Requirements for improved radar system performance, in terms of detection of small cross section targets in a high clutter environment, necessitate the achievement of a high degree of spectral purity in the microwave transmitter and receiver local oscillator signals generated in the radar exciter. Exciter signal phase noise requirements depend on the system application and on performance constraints in non-exciter hardware such as the transmitter and receiver, and they are also affected by system design considerations such as transmit signal waveform characteristics. In general, requirements for extremely low phase noise in the exciter output signals exist for frequencies extending from several hundred Hz to several MHz from the carrier [1-3].

This paper reports on the design and performance of the CW signal generator portion of a radar exciter exhibiting state-of-the-art phase noise characteristics. A summary of the primary exciter signal generator characteristics is shown in tables 1 and 2.

Table 1. Signal Generator (Coherent) Output Signals

1. Receiver 1st L.O.:	8080MHz-8560MHz (80MHz increments)
2. Receiver 2nd L.O.:	1430MHz
3. Receiver 3rd L.O.:	32.5MHz
4. System Outputs:	2.5MHz, 5MHz, 10MHz, 20MHz, 30MHz, 40MHz, 80MHz
5. Auxiliary Output:	320MHz
6. System Clocks:	10MHz, 20MHz

Table 2. Receiver 1st L.O. Signal Phase Noise Characteristics

Carrier Offset Frequency (Hz)	Phase Noise Sideband Level (dBc/Hz)	
	Stationary	Under Vibration
10	-70	-60
100	-100	-95
1K	-125	-125
10K	-135	-135
>100K	-145	-145

This work was supported by the Naval Research Laboratory, contract #N00014-91-C-2071 and #N00014-92-C-2222.

0-7803-1945-1/94 $3.00 © 1994 IEEE

The profile of the hardware vibration test spectrum is shown in Figure 1.

Signal Generator Design

The signal generator was designed to consist of three functional subassemblies: (1) Reference Signal Generation Subassembly, (2) First L.O. Signal Synthesis Subassembly, and (3) HF Signal Synthesis Subassembly. Each of these is described in detail below.

1. Reference Signal Generation

All of the signal generator output signals are derived from a single, ultra-low noise, 80MHz master oscillator. The oscillator uses four, series-connected, 3rd overtone SC-cut crystals in order to achieve excellent flicker-of-frequency noise (-143dBc/Hz at fm=100Hz) with a noise floor level of -180dBc/Hz. Total crystal power dissipation is 30mW, and the oscillator output power is 23dBm [4]. As shown in figure 2, cascaded stages of high level, diode frequency doubler-filter-amplifier combinations provide outputs at 160MHz, 320MHz, 640MHz, and 1280MHz.

In addition, a cascade of two, single-pole surface acoustic wave resonator (SAWR) filter-amplifiers act as a narrowband (20KHz) spectral cleanup filter. Operation of the SAWR filters at 20dBm SAWR dissipation provides a carrier signal phase noise floor level close to -180dBc/Hz at 320MHz. The 640MHz signal is used to drive a step recovery diode (SRD) multiplier, and a manifold filter-bank extracts 7680MHz, 8320MHz, and 8960MHz signals from the SRD output. The key to obtaining low phase noise performance in the SRD output signals involve use of a low noise master oscillator, use of low flicker-of-phase noise diode doublers and inter-stage amplifiers, use of SAWR filters selected for low flicker-of-frequency noise, use of a broadband resistive pad at the SRD output, and operation of all components at relatively high drive level [5].

The requisite, 2nd L.O. signal at 1430MHz is generated by mixing 1280MHz with 150MHz, with the 150MHz signal derived via mixing the 160MHz signal with a 10MHz signal generated in the HF Signal Synthesis subassembly. Figures 3 and 4 show the result of measurement of 1280MHz signal phase noise and 1430MHz signal amplitude noise, respectively. Phase Noise measurements were made via comparison to a second, low noise signal generator using master oscillator and frequency multiplier hardware identical to that of figure 2. For all phase noise measurements, an assumption of near-equal phase noise performance was made for both the exciter and test signal generators, indicating the performance of each was approximately 3dB better than that measured for the (phaselocked) combination of both. AM noise measurements were made using readily available diode DC detectors. In some cases, there is reason to believe the AM noise measurement is limited by detector 1/f noise. Since the AM noise measurement results were well within specification limits, no attempt to establish detector performance limitations were made.

In addition, three UHF reference signals (400MHz, 480MHz, 560MHz) are generated in the Reference Signal Generation assembly via use of a SRD frequency multiplier driven at 80MHz.

2. First L.O. Signal Synthesis

As shown in figure 5, six of the seven, X-band, first local oscillator signals are derived via individually mixing both the 7680MHz and 8960MHz signals with 400MHz, 480MHz and 560MHz UHF offset signals. The 8320MHz first L.O. signal is derived directly. Use of offset signals in the 400MHz to 600MHz range insures that in-band, unwanted mixer products are very high order and occur below -80dBc levels. It also insures that adequate suppression of mixer L.O. leakage and undesired sideband signals can be easily suppressed using practical bandpass filters at the mixer outputs. Figures 6 and 7 show results of representative First L.O. signal phase and amplitude noise measurements.

3. HF Signal Synthesis

All of the requisite signal generator HF signals were derived via use cascaded stages of regenerative frequency dividers. The use of regenerative dividers provides much lower flicker-of-phase noise and noise floor levels, compared to digital, logic level dividers [5]. Figures 8 and 9 show the phase noise levels measured at the first (80MHz-40MHz) and last (5MHz-2.5MHz) regenerative divider outputs. The discrete spurious signals at 43KHz and the harmonics thereof are associated with crosstalk from the HP Noise Measurement System HPIB bus. It should be noted that lower flicker-of-phase noise performance has been achieved for regenerative frequency dividers, compared to that shown in figure 9. The divider flicker-of-phase noise limitation is associated with either the low frequency (below 5MHz) double balanced mixer or TO-8 amplifier used.

Performance Under Vibration

Hardware Construction

In order to minimize prototype signal generator cost, the hardware was constructed via use of individual, SMA connectorized components mounted on metal mounting shelves, with component interconnections and front panel connections made using semi-rigid coaxial cable (figure 10). An analysis of the carrier signal frequency and/or phase sensitivity of the most sensitive signal generator components to vibration was made. This included the master oscillator quartz crystals, the UHF SAW resonators, and the component inter-connecting, semi-rigid coaxial cable. In order to insure no output signal spectral degradation occurred (for offset frequencies above 1KHz) due to specified vibration levels, the signal generator was constructed using the following hardware design implementations: (1) The master oscillator and SAWR filter enclosures were mounted on a single vibration-isolated plate using 12 isolators chosen to provide attenuation beyond 40Hz (figure 11); (2) Each of the three subassembly mounting shelves as well as the overall enclosure side and top panels were reinforced with Soundcoat soundfoil for damping, and a large number of shelf mounting points used (figure 10); (3) Cable tie downs were used for longer shelf-to-front-panel coaxial cable connections; (4) additional (silicone sheet sponge rubber) damping was used inside the SAWR filter enclosures; and (5) SAW resonators were individually screened for vibration sensitivity.

SAWR Vibration Sensitivity Tests

Prototype SAW resonators were provided by the SAWR vendor in each of four, distinct die designs. The vibration sensitivity of each device was evaluated via sinusoidal (shake table) excitation over a frequency range 50Hz-10KHz. In order to use a single, coaxial cable connection to the SAWR shake table fixture, the SAWR was operated with the 2nd port grounded and the connecting cable length adjusted to an odd number of multiple quarter-wavelengths. The test set-up and test results are shown in figures 12 and 13, respectively. In figure 13, same die design pairs are denoted as #1 and #3, #6 and #6b, #7 and #8, and #9 and #10. As shown in figure 13, both the highest and lowest sensitivity were exhibited by the same SAWR die design (#1 and #3), with the spread in sensitivity values ranging from $5X10^{-9}$ per g (#1) to $30X10^{-9}$ per g (#3 and #5). In all cases, the greatest degree of SAWR vibration sensitivity was for motion in the direction orthogonal to the blank face. Figure 13 also indicates a substantial mechanical resonances in the 5KHz to 8KHz region for all the devices. With a resistor mounted in place of the SAWR in the fixture, resonances in the same region occur, but the degree of resonant peaking is much less. Because of the lack of uniformity in SAWR vibration sensitivity characteristics, devices in the originally provided (#1 and #3) die design were used in both the exciter and noise test set signal generators.

Exciter Signal Generator Vibration Tests

Signal generator first L.O. signal phase noise measurements were made with the signal generator subjected to the vibration test spectrum depicted in figure 1. The signal generator itself was mounted on the shake table on an angular mounting fixture (figure 14) in order to simultaneously provide stress in directions orthogonal to both the component mounting shelves and the front panel. The phase noise measurements were conducted as previously described, via phaselock to a test set reference generator (figure 15). The initial test result is shown in figure 16. As shown in the figure, there are two carrier frequency offset regions where vibration-induced, signal spectral degradation appears to occur. In the first region (10Hz to 200Hz) spectral degradation is primarily associated with: (a) peaking in the transmission response of the isolators used to mount the master oscillator and SAWR filters, and (b) structural mechanical resonances. Spectral degradation in this region was both anticipated and allowed in the hardware performance specification. The spectral degradation in the carrier offset frequency region 1KHz-8KHz was of concern. Additional investigation confirmed that the source of this degradation was the response of the non-(acoustically) shielded SAWR filters in the test set reference generator to shake table pump and blower motor sound levels. Providing acoustic attenuation in these filters reduced the resultant signal spectral degradation in the 1KHz-8KHz region to inconsequential levels, as shown in figure 17.

Conclusions

The signal generator portion of a radar exciter has been successfully designed, fabricated, and tested.

Achievement of state-of-the-art output signal phase noise performance was made possible via use of: low (flicker-of-frequency) noise quartz crystal and SAW resonators, low flicker-of-phase noise components, regenerative frequency division, and signal generator circuit operation at relatively high drive levels.

Minimization of vibration-induced, signal spectral

degradation was achieved via vibration isolation of critical circuit subassemblies and mechanical/acoustic structural damping.

A wide, non-uniform variation in SAW resonator frequency sensitivity to mechanical vibration as well as acoustic (sound pressure) stress was observed.

References

[1] M. I. Skolnick, Introduction to Radar Systems, Second Edition, McGraw-Hill, 1980.

[2] N. Slawsby, "Frequency Control Requirements of Radars," Proc. 1994 IEEE Freq. Contr. Symp., June, 1994.

[3] Fundamental Signal Generator Requirements in Relation to Radar System Performance" Proc. 1994 Freq. Contr. Symp., June 1994.

[4] M. M. Driscoll, "Reduction of Quartz Crystal Oscillator Flicker-of-Frequency and White Phase Noise (Floor) Level and Acceleration Sensitivity via Use of Multiple Resonators," Proc. 1992 Freq. Contr. Symp., June 1992, pp. 334-339.

[5] M. M. Driscoll and T. D. Merrell, "Spectral Performance of Frequency Multipliers and Dividers", Proc. 1992 Freq. Contr. Symp., June 1992, pp. 193-200.

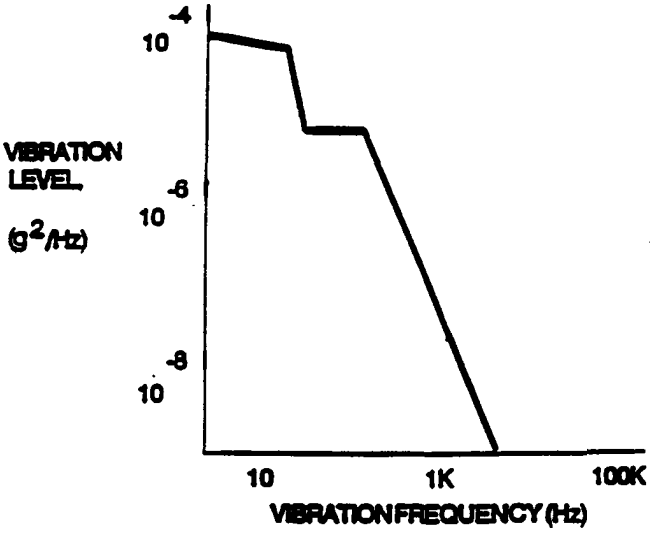


Figure 1. Vibration Test Levels for the Exciter Signal Generator

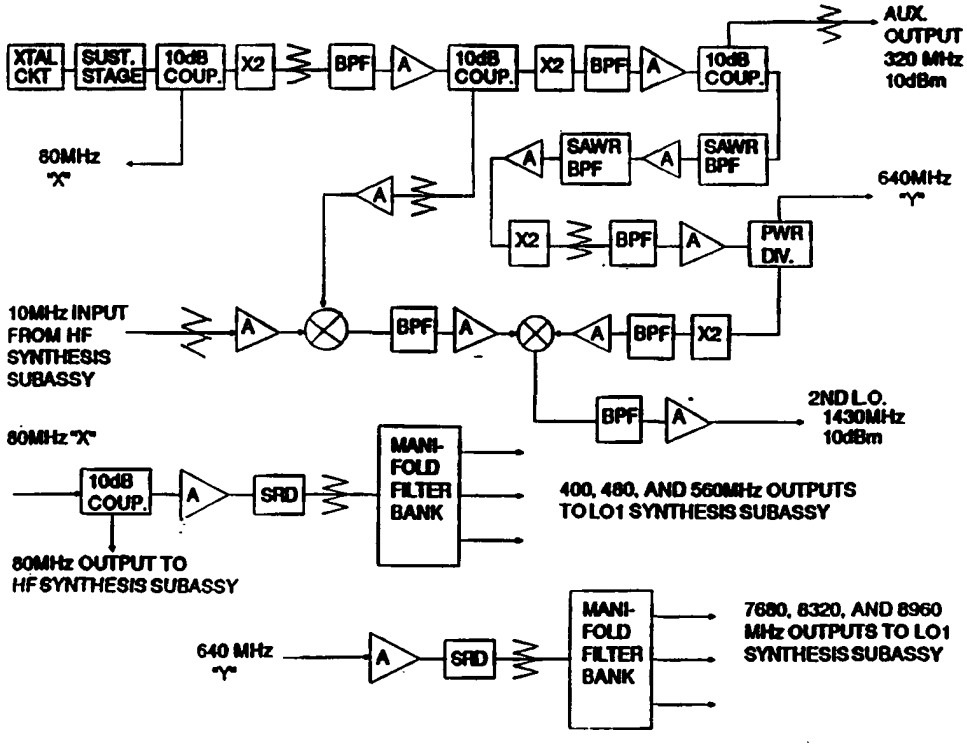


Figure 2. Reference Signal Generation Subassembly Block Diagram

A FACTOR 1600 INCREASE IN NEUTRAL ATOMIC BEAM INTENSITY USING LASER COOLING

M.D.Hoogerland [1], J.P.J.Driessen, E.J.D.Vredenbregt,
H.J.L.Megens, M.P.Schuwer, H.C.W.Beijerinck and K.A.H. van Leeuwen
Physics Department, Eindhoven University of Technology, P.O. Box 513,
5600 MB Eindhoven, The Netherlands

Abstract

Using a three-step transverse cooling scheme, a strongly diverging flow of metastable Ne*($^3P_2$) atoms is compressed into a small diameter (3 mm), well-collimated (1 mrad) atomic beam, using only 200 mW total laser power. This extremely valuable technique for frequency standard applications is applicable to many different atomic systems.

Introduction

In many applications in atomic physics, e.g., in high resolution spectroscopy and atomic beam frequency standards, a high degree of atomic beam collimation and a high beam flux are desirable. Also a small beam diameter is of paramount importance to minimize both the variation of fields across the atomic beam and the size of the holes in, e.g., an R.F. cavity used for the spectroscopy. Laser cooling provides a means to manipulate atomic beams such that all requirements can be met simultaneously. Consequently, this technique can drastically improve existing experiments and open up possibilities for new experiments in these fields.

Radiation pressure, the basis for the most simple laser cooling mechanism, provides a mean to exert large forces on neutral atoms. The maximum radiation pressure force is $F = \hbar k\Gamma/2$, with $\hbar k$ the photon momentum and Γ the spontaneous emission rate, equal to the natural linewidth of the transition used. Due to the non-conservative nature of this force, the 4-D phase space volume of the transverse position and momentum coordinates can be reduced considerably with a corresponding increase in phase space density. Liouville's theorem, that describes the conservation of phase space density in optics and particle beams, is thus 'beaten': an atomic beam can actually be 'brightened' by laser cooling.

Although the 'beam brightener' described here was developed for application in atomic scattering experiments [2], the basic properties of the 'brightened' atomic beam are also of great value for atomic beam frequency standard applications.

Experimental set-up

An atomic beam of metastable Ne*($\{3s\}^3P_2$) neon atoms is intensified in a three-step process. We use the closed-level Ne*($^3P_2$) to Ne**($^3D_3$)D transition at λ = 640 nm, with a maximum deceleration of 8 10^5 ms^{-2}. Our discharge-excited supersonic Ne* source is cooled by liquid nitrogen. The resulting axial velocity is $v_{\|}$ = 580 m/s with a spread $\Delta v_{\|}$ = 100 m/s. Although all numbers in this section are for metastable neon, the techniques described here can easily be transferred to other atoms with a strong closed-level transition.

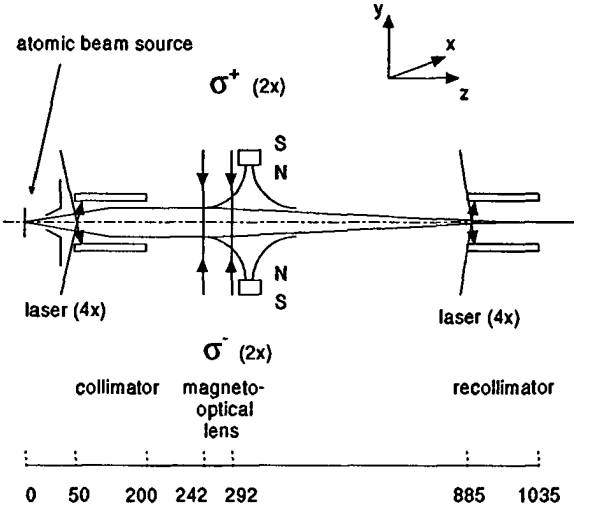


Figure 1: Schematic view of the three stages of the experimental set-up for the production of intense and bright atomic beams.

A schematic view of the experimental set up is given in Fig. 1. First, we collimate the diverging atomic beam into a parallel beam with a fairly large diameter. The capture angle of this stage is of paramount importance, because it fully determines the final beam flux. Second, this parallel beam is focussed to a small spot using a two-dimensional magneto-optical trap as the equivalent of a lens in optics. Third, near the focus of this lens, the atomic beam is recollimated into a narrow and well collimated beam. The phase space compression takes place in the first and the third stage; the action of the magneto-optical lens is mostly conservative.

Our final aim is to obtain a capture angle θ_0 = 100 mrad, with θ the polar angle with respect to the atomic beam axis. With an effective radius of the source of

0-7803-1945-1/94 $3.00 © 1994 IEEE

0.2 mm this corresponds to an initial 2-D phase space volume of 400 mm^2 mrad^2. For the final beam we demand a 1 mm radius and a divergence of 0.5 mrad, the latter determined by the Doppler laser-cooling limit $\Delta v_D = 0.23$ m/s for the transverse velocity. The final 2-D phase space volume would then be equal to 0.25 mm^2 mrad^2, resulting in a gain in 'brightness' of 1600. At 1 meter distance from the source the gain in beam flux then is equal to 10^4.

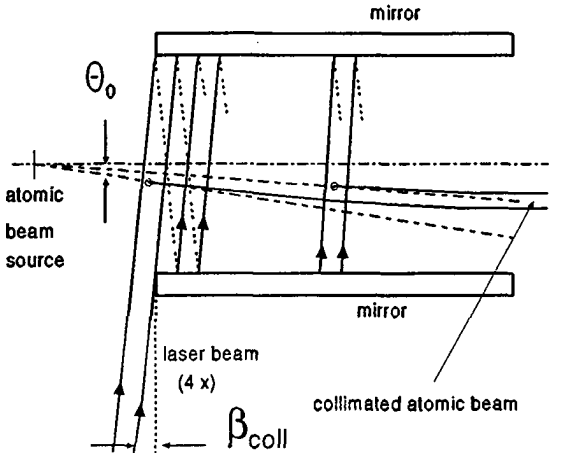


Figure 2: Schematic view of the mirror section used for (re)collimation of the atomic beam in the first an d third stage. A laser beam is injected between two nearly parallel mirrors at an angle $\beta_{0,(re)coll}$. With each reflection the angle between the laser beam and the plane perpendicular to the atomic beam is reduced by an amount α, resulting in effectively curved wave fronts. The trajectory of the atom is captured at resonance and then follows the wavefront.

In the first stage we apply a two-dimensional optical molasses to collimate the beam. To achieve a large capture angle while keeping the cooling time as short as possible, it is desirable to use curved wavefronts for the molasses laser beams [3, 4]. For zero detuning of the laser, each atom will be 'locked' to the curved wavefront once it comes into resonance, provided that the laser cooling force can match the centrifugal force. The trajectories with a large initial angle θ are thus captured at the entrance of the collimator, those with a smaller angle further downstream. The resulting beam profile is annular. By using multiple laser beam reflections between two nearly parallel mirrors, as shown in Fig. 2, we obtain both the effectively curved wavefronts and a large interaction time at a bargain price in laserpower. The offset angle between the 150 mm long mirrors (60 mm apart) is $\alpha = 1.5$ mrad.

We inject the laser light at an angle $\beta_{0,coll} = 100$ mrad with respect to the plane perpendicular to the atomic beam axis. With each reflection, the angle β between the laser beam and the atomic beam axis is reduced by an amount α. Since an exit angle $\beta_{N,coll} = 0$ would cause the laser beams to retrace its initial path and subsequently heat the atomic beam, a finite exit angle $\beta_{N,coll} > 0$ is used. We compensate the Doppler shift $kv_{\|}\ \beta_{N,coll}$ corresponding to this exit angle by detuning the laser frequency in this section by $\Delta\omega_{coll} = +35$ MHz from resonance.

The same principle holds for two dimensions and two sets of mirrors. For each mirror set, the light is injected from both sides to obtain a complete coverage of the interaction region. For the four laser beams in the first stage together, we use 30 mW of laser power. This way, we collimate an atomic beam with an initial HWHM divergence $\Delta\theta_0 = 100$ mrad to a well-collimated beam with a HWHM radius $\Delta r = 10$ mm and a residual HWHM divergence $\Delta\theta_1 = 2$ mrad.

To focus the collimated atomic beam in the second stage we use a two-dimensional 'magneto-optical trap' [5, 6]. In a quadrupole magnetic field $B_x = G(z)\ x$ and $B_y = -G(z)\ y$ perpendicular to the beam axis, we illuminate the atomic beam with counter-propagating laserbeams with orthogonal circular polarization. The gradient $G(z)$ increases approximately linear with the axial position z in the lens, from $G = 0.2$ Gauss/mm to $G = 1.2$ Gauss/mm at the exit. Because the characteristic cooling time of the transverse velocity is short compared to the total interaction time in the lens, each atom is effectively 'locked' to a transverse velocity where the Doppler detuning along it's trajectory equals the Zeeman detuning of the magnetic field. For the x-component this means $kv_{\perp,x}(z) = -(g\mu/\hbar)\ G(z)\ x$. The final action is a converging trajectory with angle $\theta_2 = v_{\perp,x}(z_{exit}\ /\ v_{\|} \sim x$. Consequently, this configuration acts as a lens in optics. Its action is rather insensitive for the laser power used.

The interaction length is 50 mm, which corresponds to an interaction time of $\approx 100\ \mu$s. A typical cooling time is less than 1μs. We use in-vacuum polarization-preserving mirrors to illuminate the atomic beam from four sides with circular polarization using a single input beam. The laserpower used in this stage is thus limited to $\approx$ 100 mW; the detuning is equal to $-\Gamma$ with respect to ω_0. The magnetic field is obtained using permanent ferrite magnets outside the vacuum, at the downstream end of the interaction region. Thus the quadrupole magnetic field gradient increases approximately linear with the axial position.

The focal length is $f_{MOL} = 0.70$ m for $v_{\perp} = 580$ m/s, in agreement with the analytical prediction. Because the action of the lens is proportional to $v_{\|}^{-1}$, it has a chromatic aberration. Consequently, the size of

the focus and thus the size of the compressed atomic beam is limited to a radius of 1.4 mm by the velocity spread of the atomic beam.

Near the focus, we recollimate the atomic beam using a mirror set which is identical to the first stage. However, the angle between the mirrors $\alpha = 0$ mrad and the coupling angle $\beta_{0,recoll}$ mrad are chosen differently to match the incoming atomic beam and the detuning of the laser beam. The latter is chosen equal to the detuning in the first stage to avoid using an extra AOM. The laser power used is 10 mW, less than in the first stage which is due to the much smaller atomic beam width.

We use two scanning wire-detectors for measuring 1-D integrated beam profiles. A stainless steel wire is scanned through the atomic beam by a stepper motor; at each lateral position the metastable atom induced Auger emission is monitored by measuring the cur rent from the wire. The wire detectors can be inserted (horizontally or vetically) at different positions along the beam line, allowing a measurement of the effective beam divergence. For monitoring the profile at the end of the beam line a Micro Channel Plate electron multiplier (MCP) with a phosporous screen anode was used.

Results

In Fig. 3(a) the atomic beam profile is shown with only the first collimator stage in operation, for two positions z= 200 and 650 mm downstream of the source ($z = 0$), respectively. The diverging atomic beam is collimated to a beam with a HWHM radius $\Delta r = 10$ mm and a residual HWHM divergence $\Delta\theta_1$ =2 mrad.

When the second, magneto-optical stage is activated, the downstream atomic beam profile is focussed into a sharp peak. In Fig. 3(b) we show the HWHM of the atomic beam profile at position $z = 1040$ mm, measured as a function of the strength of the lens, i.e. the focal length. This variation is achieved by partial blocking of the down stream end of the laser beam entering the magneto-optical lens. We observe a clear minimum in the atomic beam half-width as we go from an underfocussed atomic beam to an overfocussed atomic beam. The minimum value of 1.5 mm is fully determined by the velocity spread of the atomic beam source. The focus can be positioned laterally by small changes in the position of the permanent magnets, i.e. by adjusting the axis of the quadrupole magnetic field.

Finally, the third, recollimation stage is inserted in the beam line at position $z = 960$ mm, i.e. centered on the focus of stage two. The beam profile at $z = 1380$ mm is shown in Fig. 3(c): its HWHM radius of 2.5 mm is only slightly larger than the minimum value given in Fig. 3(b) for the focussed beam. This demonstrates the effective operation of the recollimation stage. We can minimize the beam divergence to a value close to the Doppler limit $\Delta\theta_D = v_D/v_{parallel} =$ 0.5 mrad at the cost of a slight increase in beam half-width, resulting in the maximum 2-D phase space compression factor of 400. This is a factor four lower than discussed in the introduction, mainly due to the beam width which is a factor two larger.

Summary

Summarizing, we presented the operation of a three stage atomic beam brightener for thermal atomic beams. For atomic beam frequency standards, this means that with a single laser, the number of atoms in a collimated, small diameter atomic beam, and thus the signal, can be increased by a factor 1600, leading to shorter integration times, and thus a better frequency standard. This technique is easily portable to different atomic systems.

Acknowledgements

The authors would like to thank H. Metcalf for stimulating discussions. This work was financially supported by the Dutch Foundation for Fundamental Research on Matter (FOM).

References

[1] Present address: National Institute of Standards and Technology, Physics A167, Gaithersburg MD 20899, U.S.A.

[2] M.D. Hoogerland, J.P.J. Driessen, E.J.D. Vredenbregt, H.J.L Megens, M.P. Schuwer, H.C.W. Beyerinck and K.A.H. van Leeuwen, submitted to PRL.

[3] F. Shimizu, K. Shimizu, and K. Takuma, Chem. Phys. **145**, 327 (1990).

[4] A. Aspect, N. Vansteenkiste, R. Kaiser, H. Haberland, and M. Karrais, Chem. Phys. **145**, 307 (1990).

[5] E. L. Raab, M. Prentiss, A. Cable, S. Chu, and D. E. Pritchard, Phys. Rev. Lett. **59**, 2631 (1987).

[6] J. Nellesen, J. Werner, and W. Ertmer, Opt. Commun. **78**, 300 (1990).

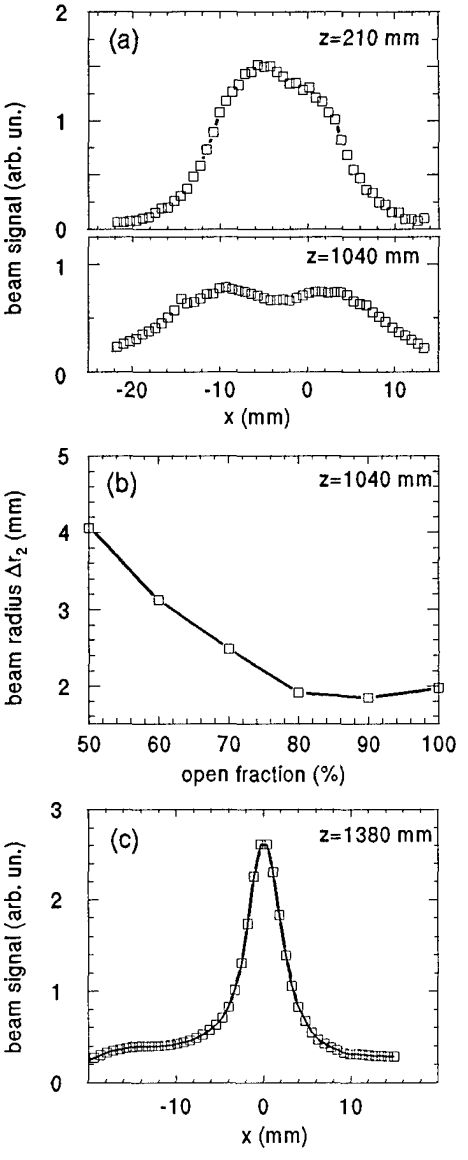


Figure 3: (a) Atomic beam profile with only the first stage (collimator) activated, measured with the scanning wire detector at $z = 210$ mm (top) and at $z = 1040$ mm (bottom). The block shaped beam profile directly beyond the collimator (top) spreads out on its downstream path (bottom). (b) Atomic beam width at $z = 1040$ mm with the first and the second stage active, as a function of the open fraction of the laser beam of the magneto-optical lens, i.e. fraction of the maximum strength of the lens. We observe a clear minimum when going from an underfocussed to an overfocussed atomic beam. The recollimator is not in operation. (c) Atomic beam profile with all three stages in operation, measured with the scanning wire detector at $z = 1380$ mm. The width is only slightly larger than the minimum size of the focus of the magneto-optical lens as given in part (b).

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

OPTICAL LATTICES FOR ATOMIC FOUNTAIN FREQUENCY STANDARDS

W. D. Phillips, W. M. Golding†, A. Kastberg, S. L. Rolston, and R. J. C. Spreeuw*

Atomic Physics Division, PHY A167, National Institute of Standards and Technology, Gaithersburg, MD 20899.

Abstract

Atomic fountain frequency standards require extremely low temperature atomic samples. Laser cooling of neutral atoms has achieved temperatures of a few microkelvin, but even such low temperatures still present difficulties. Here we report laser cooling in an optical lattice, along with cooling by adiabatic expansion, that gives a temperature of 0.7 μK. Such low temperatures should improve the performance of atomic fountain frequency standards.

Introduction

This paper is based on a presentation made at the 1994 IEEE International Frequency Control Symposium, 1-3 June, in Boston, MA. This is not intended as a comprehensive treatment of atomic fountains and optical lattices, but rather as a summary of the presentation and a guide to the literature on the subject.

Because many of the factors that limit the accuracy and precision of atomic frequency standards are related to thermal motion of the atoms, the use of laser cooling to reduce that thermal motion has long been proposed. Some of these ideas are reviewed in reference [1]. Laser cooling of cesium, the atom on which the definition of time is based, has led to three dimensional temperatures below 3 μK [2]. Such low temperatures are achieved by virtue of the recently discovered "polarization gradient" or "Sisyphus" mechanisms for laser cooling below the Doppler-cooling limit [3].

Atomic Fountains

Traditional atomic cesium clocks employ the Ramsey method of separated oscillatory fields [4] in which an atomic beam is sent successively through two microwave cavities operating near the hyperfine resonance frequency of about 9.2 GHz. The linewidth of the Ramsey resonance is inversely proportional to the flight time of the atoms between the two cavities. In most laboratory frequency standards, with thermal atomic beams, this time is on the order of 10 ms.

Atoms with temperatures in the microkelvin range cannot form an atomic beam in the usual sense. They simply fall under the influence of gravity. The usual procedure for making use of such cold atoms is in an atomic fountain geometry. First proposed by Zacharias [5], it was only recently demonstrated [6], once laser cooling had become available. Atoms in a fountain travel vertically upward through a microwave cavity, reach the apex of their ballistic trajectory and fall back through the same cavity, completing the second interaction of the Ramsey method. For a modest trajectory height of one meter, the time between interactions with the cavity is on the order of a second, two orders of magnitude longer than is typical of a traditional atomic beam standard. A number of groups have demonstrated Ramsey-Zacharias resonance in a cesium atomic fountain where atoms pass entirely through the microwave cavity both upward and downward [7-10]. This double passage in opposite directions insures the cancellation of several possible systematic errors.

U.S. Government work not protected by U.S. copyright

With atomic temperatures of 3 μK, which implies a thermal velocity of 1.4 cm/s along a given axis, after 1 second even a point source of cesium atoms will spread out to have a Gaussian diameter (at the $1/e^{1/2}$ points) of 2.8 cm. Design considerations for the 9.2 GHz microwave cavity limit the opening in the cavity (to allow passage of the atoms) to about a 1 cm diameter. Therefore, only a small fraction of the atoms initially launched up through the cavity will return through the cavity to complete the Ramsey interaction and contribute to the signal.

While one might imagine compensating for this reduction in signal by increasing the number of atoms launched, this in fact presents a difficulty because of collisions between cold atoms in the launched sample. The collisional hyperfine frequency shift for atoms in the $m_F = 0$ states has been measured to be 5.5 mHz at a density of 3.5×10^8 cm^{-3}, or a fractional shift of 1.7×10^{-15} at 10^6 cm^{-3}. Thus, to achieve relative accuracy on the order of 10^{-16}, an often-stated goal for atomic fountains, one will need to control and limit the density. On the other hand, lower density will mean a smaller number of atoms and a lower shot-noise-limited signal-to-noise ratio. This will in turn limit the ability to achieve high precision or accuracy.

Atoms that are launched upward through the cavity, and that do not return, contribute to collisional frequency shifts but not to the signal. Thus it is essential to ensure that as large a fraction as possible of launched atoms return. This means the transverse velocity spread should be as small as possible. Among the possibilities for achieving this are: velocity selection by stimulated Raman transitions [11], cooling by velocity selective coherent population trapping [12], and Raman cooling [13]. Here we report results of a different technique for laser cooling of cesium using adiabatic expansion in an optical lattice.

Optical Lattices

Atoms laser cooled to sufficiently low temperatures can become trapped in the sub-wavelength-size potential wells formed when interfering laser beams create spatially varying light-shift potentials for the atoms [14-16]. This regular arrangement of potential wells is an optical lattice. Atoms trapped in the wells exhibit oscillatory, quantized motion. A key feature of optical lattices is that the character of the interference pattern forming the lattice does not vary in time. This is achieved either by active control of the relative phase of the interfering laser beams [15] or by a choice of geometry that makes only the position but not the character of the lattice sensitive to phase variations [14]. Using the latter approach, with four intersecting laser beams in a three-dimensional configuration, we have cooled cesium atoms to temperatures as low as one microkelvin in quasi-equilibrium.

Additional cooling of the atoms is obtained by allowing adiabatic expansion of the atoms initially localized near the bottoms of the microscopic potential wells. If the strength of the potential wells is reduced, a localized atom expands to fill more of its well, resulting in adiabatic cooling of the atoms. We have observed cooling of cesium by this technique to temperatures of 700 nK, corresponding to an rms velocity less than 7 mm/s along a given axis [17,18].

The optical lattice technique should be easily adapted for use in an atomic fountain. The "moving optical molasses" now used to launch atoms in a fountain [7] would be replaced with a moving lattice, achieved, as in the moving molasses, by offsetting the frequencies of some of the interfering laser beams. The adiabatic cooling can be easily accomplished while the atoms are being launched. Furthermore, we believe that launching atoms trapped in a moving optical lattice will be more effective than in a moving molasses where the atoms are less likely to be microscopically localized.

Acknowledgements

This work is supported in part by the Office of Naval Research and by the National Science Foundation grant no. PHY 9312572. A. K. thanks the Swedish Natural Science Research Council (NFR) and R. S. thanks the Niels Stensen Stichting for financial support.

References

†Also with the Naval Research Laboratory and George Washington University

*Present address: Universität Konstanz, Fakultät für Physik.

[1]. S. L. Rolston and W. D. Phillips, "Laser-Cooled Neutral Atom Frequency Standards," Proc. IEEE, vol 79, pp. 943-951, 1991.

[2]. C. Salomon, J. Dalibard, W. D. Phillips, A. Clairon and S. Guellati, "Laser Cooling of Cesium Atoms below 3 μK," Europhys. Lett., vol. 12, pp. 683-688, 1990.

[3] C. Cohen-Tannoudji and W. D. Phillips, "New Mechanisms for Laser Cooling," Physics Today, pp. 33-40, October 1990.

[4]. N. Ramsey, Phys. Rev. vol 76, p. 996, 1949.

[5]. J. Zacharias, Phys. Rev. vol. 94, p. 751T, 1954.

[6]. M. Kasevich, E. Riis, S. Chu and R. DeVoe. "RF Spectroscopy in an atomic fountain," Phys. Rev. Lett., vol. 63, pp. 612-615, 1989.

[7]. A. Clairon, C. Salomon, S. Guellati, and W. Phillips, "Ramsey Resonance in a Zacharias Fountain," Europhys. Lett., vol. 16, pp. 165-170, 1991.

[8]. K. Gibble and S. Chu, "Laser-cooled Cs frequency standard and a measurement of the frequency shift due to ultracold collisions," Phys. Rev. Lett., vol 70, pp. 1771-1774, 1993.

[9]. A. Michaud, M. Chowdhury, K. Zetie, C. Cooper, G. Hillenbrand, V. Lorent, A. Steane, and C. Foot, "Realisation of a frequency standard using an atomic fountain," Proceedings of the 7th European Frequency and Time Forum, Neuchâtel, Switzerland, March 16-18, 1993 (Swiss Foundation for Research in Microtechnology, Neuchâtel) pp. 525-530.

[10]. A. Clairon, Ph. Laurent, A. Nadir, G. Santarelli, M. Drewsen, D. Grison, B. Lounis, and C. Salomon, "A laser cooled cesium fountain clock: design and expected performances," SPIE vol 1837, pp. 306-313, 1992; see also: A. Clairon et al., "A simple and compact source of cold atoms for cesium fountains and microgravity clocks,' Proceedings of the 6th European Frequency and Time Forum, p. 27, 1992.; S. Lea et al., "Laser cooling and trapping of atoms: new tools for ultra-stable caesium clocks," Physica Scripta, vol. T51, p. 78, 1994; G. Santarelli et al. "Preliminary results in a cesium atomic fountain frequency standard," in Proceedings of the 6th European Frequency and Time Forum, Munich 1994.

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

[12]. A. Aspect, E. Arimondo, R. Kaiser, N. Vansteenkiste, and C. Cohen-Tannoudji, "Laser cooling below the one-photon recoil energy by velocity-selective coherent population trapping," Phys. Rev. Lett., vol 61, pp. 826-829, 1988.

[13]. N. Davidson, H. Lee, M. Kasevich and S. Chu, "Raman cooling of atoms in two and three dimensions," Phys. Rev. Lett., vol 72, pp. 3158-3161, 1994.

[14]. P. Verkerk, B. Lounis, C. Salomon, and C. Cohen-Tannoudji, "Dynamics and spatial order of cold cesium atoms in a periodic optical potential, Phys. Rev. Lett., vol. 68, pp. 3861-3864, 1992. and G. Grynberg, B. Lounis, P. Verkerk, J.-Y. Courtois, and C. Salomon, "Quantized motion of cold cesium atoms in two- and three-dimensional Optical Potentials," Phys. Rev. Lett., vol. 70, pp. 2249-2252, 1993.

[15]. A. Hemmerich and T. Hansch,"Two-dimensional atomic crystal bound by light," Phys. Rev. Lett., vol. 70, pp. 410-413, 1993.

[16]. P. Jessen, C. Gerz, P. Lett, W. Phillips, S. Rolston, R. Spreeuw, and C. Westbrook, "Observation of quantized motion of Rb atoms in an optical field," Phys. Rev. Lett., vol 69, pp. 49-52, 1992).

[17]. R. Spreeuw, P. Jessen, A. Kastberg, W. Phillips and S. Rolston, "Sub-μK cesium by adiabatic expansion in an optical lattice," post-deadline presentation at the 1993 Annual Meeting of the Optical Society of America, October 3-8, Toronto.

[18]. A. Kastberg, W. Phillips, S. Rolston, R. Spreeuw, and P. Jessen, "Adiabatic cooling of cesium to 700 nK in an optical lattice," to be published.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Velocity Distributions of Atomic Beams by Gated Optical Pumping

W. D. Lee, J. H. Shirley, and R. E. Drullinger
Time and Frequency Division
National Institute of Standards and Technology
Boulder, Colorado 80303, USA
Tel: 303-497-7674

Abstract— **In evaluating the accuracy of a cesium-beam frequency standard, accurate measurement of the atomic velocity distribution is important. In frequency standards which employ atoms with thermal velocities, the measured atomic resonance frequency differs from the true resonance by several parts in 10^{13} due to the second-order Doppler shift. To achieve the frequency accuracy goal for NIST-7 of $1 \cdot 10^{-14}$, the uncertainty in the second-order Doppler shift must be no more than a few parts in 10^{15}. This requirement establishes an upper bound on the uncertainty of the mean-square atomic velocity of about 1 percent. We present the results of experiments designed to measure the velocity distribution of NIST-7 using two independent techniques: gated optical pumping, and Ramsey fringe inversion. We show that these techniques yield velocity distributions and corresponding second-order Doppler shifts that agree within the stated tolerances.**

I. Introduction

A variety of techniques have already been developed to measure the velocity distribution of atomic beams in atomic frequency standards. Boulanger's method[1] is based upon the measurement of the amplitude of the central maximum of the Ramsey resonance as a function of the microwave power level. This technique is probably not appropriate for NIST-7. At the present static magnetic field (C-field) magnitude, the peak amplitude of the Ramsey fringe does not settle to a constant value (as microwave power is increased) before adjacent transitions begin to interfere with the measurement. A novel method introduced by Hellwig[2] yielded velocity selection of the atomic beam by application of short microwave pulses whose widths were equal to the transit time of a given velocity group through one end of the Ramsey cavity. The pulse separation was the transit time across the entire cavity. This technique is, however, ineffective for evaluating the velocity distribution of NIST-7. The "mono-velocity" Ramsey lineshapes obtained using this technique are contaminated by Ramsey fringes from harmonics of the velocity group being studied. This effect is quite pronounced in NIST-7 which exhibits a full Maxwellian velocity distribution. Jarvis[3] introduced a technique whereby the velocity distribution was obtained from Fourier transforms of the Ramsey lineshape at each of several known microwave power levels. We have modified and extended this technique[4] to serve as our primary tool for measuring velocity distributions. However, for a thorough evaluation of the frequency accuracy of NIST-7, we desire two independent measurement techniques to estimate the major frequency biases present. Thus, we have developed the additional technique of gated optical pumping.

Contribution of the U. S. Government, not subject to copyright.

II. Experimental Method

The unique design of the NIST-7 primary frequency standard, specifically the optical pumping scheme,[5] is well suited for this measurement technique. Figure 1 is a schematic of NIST-7.

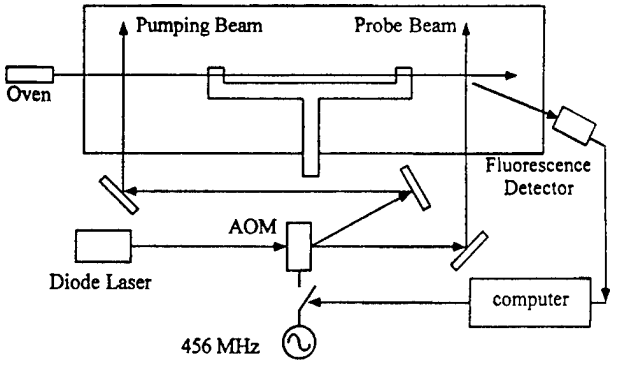


Fig. 1. Schematic of NIST-7.

The oven emits a collimated beam of cesium atoms which are uniformly distributed among the $F = 3$ and $F = 4$ ground states. (See figure 2).

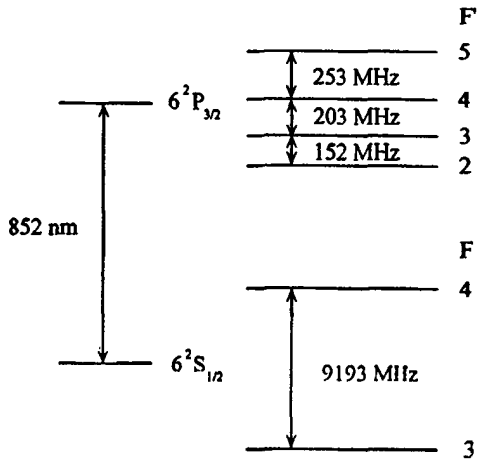


Fig. 2. Cs Energy Level Diagram.

In the optical pumping region virtually all of the atoms are pumped into the $F = 3$ ground state by a laser which is tuned to the optical transition $F = 4 \rightarrow F' = 3$. The atoms then traverse the Ramsey cavity and those that undergo a transition to the $F = 4$ ground state are detected by a probe laser which is tuned to the optical cycling transition $F = 4 \rightarrow F' = 5$. The resulting fluorescence is collected and detected by a photodiode.

Both the optical pumping and probe laser beams are derived from a single external-cavity $Al_xGa_{1-x}As/GaAs$ diode laser. The laser is frequency locked to the $F = 4 \rightarrow F' = 5$ cycling transition using saturation spectroscopy in a cesium vapor cell. The output passes through an acousto-optic modulator (AOM) where the undiffracted beam serves as the probe beam. When RF power is applied to the AOM a diffracted beam is generated that is shifted down in frequency by an amount equal to the frequency separation of the $F' = 3$ and $F' = 5$ levels in the excited state. This beam serves as the optical pumping beam.

To implement the gated optical pumping method, no microwave field is applied to the cavity; the time-of-flight distribution is recorded using lasers only. In steady-state, when the optical pumping beam is on, virtually all of the atoms are pumped into the $F = 3$ hyperfine level. Since no microwave transitions take place, there exists no significant fluorescence signal from the probe region. The atomic transit-time distribution is measured by switching off the optical pumping beam for a time interval τ using the computer-controlled RF switch connected to the AOM. During this time interval, a small group of atoms (uniformly distributed among the $F = 3$ and $F = 4$ hyperfine levels) are allowed to transit the optical pumping region unpumped. When the optical pumping beam is switched back on, subsequent atoms are transferred to the lower ground state. This short burst of unpumped atoms spreads spatially as it propagates down the beam tube due to the distribution of atomic velocities. These atoms then arrive at the detection beam at times determined by their velocities. The resulting fluorescence-pulse represents the atomic transit-time distribution.

We adjust τ so that the slowest atoms of interest have sufficient time to traverse the optical pumping region. This sets a lower limit on τ. Increasing τ also increases the number of atoms contributing to the measurement; thus the signal to noise ratio improves. However, large values of τ reduce the resolution of the transit-time measurement.

The output of the photodiode amplifier represents the fluorescence-pulse convolved with the impulse response of the photodiode and its associated amplifier. To recover the shape of the fluorescence-pulse, the effects of the photodetector's transfer function must be removed. Figure 3 shows the measured frequency response of the photodetection system.

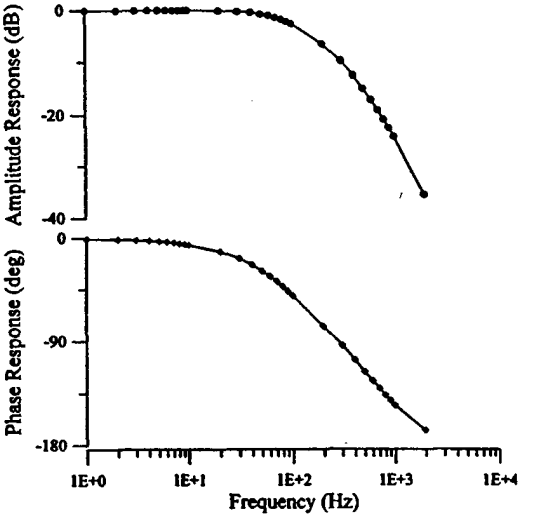


Fig. 3. Frequency Response of the Photodetection System.

The transfer function was determined by measuring the amplitude and phase response of the photodetection system to a sinusoidal optical excitation. The cutoff frequency of $\approx$ 100 Hz is largely determined by the 10 GΩ feedback resistance used in the photodetector's transimpedance amplifier. The spectrum of the photoelectric pulse was corrected for the non-

uniform amplitude and phase delay characteristics of the photodetector and its amplifier. Figure 4 shows a photoelectric pulse before and after this deconvolution operation was performed.

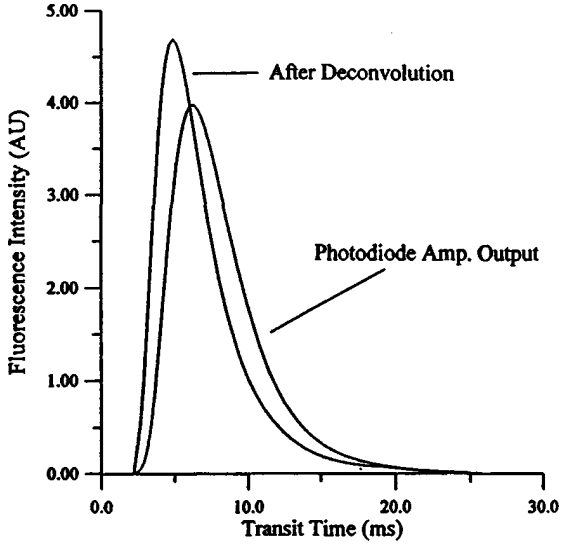


Fig. 4. Raw and Deconvolved Photoelectric Pulses. τ = 500 microseconds. 4000 pulses were averaged for this measurement.

The rise time of the pulse has been shortened significantly with the restoration of the high frequency components of the pulse. The time axis of this figure has been normalized so that after deconvolution this pulse represents the distribution of atomic transit-times across the Ramsey cavity, rather than from pump-laser to probe-laser. The velocity distribution $\sigma(v)$ is obtained from the transit-time distribution $\rho(T)$ by the relation

$$\sigma(v) = (L/v^2)\rho(L/v), \qquad (1)$$

where v is the atomic velocity and L is the length of the Ramsey cavity. The corresponding second-order Doppler correction for this distribution under normal operating conditions is $-3.60 \cdot 10^{-13}$.

Atomic transit-time distributions from the gated optical pumping technique were compared to those obtained using a Ramsey fringe inversion method.[4] This method is optimized for use with long laboratory standards where the excitation length is much less than the drift region length. The transit-time distribution is obtained from a set of Ramsey lineshapes recorded at various microwave power levels. This data is collected using the digital servo of figure 5.

A computer controls both the microwave frequency and power level while monitoring the beam fluorescence from the probe region. A computer program then sweeps the microwave frequency and records the Ramsey lineshape, incrementing the microwave power between sweeps. Figure 6 shows a typical set of Ramsey lineshapes.

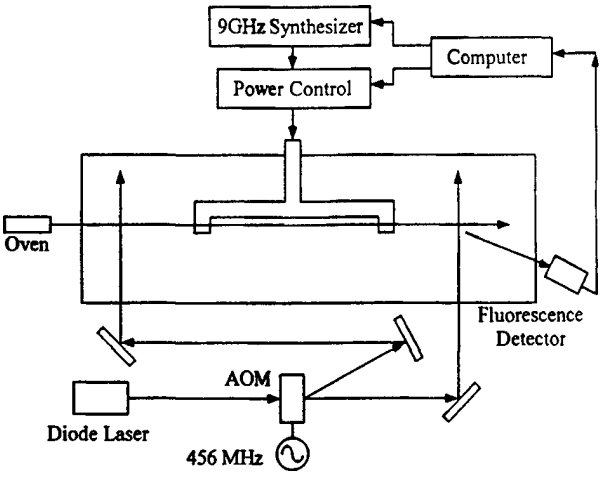


Fig. 5. Schematic Diagram of the Digital Servo.

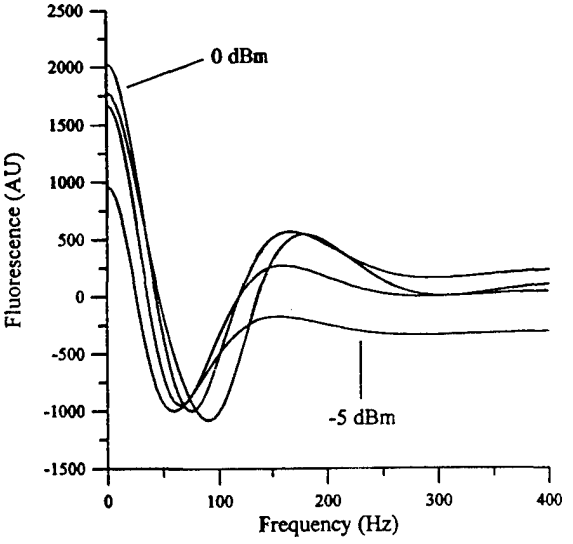


Fig. 6. Ramsey Lineshapes as a Function of Microwave Power.

The corresponding transit-time distribution is shown in figure 7.

The second-order Doppler correction for this distribution under normal operating conditions is $-3.58 \cdot 10^{-13}$, in excellent agreement with the results obtained from gated optical pumping.

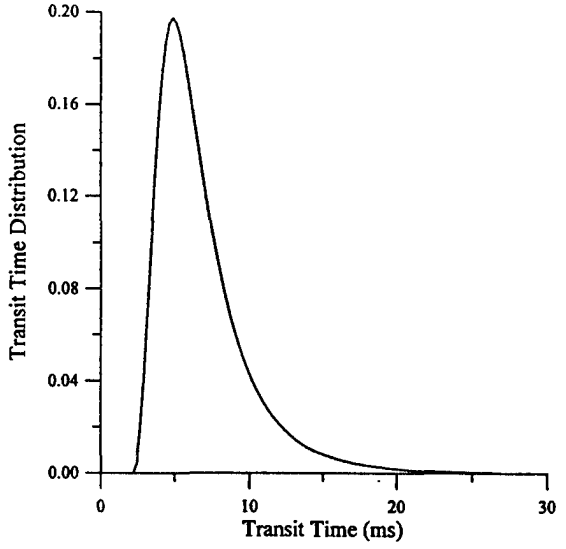


Fig. 7. Transit-time Distribution from Ramsey Fringe Inversion.

III. Conclusion

During the evaluation of the frequency accuracy of NIST-7, we desire two or more independent techniques for estimating the major frequency biases present in the standard. We have developed two techniques for measuring the velocity distribution, required for computing the second-order Doppler correction: gated optical pumping and Ramsey fringe inversion. These two techniques yield velocity distributions in excellent agreement. The second-order Doppler correction was computed for several microwave power levels using transit-time distributions obtained from the two methods. While the corrections varied by nearly 2 parts in 10^{13} over a 7.5 dB microwave power range, the computed corrections from the two techniques differed by no more than $2 \cdot 10^{-15}$.

References

[1] J. S. Boulanger, "A New Method for the Determination of Velocity Distributions in Cesium Beam Clocks," *Metrologia*, vol. 23, pp. 37-44, 1986.

[2] H. Hellwig, S. Jarvis, Jr., D. Halford, and H. E. Bell, "Evaluation and Operation of Atomic Beam Tube Frequency Standards Using Time Domain Velocity Selection Modulation," *Metrologia*, vol. 9, pp. 107-112, 1973.

[3] S. Jarvis, "Determination of Velocity Distributions in Molecular Beam Frequency Standards from Measured Resonance Curves," *Metrologia*, vol. 10, pp. 87-98, 1974.

[4] J. H. Shirley, "Velocity Distributions From the Fourier Transforms of Ramsey Line Shapes," in *Proc. 43rd Ann. Symp. on Freq. Control*, 1989, pp. 162-167.

[5] R. E. Drullinger, "Optically Pumped Primary Frequency Standards," in *Proc. 44th Ann. Symp. on Freq. Control*, 1990, pp. 76-81.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A Hybrid Digital/Analog Servo for the NIST-7 Frequency Standard

J.P. Lowe, W.D. Lee, F.L. Walls, R.E. Drullinger

Time and Frequency Division

National Institute of Standards and Technology

325 Broadway

Boulder, CO 80303

Abstract

This paper describes a hybrid (analog/digital) approach to the servo electronics for the NIST-7 cesium-beam, frequency standard. A digital servo system has been added to the existing analog system to enhance the capabilities of both servos. The analog servo allows a 5 MHz oscillator to be frequency locked directly to the atomic transition frequency. The digital servo reduces the frequency perturbations due to changes in the ambient magnetic field. Also, the existence of separate servo systems allows independent measurements of systematic errors. Design specifications are given and stability measurement results are presented.

Introduction

A preliminary evaluation of the NIST-7 optically pumped, primary frequency standard yielded a preliminary evaluation with an uncertainty of a few parts in 10^{14}.[1] That evaluation was completed with an analog servo system specifically designed for NIST-7. In order to obtain the full design uncertainty of one part in 10^{14} and adequately address all potential systematic errors, several new servo systems, including digital systems, have been developed. All of these systems will be used to investigate the various systematic errors. The digital and analog systems have been combined to form a hybrid servo that uses the best features of both.

Analog Servo

The analog electronic system for NIST-7 employs sine-wave phase modulation, square-wave demodulation and a second-order integration filter. The design of the beam-tube produces a clock-signal line Q of just over 10^8. To meet the design uncertainty goal of 1×10^{-14} and, thus, a line-center resolution of 10^6, a new modulation technique with second harmonic distortion below -120 dBc was developed.[2] The microwave radiation is synthesized by direct multiplication from the local oscillator at 5.006880 MHz (see Fig. 1). The multiplier is characterized by a phase noise $\mathcal{L}(1\ \text{Hz}) = -140$ dBc/Hz, a noise floor $S_\phi \leq -170$ dBc/Hz, 5 MHz sidebands < -40 dBc and sidebands from the power mains < -55 dBc. The unbalance of the power main sidebands has been shown to be less than -100 dBc. The spectral purity of the microwave radiation and the low second-harmonic distortion in the modulator lead to errors much less than 10^{-14}. The DC offsets in the electronics have been measured by changing the AC and DC gain by as much as 2. No measurable frequency shifts were observed at the 10^{-14} level.

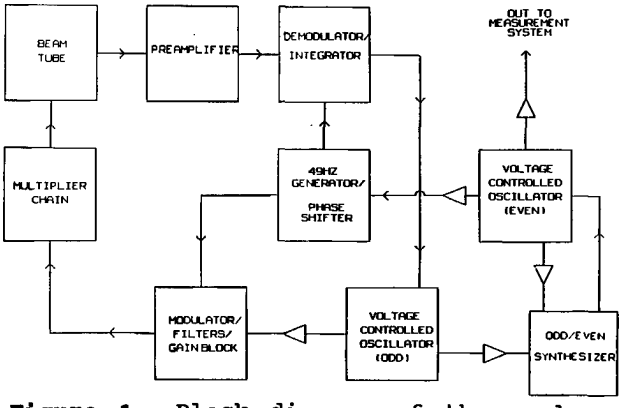


Figure 1 Block diagram of the analog servo system.

Digital Servo

Digital synthesizers have been adapted for use in frequency control servos for cesium frequency standards.[3] The agility of a digital servo allows parameters such

Contribution of the U.S. Government, not subject to copyright.

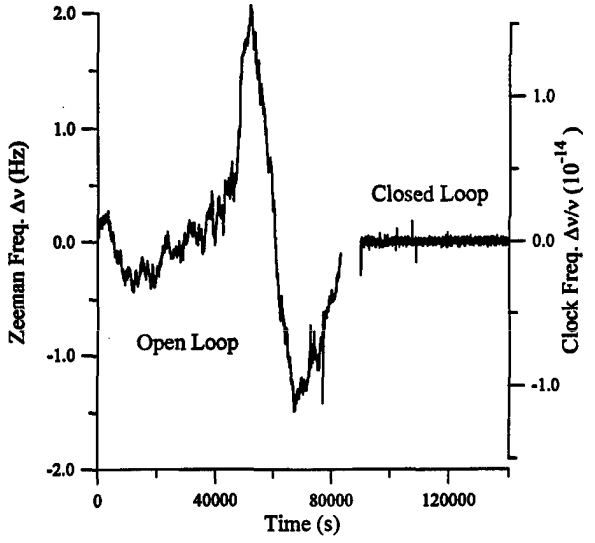


Figure 2 The frequency fluctuations of the Zeeman transition under open-loop and closed-loop conditions.

as the C field to be under closed-loop control. The NIST-7 digital servo uses a hydrogen maser as a reference frequency. A low phase-noise synthesizer developed at NIST[4] is used to generate the microwave interrogation frequency. A slow square-wave modulation technique is used to probe the central Ramsey peak and determine line-center. The difference between the line-center frequency and the maser frequency is logged by a computer for subsequent analysis. Care has been taken to reduce frequency pulling introduced by spurious components of the microwave spectrum that are coherent with the slow square-wave modulation. NIST-7 employs a longitudinal C-field. Consequently, there is a slight sensitivity to changes in the ambient magnetic field due to the reduced effectiveness of the magnetic shields in the longitudinal direction. Figure 2 shows the variation of the Zeeman frequency (and the resulting fractional frequency change of the clock transition) due to changes in the ambient magnetic field. For times less than about 80 000 seconds, the figure shows typical changes in the Zeeman frequency due to a diurnal variation in the local east-west component of the terrestrial magnetic field. This variation is due to an interaction of the solar wind with the Earth's ionosphere. These data show that this diurnal fluctuation in the ambient magnetic field can give rise to a fractional frequency change of a couple parts in 10^{14}. This is twice the uncertainty goal for NIST-7. These frequency fluctuations have been reduced

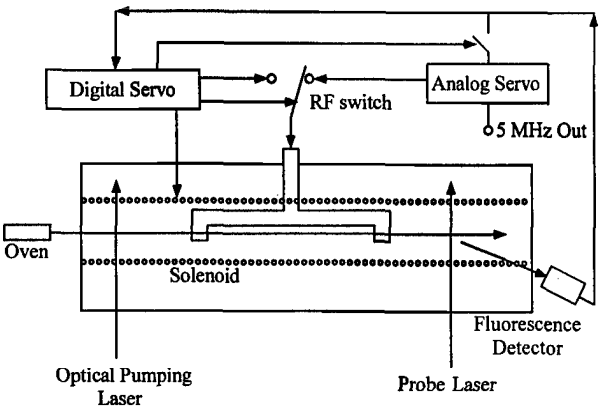


Figure 3 Block diagram of the hybrid servo system.

using a C-field servo, which was implemented using the computer-controlled digital synthesizer. The frequency agility of the digital servo allows monitoring of a field-sensitive transition. The computer uses these data to provide closed-loop control of the C field by adjusting the current in the C field solenoid. Figure 2 shows the performance of the C-field servo for times greater than about 90 000 seconds. Fractional frequency changes in the clock transition due to fluctuations in the ambient magnetic field have been reduced to a few parts in 10^{15}.

Hybrid Servo

The analog servo has a higher modulation rate and, thus, is less sensitive to 1/f noise present in the system than the slow digital servo. The digital servo has the ability to provide closed-loop control of the C-field, which is not possible for the present analog servo system. Therefore, a hybrid servo system has been developed. Figure 3 shows the hybrid servo block diagram. A switch controls the RF input to the Ramsey cavity, selecting either the analog or the digital servo. The output power level of these two sources has been closely matched. First, the analog servo interrogates the central Ramsey peak and steers a 5 MHz oscillator to its center. Next, the digital servo interrogates the Zeeman frequency and controls the C field. Finally, the analog servo regains control and continues to probe the Ramsey peak. The 5 MHz oscillator is free running during the digital interrogation interval; therefore, the digital servo should perform its operation in a time that is small compared to the interrogation interval of the analog servo. The time required by the digital servo is less than one second, which does not significantly degrade the stability

of the quartz oscillator. Both the analog and digital portions of the hybrid servo monitor the atomic beam fluorescence. The digital servo does not digitize this signal while the analog servo is operating. However, the analog servo must be electronically disconnected from the beam tube while the digital servo is running. This is necessary because the fluorescence generated by the slow square-wave frequency modulation of the digital servo can degrade the open-loop stability of the analog servo. Similarly, switching transients occurring during the removal of the fluorescence signal can degrade open-loop stability. This is a result of the design of the dual-integrator, frequency-control loop. A switching transient can generate a non-zero error (briefly) which drives the output of the first integrator to some non-zero, constant voltage. Since the analog servo is open loop, the output of the second integrator will grow linearly with time, introducing a linear frequency drift. The upper trace of figure 4 shows the frequency drift of the analog servo under open-loop conditions when the fluorescence signal is disconnected at some arbitrary point in its modulation cycle. The slope of the linear frequency drift can be reduced dramatically if the fluorescence switch timing is phase-locked to the 49 Hz phase modulation clock of the analog servo. Figure 5 shows the timing diagram of the improved system. The time delay, Δt, from the falling edge of a synchronization marker

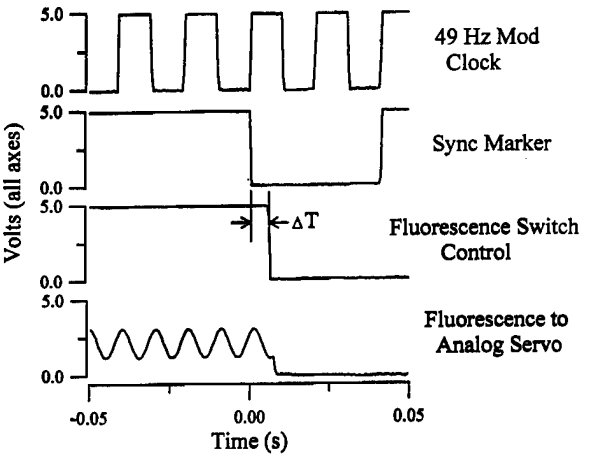


Figure 5 The timing diagram for the hybrid servo switch controls.

to the activation of the fluorescence switch can be adjusted with a resolution of 100 microseconds. This permits fine adjustment of the switching transients and the virtual elimination of the frequency drift of the analog servo while open-loop. The lower trace of figure 4 shows that the residual frequency fluctuations of the analog servo have been reduced to nearly the noise floor of the measurement system. Furthermore, figure 6 indicates that the implementation of the hybrid servo has not significantly degraded the short-term frequency stability of the analog servo.

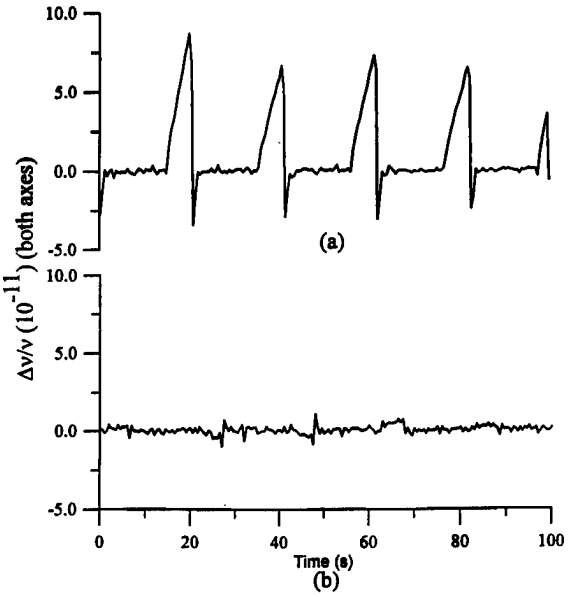


Figure 4 The fractional frequency deviations due to
a) asynchronous and
b) synchronous switching.

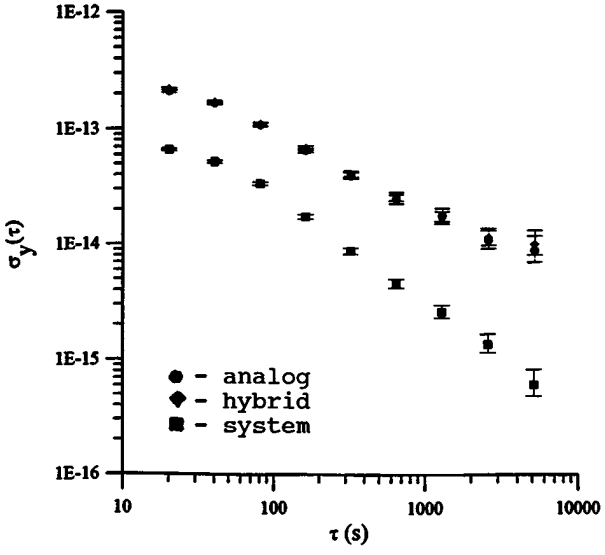


Figure 6 The stability diagram for the hybrid servo.

Conclusion

Using a digital frequency control servo, we have found that active control of the C field is necessary in order to achieve the frequency uncertainty goal for NIST-7. A hybrid digital/analog servo has been designed and implemented. This servo suppresses frequency fluctuations due to the second order Zeeman effect to a few parts in 10^{15} without significantly degrading short-term stability. This hybrid servo also interrogates the central Ramsey peak at a high modulation rate, thus avoiding potential 1/f noise. The addition of the digital servo to the existing analog servo system has reduced the uncertainty and has helped facilitate the systematic error analysis.

References

[1] R.E. Drullinger, J.H. Shirley, J.P. Lowe, and D.J. Glaze, "Error Analysis of the NIST Optically Pumped Primary Frequency Standard", IEEE Transactions on Instrumentation and Measurements, vol. 42, no. 2, pp. 453-456, April 1993

[2] J.P. Lowe, and F.L. Walls, "Ultralinear Small-angle Phase Modulator", Proceedings of the 5th European Frequency and Time Forum, pp. 461-464, 1991

[3] L.S. Cutler, and R.P. Giffard, "Architecture and Algorithms for New Cesium Beam Frequency Standard Electronics", Proceedings of the IEEE Frequency Control Symposium, pp. 127-133, 1992

[4] C.W. Nelson, F.L. Walls, F.G. Ascarrunz, and P.A. Pond, "Progress on Prototype Synthesizer Electronics for (199)Hg", Proceedings of IEEE Frequency Control Symposium, pp.64-69, 1992

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

HIGH-FREQUENCY OSCILLATORS USING PHASE-LOCKED ARRAYS OF JOSEPHSON JUNCTIONS*

S. P. Benz and P. A. A. Booi,
National Institute of Standards and Technology, Division 814.03,
325 Broadway, Boulder, CO 80303

Abstract--We present a basic description of Josephson junctions and discuss their use as GHz and THz oscillators. The resistively shunted junction model is used to calculate the available power, linewidth and operating frequency of the oscillators. We discuss how phase-locked arrays of junctions are used to achieve higher power and narrower linewidth. Two experimental examples of phase-locked emission are shown: one from on-chip detection circuits at 150 GHz and one detected off-chip showing a 13 kHz linewidth at 88.8 GHz.

I. INTRODUCTION

We are interested in developing the technology to synthesize and measure sub-THz and THz frequencies with the same ease and precision now available in the microwave region. This is important for bridging the frequency band between the microwave and optical regions. This effort is intended to support and encourage emerging instrumentation in high-speed optical signal processing and precision optical metrology including standards for frequency and length.

We believe this frequency band can be filled by exploiting the intrinsic high-frequency properties of superconductive Josephson-junction devices. Josephson junctions and circuits are naturally suited to sub-THz and THz frequency operation, because they intrinsically generate currents at frequencies up to a few times the gap frequency of the superconducting material. Low-temperature superconductor (LTS) devices are typically made from Nb, a refractory metal with a superconducting transition temperature T_c of 9 K, and are operated at 4 K. Nb junctions have a gap frequency of about 1.4 THz. LTS devices have been demonstrated to operate at 8 THz.[1] High-temperature superconductor (HTS) devices, such as YBaCuO, have $T_c \approx 90$ K, allowing 77 K liquid nitrogen operation. HTS devices have demonstrated operation at 8 THz,[2] and with a gap frequency of approximately 15 THz, are expected to perform frequency translation functions up to at least 30 THz.

By developing small, lithographically-fabricated devices based on Josephson junctions we hope to achieve an unprecedented level of stability and reliability over previous technologies based on gas lasers and whisker-contacted diodes. Lithography also enables the integration of multiple devices into circuits in order to perform more complex frequency translation functions. LTS fabrication technology is well-developed. However, HTS multilayer thin-film fabrication is still under development.

* U.S. Government work, not subject to U.S. copyright.

Many different devices can be implemented using Josephson technology, such as mixers, oscillators, detectors, and logic elements. This paper will focus on Josephson oscillators and their application to the submillimeter (> 300 GHz) and THz frequency regimes. We present experimental results from LTS devices that were designed to demonstrate the fundamental principles of Josephson oscillators at accessible microwave frequencies.

II. PRINCIPLES OF JOSEPHSON OSCILLATORS

We use the zero-capacitance, resistively-shunted junction (RSJ) model[3,4] to compare Josephson oscillators based on single junctions and arrays of junctions. Our discussion will be limited to nonresonant junctions, so no comments will be made on flux-flow oscillators.[5] Furthermore, we assume lumped arrays having physical dimensions smaller than 1/8 of the emission wavelength. Finally we assume that ground planes are used with all circuits, preventing self-field coupling between devices. Within this model, we discuss specific design criteria required to achieve impedance matching, high output power and narrow linewidth at a given operation frequency.

A. Single Junctions

A RSJ is characterized by a critical current I_c and resistance R, and is biased with a dc current source I. The equations describing a RSJ are

$$I = \frac{\Phi_0}{2\pi R}\dot{\phi} + I_c \sin\phi, \tag{1a}$$

$$V = \frac{\Phi_0}{2\pi}\dot{\phi}, \tag{1b}$$

where $\phi(t)$ is the macroscopic phase difference across the junction, $\Phi_0 = h/2e \approx 2.07$ mV/ THz is the magnetic flux quantum, V is the instantaneous voltage across the junction, and the overdot denotes differentiation with respect to time.

Equations (1) can be solved analytically. Figure 1(a) shows the dc I-V curve of a RSJ. For $I \le I_c$, the junction is in the zero-voltage state, while for $I > I_c$, a voltage develops across the junction. In the voltage state, the supercurrent $I_c \sin\phi$ of a junction oscillates at a well-defined fundamental frequency $f = V_{dc}/\Phi_0$. This *ac Josephson effect,* where the junction supercurrent oscillates at $1/\Phi_0 \approx 483$ GHz per mV of bias voltage, means that the

junction is intrinsically a high-frequency dc-to-ac converter and can be used as a voltage-controlled oscillator.

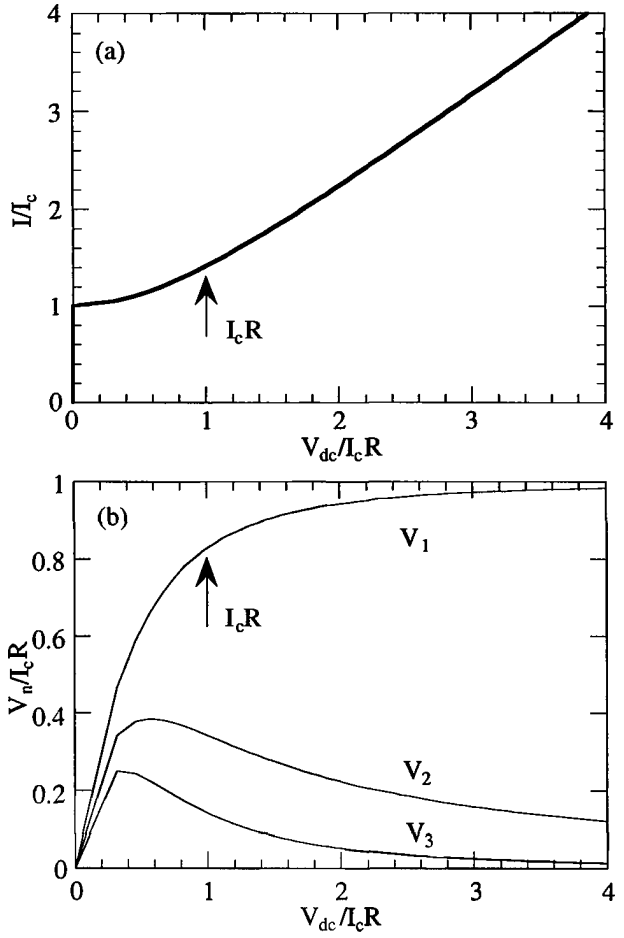


Fig. 1. (a) I-V curve of a RSJ. (b) Harmonic amplitudes V_n as a function of bias voltage.

Figure 1(b) shows the amplitude of the fundamental Josephson oscillation and the first two higher harmonics as a function of voltage. When the junction is biased at voltages greater than the characteristic voltage $V_c \equiv I_cR$, the time-dependent voltage is nearly sinusoidal because the amplitudes of the higher harmonics V_n ($n \geq 2$) are much smaller than the amplitude of fundamental. The peak voltage amplitude at the fundamental frequency approaches a maximum $V_1 \approx V_c$ for $V_{dc} \geq V_c$, and the junction impedance becomes resistive, $Z \approx R$. In order to maximize the fundamental power and minimize the contribution from harmonics, the characteristic frequency, $f_c = V_c/\Phi_0$, should be fixed at the minimum operating frequency and the device should be operated at $V_{dc} \geq V_c$. Under these conditions the *rms available power* from the junction is

$$P_{JJ} = \frac{V_c^2}{8R} = \frac{I_cV_c}{8}. \qquad (2)$$

If we assume that all Josephson junction technologies have a maximum critical current, then the available power is limited by the maximum I_c and the minimum operating frequency. Typically it is assumed that I_c is limited to a few milliamperes by nonuniform current flow due to self-induced magnetic fields.[6] The characteristic voltage is limited for a given technology by the gap frequency of the superconducting material. Typical Nb junctions have a maximum characteristic voltage $V_c \approx 2.8$ mV. Under these *optimal* conditions, the available power from a single junction could approach $P_{JJ} \approx 1$ μW for operation above 1.4 THz. Lower minimum operating frequencies will yield lower power.

If the junction is coupled to a resistive load R_L and $V_{dc} > V_c$, then the most efficient power transfer occurs when the junction and load resistances are matched ($R = R_L$). The rms power delivered to this *matched load* is then equal to the available power. Demanding $R_L = 50\ \Omega$ load and $f_c = 1.4$ THz limits the critical current to only $I_c = V_c/R_L = 58$ μA. The power delivered to a matched 50 Ω load for this example is only 20 nW. In the highest frequency demonstration of a single junction Josephson oscillator, a tunnel junction delivered 10 nW at 1 THz to another junction on the same chip.[7]

Many oscillator applications require linewidths < 100 kHz. The theoretical linewidth (full width at half-maximum power) at the fundamental frequency for a single junction is[8,9]

$$\Delta f_{JJ} = \frac{\pi R_d^2 S_I(0)}{\Phi_0^2} \approx \left(\frac{41\ \text{MHz}}{\Omega K}\right)\frac{TR_d^2}{R}, \qquad (3)$$

where $R_d = dV_{dc}/dI$ is the dynamic resistance of the junction, $S_I(0) = 4kT/R$ is the Johnson current noise spectral density, k is Boltzmann's constant, and T is temperature. With a typical $R \approx R_d \approx 1\ \Omega$ at T = 4 K, we obtain a ~160 MHz linewidth for a single junction, which is too wide for most applications.

Applications of single junction oscillators have been difficult to implement because they have low output power (10 nW) and their low (≤ 1 Ω) impedances do not match to typical high-impedance loads. In order to achieve useful power (> 0.1 mW) and impedance matching, arrays of phase-locked junctions are needed for most applications. One-dimensional (1D)[10,11] and two-dimensional (2D)[12-17] arrays of phase-locked junctions offer potential solutions to these problems. When the junctions in an array are phase-locked, arrays can provide higher output power with narrower linewidths to a matched load impedance. However, phase-locking is difficult because of the complex dynamics of these multi-dimensional systems.

B. Phase-locked Arrays

Arrays of junctions have better impedance matching, higher power, and narrower linewidth than single junctions. Figure 2 shows the geometry of a 2D array. The array impedance depends on its width and length, M

columns each with N junctions in the voltage state. The available power from a phase-locked 2D array in the sinusoidal operating regime ($V_{dc} \geq V_c$ for each junction), is proportional to the number of phase-locked junctions, $P_A = MNI_cV_c/8$. The length-to-width ratio can be chosen to achieve impedance matching $N/M = R_L/R$, so that the power delivered to a matched load is[12,13]

$$P^L_A = \frac{NMI_cV_c}{8} = \frac{M^2I_c^2R_L}{8}. \quad (4)$$

With V_c fixed and I_c maximized (2.8 mA for Nb junctions), the total power delivered to a matched load is proportional to the square of the number of columns, M. This gives 2D arrays an additional degree of freedom over 1D arrays to increase and tune the power delivered to a fixed load. For the 1.4 THz minimum operating frequency (V_c = 2.8 mV), the power can be tuned to 4.9 mW by choosing M = 10 and N = 500 to match the array to a 50 Ω load. This delivered power is sufficient for both on-chip or off-chip applications, and the 5000-junction integration is achievable with present fabrication technology. Even higher power levels can be achieved by increasing the number of junctions accordingly.

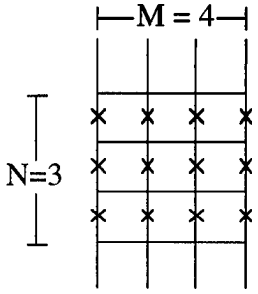


Fig. 2. Schematic layout of a 4 x 3 array. The crosses depict Josephson junctions. The junctionc are connected by superconducting wires with inductance.

For phase-locked arrays of junctions, the linewidth is reduced by the total number of phase-locked junctions[6,13]

$$\Delta f_A = \frac{\Delta f_{JJ}}{MN}. \quad (5)$$

Arrays with M = 10 and N = 500 would have a linewidth of ~33 kHz at 4 K which is sufficiently narrow for most applications.

III. EXPERIMENTAL RESULTS

The first demonstration of complete phase locking in 2D arrays was accomplished by using single Josephson junctions as on-chip detectors.[12-13] Emission was coupled from a 2D array to a detector junction through a dc blocking capacitor. The capacitor enabled the array and the detector junction to be biased independently. When the array was biased into the voltage state, emission was coupled to the detector causing the detector's intrinsic Josephson oscillations to lock to the array's emission. This phenomenon is observed in the detector's I-V curve as flat voltage steps at voltages $V_{dc} = n\Phi_0 f$, where f is the drive frequency and n = 0, 1, 2, ... is the step number. These steps are called Shapiro steps. The number of steps and their widths depend on the amplitude of the drive.

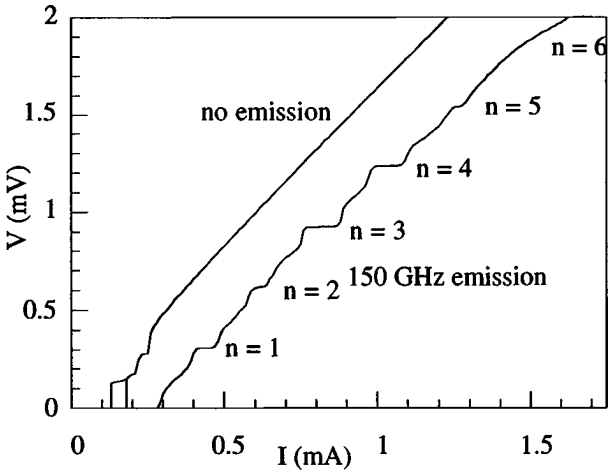


Fig. 3. Detector I-V curve displaying Shapiro steps.

Figure 3 shows the detector junction I-V curve for a 10-by-10 junction 2D array coherently emitting at 150 GHz.[12,13] Since the 0.4 μW coupled to the detector junction at this frequency agrees with the expected theoretical power, all 100 junctions in the array are phase locked. This array showed Shapiro steps at voltage tunable frequencies ranging from 60 to 210 GHz, indicating broad band tunability.

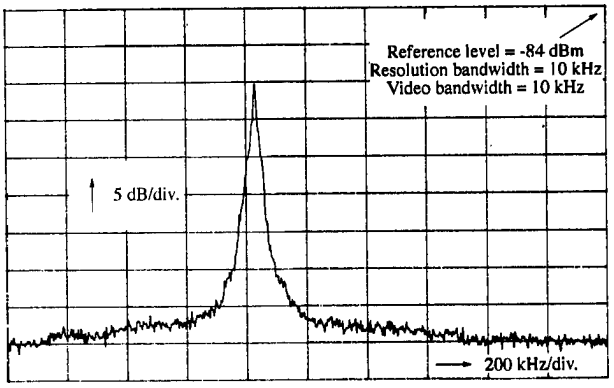


Fig. 4. Emission peak after 100 video averages at 88.844 GHz showing a ~13 kHz linewidth.

More recently,[15] the emission from 10 x 10 arrays has been coupled off-chip to room-temperature mixers through a fin-line antenna and WR-12 waveguide. A single voltage-tunable peak was detected in the frequency range from the 53 GHz cut-off frequency of the waveguide up to 230 GHz. Figure 4 shows an example spectral peak with a linewidth of ~13 kHz.

These results were the first off-chip linewidth and

power measurements of phase-locked 2D arrays where a detailed comparison was made with theory. It was found that a stripline resonance in the antenna reduced the array dynamic resistance and thereby the emission linewidth [see Eq. (3)] from the 2 MHz linewidth expected for a non-resonant-coupled array. From 84 to 90 GHz, the linewidth varied approximately one order on magnitude.

IV. CONCLUSIONS

Josephson oscillators are voltage tunable with potential emission at GHz and THz frequencies. LTS 2D arrays have demonstrated complete phase locking and narrow linewidths over broadly-tunable microwave frequencies. HTS technology is desirable for Josephson oscillators because the higher intrinsic frequencies would enable higher-frequency oscillators. The impact of HTS Josephson devices on precision frequency synthesis and measurement depends largely upon improvement of HTS thin-film fabrication technology.

ACKNOWLEDGMENT

This research is supported in part by the Office of Naval Research contract No. N00014-92-F-0040, and by the BMDO Office of Innovative Science and Technology under contract number F30602-93-F0121 with technical program management from Rome Laboratory.

REFERENCES

[1] D. G. McDonald, V. E. Kose, K. M. Evenson, J. S. Wells, and J. D. Cupp, "Harmonic generation and submillimeter wave mixing with the Josephson effect," *Appl. Phys. Lett.*, vol. 15, pp. 121-122, Aug. 1969.

[2] 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.*, vol. 63, pp. 1984-1986, Apr. 1993.

[3] W. C. Stewart, "Current-voltage characteristics of Josephson junctions," *Appl. Phys. Lett.*, vol. 12, pp. 277-280, Apr. 1968.

[4] D. E. McCumber, "Effect of ac impedance on dc voltage-current characteristics of superconducting weak-link junctions," *J. Appl. Phys.*, vol. 39, pp. 3113-3118, June 1968.

[5] K. Yoshida and F. Irie, "Frequency conversion in a long Josephson junction with a moving vortex array," *Appl. Phys. Lett.*, vol. 27, pp. 469-470, Oct. 1975.

[6] A. K. Jain, K. K. Likharev, J. E. Lukens, and J. E. Sauvageau, "Mutual phase-locking in Josephson junction arrays," *Phys. Rep.*, vol. 109, pp. 309-426, 1984.

[7] R. P. Robertazzi and R. A. Buhrman, "Josephson terahertz local oscillator," *IEEE Trans. Magn.*, vol. 25, pp. 1384-1387, Mar. 1989.

[8] A. I. Larkin and Yu. N. Ovchinnikov, "Radiation line width in the Josephson effect," *Sov. Phys. JETP*, vol. 26, pp. 1219-1221, June 1968.

[9] M. J. Stephen, "Noise in the ac Josephson effect," *Phys. Rev.*, vol. 182, pp. 531-538, June 1969.

[10] B. Bi, S. Han, and J. E. Lukens, "Radiation linewidth of phase-locked distributed array in the submillimeter wave range," *Appl. Phys. Lett.*, vol. 62, pp. 2745-2747, May 1993.

[11] S. Han, B. Bi, W. Zhang, and J. E. Lukens, "Demonstration of Josephson effect submillimeter wave sources with increased power," *Appl. Phys. Lett.*, vol. 64, pp. 1424-1426, Mar. 1994.

[12] S. P. Benz and C. J. Burroughs, "Coherent emission from two-dimensional Josephson junction arrays," *Appl. Phys. Lett.*, vol. 58, pp. 2162-2164, Apr. 1991.

[13] S. P. Benz and C. J. Burroughs, "Two-dimensional arrays of Josephson junctions as voltage-tunable oscillators," *Proceedings of the Third International Superconductive Electronics Conference*, Glasgow, pp. 230-237, June 1991; *Supercon Sci. Technol.*, vol. 4, pp. 561-567, 1991.

[14] P. A. A. Booi, S. P. Benz, T. Doderer, D. Hoffmann, J. Schmidt, S. Lachenmann, and R. P. Huebener, "Frequency dependence of the emission from 2D array Josephson oscillators," *IEEE Trans. Appl. Supercon.*, vol. 3, pp. 2493-2495, Mar. 1993.

[15] P. A. A. Booi and S. P. Benz, "Emission linewidth measurements of two-dimensional array Josephson oscillators," *Appl. Phys. Lett.*, vol. 64, pp. 2163-2165, Apr. 1994.

[16] K. Wiesenfeld, S. P. Benz, and P. A. A. Booi, "Phase-locked oscillator optimization for arrays of Josephson junctions," *J. Appl. Phys.*, to appear Sept. 1994.

[17] P. A. A. Booi and S. P. Benz, "Resonances in two-dimensional Josephson-junction arrays," *J. Appl. Phys.*, submitted, 1994.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ATOMIC HYDROGEN SPIN-EXCHANGE FREQUENCY SHIFT CROSS SECTION AT 0.5 KELVIN

M. E. Hayden,* M. D. Hürlimann,† and W. N. Hardy

Department of Physics, University of British Columbia, Vancouver, B.C. CANADA V6T 1Z1

We present an accurate measurement of the dominant hydrogen atom spin-exchange frequency shift cross section at 0.5 K, made using a recirculating cryogenic hydrogen maser. This parameter is sensitive to the detailed form of the H-H interaction potentials, particularly at large interatomic separations, and thus provides a means for testing the accuracy of ab-initio potentials. We find $\overline{\lambda}_0 = -21.7 \pm 2.8$ Å^2, *whereas the most sophisticated theoretical calculations to date give* $\overline{\lambda}_0 = -11.8$ Å^2.

Spin-exchange collisions broaden and shift the atomic hyperfine transition upon which the operation of a hydrogen (H) maser is based, thereby coupling fluctuations in the atomic collision rate to the maser oscillation frequency f. A key development in the evolution of modern H masers was the introduction of 'spin-exchange tuning' whereby f is made insensitive to changes in the atomic collision rate by detuning the maser microwave cavity in a prescribed manner [1, 2]. Tuned in this fashion, state-of-the-art H masers have attained frequency stabilities better than 1 part in 10^{15} over time periods of several hours [3].

The success of the spin-exchange tuning procedure relies on the spin-exchange induced frequency shift being proportional to the population inversion of the masing states within the maser cavity [1, 2, 4, 5]. Several years ago, a careful reexamination of the H atom spin-exchange process by B. J. Verhaar and a number of collaborators [6, 7] revealed that when hyperfine interactions are properly accounted for, the resulting frequency shifts actually depend on the detailed occupancy of the individual hyperfine states of the colliding atoms. The implication of this is that spin-exchange tuning cannot be as efficient as it was once thought to be, particularly at low temperatures where atomic collision energies are comparable to the hyperfine interaction energy. This prediction is substantiated, at least in part, by two recently reported experiments, one at room temperature in a conventional H maser [8, 9], and the other at 0.5 K in a low temperature, superfluid liquid helium (ℓ-^{4}He) lined H maser [10, 11]. In both cases a frequency shift *beyond* that predicted by classical[1] theories of spin-exchange was observed. This effect has been interpreted as being a direct consequence of hyperfine interactions during spin-exchange collisions and we thus refer to it as the *hyperfine-induced* (*h-i*) frequency shift. We note, however, that there are serious quantitative disagreements between the experimental observations of the *h-i* frequency shift and the most recent theoretical predictions [6, 7]. These discrepancies must be resolved if one hopes to fully understand spin-exchange collisions and their influence on the ultimate stability of an oscillating hydrogen maser.

In this paper we examine the data from the low temperature maser experiment [10, 11] to extract the dominant spin-exchange frequency shift cross section $\overline{\lambda}_0$, which is sensitive to the long range part of the H-H interatomic potentials [4] yet essentially independent of the influence of hyperfine interactions [6, 7] at 0.5 K. By comparing theoretical calculations of $\overline{\lambda}_0$ with our experimental result, one obtains a gauge against which the accuracy of the theoretical potentials can be mea-

*Present address: Department of Condensed Matter and Thermal Physics, Los Alamos National Laboratory, Mail Stop K764, Los Alamos NM 87545, USA.

†Present address: Schlumberger-Doll Research, Old Quarry Road, Ridgefield CT 06877, USA.

[1] We refer to theories which ignore hyperfine interactions as being 'classical' only to distinguish them from more complete spin-exchange theories [6, 7, 12]. Technically, the approximation which is invoked is the degenerate-internal-states or DIS approximation.

0-7803-1945-1/94 $3.00 © 1994 IEEE

sured. Suitable revisions of the existing potentials to bring the theoretical value for $\overline{\lambda}_0$ into agreement with experiment should be reflected in a modification of the predictions for the *h-i* frequency shift. Full details of the experimental results outlined here have been reported elsewhere [10, 13].

Spin-Exchange Frequency Shifts

Hydrogen masers traditionally operate on the $\Delta F = 1$, $\Delta m_F = 0$ hyperfine transition of the ground electronic state of the H atom which occurs at 1420 MHz near zero magnetic field. We adopt the notation whereby the four hyperfine states of the H atom are labeled $|\ a\ \rangle$, $|\ b\ \rangle$, $|\ c\ \rangle$, and $|\ d\ \rangle$ in order of increasing energy in low magnetic fields, and thus we denote the transition of interest the *a-c* transition. The frequency ω at which a hydrogen maser operates is shifted slightly from the unperturbed *a-c* transition by a number of processes. Collisions with the walls of the storage bulb and other one body processes lead to a density independent shift and are not of interest in the present study. We thus reference all frequency shifts to the linewidth independent frequency ω_0 at which radiation is emitted in the limit of zero H density and negligible coupling between the atoms and the microwave cavity.

Spin-exchange collisions further shift the oscillation frequency of a hydrogen maser via two distinct mechanisms. The first of these is a direct frequency shift which is expected to be a function of the distribution ρ of atoms between the $|\ a\ \rangle$ and $|\ c\ \rangle$ states [6, 7]. The second contribution comes about as a consequence of the coupling between the maser cavity and the atomic system. Any detuning of the cavity resonance frequency ω_c with respect to the atomic transition frequency leads to a frequency shift proportional to the atomic linewidth Γ [14]. This effect is generally referred to as 'cavity pulling'. In terms of the dimensionless detuning parameter

$$\Delta = Q_\ell \left(\frac{\omega_c}{\omega} - \frac{\omega}{\omega_c} \right) \approx 2Q_\ell \left(\frac{\omega_c - \omega}{\omega} \right) \qquad (1)$$

where Q_ℓ is the quality factor of the loaded microwave cavity, this frequency shift is simply

$$\delta\omega = \Delta\Gamma\,. \qquad (2)$$

Since spin-exchange collisions contribute to Γ, they indirectly cause ω to be shifted from ω_0 via cavity pulling. For convenience we write Γ as the sum of two broadening terms

$$\Gamma = \Gamma_o + \Gamma_c\,. \qquad (3)$$

Γ_o includes all contributions to the linewidth which do not depend on the atomic density while Γ_c contains all collision dependent contributions. Note that in the low density regime relevant to H masers, one fully expects Γ_c to be proportional to the H atom density n.

The combined frequency shift due to spin-exchange and cavity pulling (*i.e.* the offset $\delta\omega$ between ω and ω_0) for an oscillating H maser can be written[2] [6, 7]

$$\delta\omega = \left[\Delta + \beta\overline{\lambda}_0 \left(1 + \Delta^2\right)\right] \Gamma - \Omega\Gamma_c \qquad (4)$$

where the parameter β characterizes the coupling between the atoms and the cavity. β is given by

$$\beta = \frac{4\overline{v}}{\mu_0 \eta Q_\ell \hbar \left(\gamma_e + \gamma_p\right)^2} \qquad (5)$$

where $\overline{v}$ is the thermally averaged relative speed of the colliding atoms, η is the magnetic filling factor for the storage bulb within the confines of the microwave resonator, and γ_e and γ_p are the gyromagnetic ratios for the electron and the proton respectively.

The two spin-exchange parameters $\overline{\lambda}_0$ and Ω which appear in equation 4 can in principle be accurately calculated from the interatomic potentials. $\overline{\lambda}_0$ is the thermally averaged spin-exchange frequency shift parameter common to both the classical and the fully quantum mechanical spin-exchange theories; it is the parameter upon which we focus our attention in the current article. In the degenerate-internal-states (DIS) approximation where hyperfine interactions are neglected, $\overline{\lambda}_0$ is given by the thermal average of [4]

$$\lambda_0 = \frac{\pi}{2k^2} \sum_{\ell\ \text{even}} (2\ell + 1) \sin\left[2\left(\eta_\ell^T - \eta_\ell^S\right)\right] \qquad (6)$$

where η_ℓ^S and η_ℓ^T are the phase shifts for scattering from the singlet and triplet H_2 potentials, evaluated at the collision energy of two atoms in their centre of mass frame, and the summation is taken over even angular momentum states. While the inclusion of hyperfine interactions fundamentally alters the rotational symmetry of the scattering problem, $\overline{\lambda}_0$ remains essentially unmodified at 0.5 K. It should thus provide a good probe for the interatomic potentials.

The other parameter, Ω, which appears in equation 4 is the dimensionless *h-i* spin-exchange frequency shift

[2] In writing equation 4 it is implicitly assumed that the *only* contribution to Γ_c is that due to spin-exchange.

parameter which comes about as a direct consequence of the inclusion of hyperfine interactions during spin-exchange collisions. In the DIS approximation Ω is identically equal to zero and therefore there exists a cavity tuning $\Delta' \approx -\beta\overline{\lambda}_0$ for which the maser oscillation frequency becomes independent of the *full* atomic linewidth Γ: this is the classical spin-exchange tuning criterion [1, 2]. Note that in fact the frequency shift $\delta\omega$ is predicted to be equal to zero under these conditions. For $\Omega \neq 0$, it is not possible to decouple $\delta\omega$ from Γ to the same extent. One can define a modified spin-exchange tuning criterion $\Delta'' \approx -(\beta\overline{\lambda}_0 - \Omega)$ for which $\delta\omega$ is independent of the collision dependent part of the linewidth Γ_c, however, ω remains offset from ω_0 by a residual frequency shift $\delta\omega = \Omega\Gamma_c$. Unfortunately Ω is expected to be a complicated function of the hyperfine level populations thus making it a difficult parameter to stabilize in an oscillating maser. Γ_c is likewise a parameter over which one has only imperfect control. In practice it turns out that Ω is small enough at room temperature that spin-exchange tuning is indeed an efficient means for enhancing the stability of a hydrogen maser. Of concern for the operation of cryogenic hydrogen masers (CHM's) below 1 Kelvin is the fact that Ω is considerably larger; this situation poses a real limitation to the ultimate stability one can hope to achieve with a CHM. As previously mentioned, recent measurements of Ω are inconsistent with theory.

Experimental Approach

There is an extensive literature describing the design, operation, and characterization of hydrogen masers [15]. Our studies have been performed using a CHM designed and built at the University of British Columbia (UBC) [16, 17, 18]. This device utilizes a superfluid film of ℓ-^{4}He to suppress the adsorption of H atoms to cold surfaces, and has been successfully operated over the temperature range 230 mK to 660 mK. Its design is unique in that it makes use of the very long recombination lifetime of a gas of H atoms at low temperatures and densities. Rather than eliminating 'spent' atoms which leave the storage bulb of the maser, these atoms are manipulated with dc magnetic field gradients, inverted using microwave fields, and re-injected into the storage bulb. Furthermore, the use of a low temperature discharge source of H atoms [19] allows one to operate the entire device at a relatively uniform temperature, thereby avoiding numerous technical difficulties associated with refluxing ^{4}He atoms driven by temperature gradients [20].

For the work described here, the 1420 MHz resonator of the UBC CHM was outfitted with an electronic tuning system [10, 13] which allows highly reproducible and independent changes to be made to the tuning and the quality factor of the resonator. Measurements of the maser oscillation frequency were made as a function of Q, Δ, and the H atom density in the storage bulb, and compared to the density independent oscillation frequency ω_0. The H atom density was determined by measuring the amplitude of a free induction decay (FID) response to a $\pi/2$ magnetic resonance tipping pulse at the a-c transition frequency after suppressing maser oscillations. ω_0 was obtained by extrapolating the frequency of the FID to zero density.

An essential part of the experiment is the ability to measure the tuning of the resonator. Due to technical constraints it was necessary to measure both the tuning and the quality factor of the microwave resonator by injecting a swept frequency signal and observing the cavity response in reflection. This technique is inherently susceptible to distortion induced by standing waves in the external circuitry, particularly in a cryogenic environment. Rather than rely solely upon these measurements we have employed an alternate means for determining Δ. ***Relative*** changes in Δ are accurately determined from magnetic resonance experiments by examining the initial phase ϕ of a FID with respect to the phase of the local oscillator used to generate the tipping pulse. ϕ is proportional to small changes in Δ [10, 11, 13]; by monitoring changes in ϕ, a measure of changes in the cavity tuning as experienced by the radiating atoms is obtained. The result is not sensitive to standing waves in the external circuitry.

Examination of equation 4 shows that a residual frequency shift $\delta\omega = \beta\overline{\lambda}_0\Gamma_c$ remains even in the limit of zero cavity detuning and negligible collisional broadening (*i.e.* $n \rightarrow 0$). This shift should drop to zero as the quality factor of the maser resonator is increased such that $\beta \rightarrow 0$. If, however, a small offset Δ_0 exists between the measured and the actual cavity detunings, a residual frequency shift $\delta\omega = -\Delta_0\Gamma_c$ remains even in the limit $\beta \rightarrow 0$. The presence of such an offset can be determined by measuring the frequency shift $\delta\omega$ of the maser as a function of Q^{-1} near $\Delta = 0$ and extrapolating to $Q^{-1} = 0$ and $n = 0$. As an illustration of this procedure, we show the data in figure 1. In this particular example, neglecting terms in Δ^2, we find a residual shift of -27.5 ± 3.4 mHz. From a subsequent determination of Γ_c (described in the following section), which to first approximation depends only on *relative* changes in Δ, one can obtain a self consistent

measurement of Δ_0 and hence the *absolute* resonator detuning.

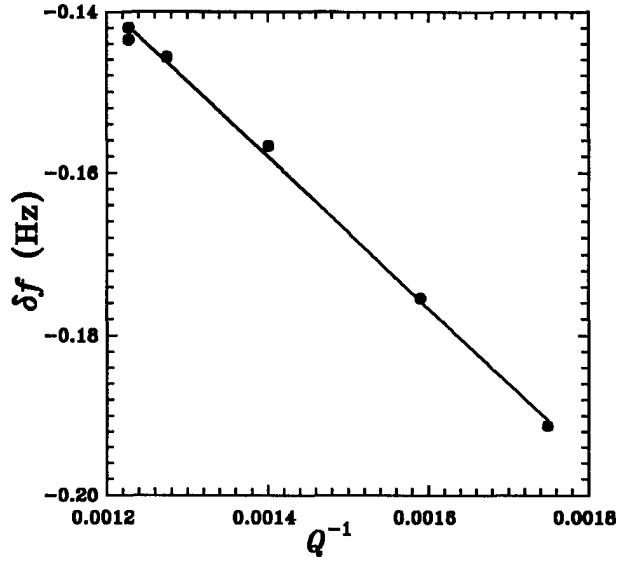


Figure 1: *An example of the shift $\delta f = \delta\omega/2\pi$ of the maser oscillation frequency measured with respect to the density independent frequency $f_0 = \omega_0/2\pi$ and shown as a function of the inverse quality factor Q of the 1420 MHZ resonator near $\Delta = 0$. Extrapolation of the data to $Q^{-1} = 0$ yields a residual frequency shift of -27.5 ± 3.4 mHz which is interpreted as being due to an offset Δ_0 between the measured and the absolute resonator tuning. The uncertainty in each data point is of order one or two mHz.*

The final component of our measurement of $\overline{\lambda}_0$ is an accurate determination of the coupling parameter β which is given by equation 5. The velocity $\overline{v}$ is set by the storage bulb temperature which was maintained at 0.500(3) K throughout all phases of this experiment. This leaves η as the only yet undetermined parameter upon which β depends. Our measurement of η involves the use of a perturbation technique wherein the storage bulb is alternately filled with liquid O_2 (which is paramagnetic) and liquid Ar [10, 21]. In both cases the resonant frequency ω_c of the 1420 MHz resonator is shifted from the frequency of the empty storage bulb/resonator configuration. The fact that the dielectric constants of these two liquids are very nearly the same allows one to separate out the magnetic and electric contributions to the frequency shift, and hence to deduce η.

Frequency Shift Cross Section

A typical example of the measured dependence of the maser frequency shift $\delta\omega$ on the H atom density within the storage bulb is plotted in figure 2. A linear extrapolation of the data to zero density leaves a residual frequency shift of about -22 mHz. Note that this quantity

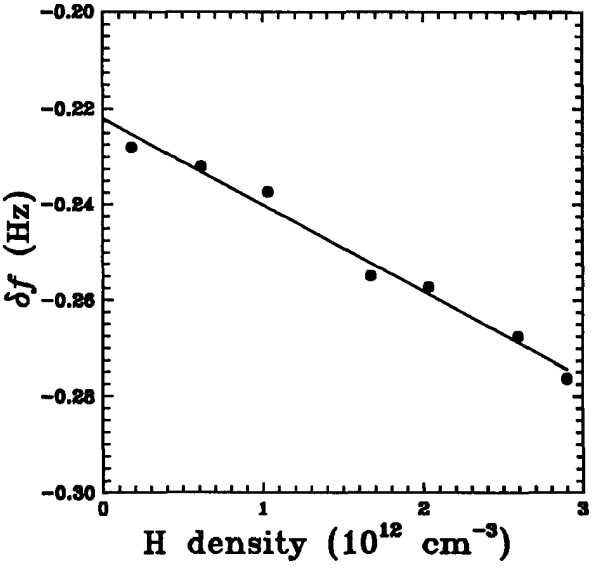


Figure 2: *An example of the density dependence of the maser oscillation frequency measured at constant Q and Δ. Extrapolation of the data to zero density yields a residual frequency shift which does not depend on the absolute calibration of the hydrogen density and which can be used to extract the spin-exchange frequency shift cross section $\overline{\lambda}_0$ (see figure 3). The uncertainty in each data point is of order one or two mHz.*

does not depend upon the absolute calibration of the H atom density. Examination of equation 4 indicates that the residual shift so obtained is given by

$$\lim_{n\to 0} \delta\omega = \left[\Delta + \beta\overline{\lambda}_0 \left(1 + \Delta^2\right)\right] \Gamma_a \qquad (7)$$

from which both $\overline{\lambda}_0$ and Γ_a can be determined. The residual frequency shift taken in the limit of zero H atom density has been plotted as a function of the *absolute* resonator tuning in figure 3 along with a fit of equation 7 to the data. The slight negative curvature indicated by the fit is significant and indicates that $\overline{\lambda}_0 < 0$. We note that to first approximation the value

of Γ_o determined in this manner is independent of the absolute cavity detuning. The absolute cavity tuning was in fact determined from a self consistent evaluation of data such as that shown in figure 1 and figure 2. The fit to the data shown in figure 3 yields a spin-exchange frequency shift cross section $\overline{\lambda}_0 = -21.7 \pm 2.8$ Å^2 where the error includes all experimental uncertainties.

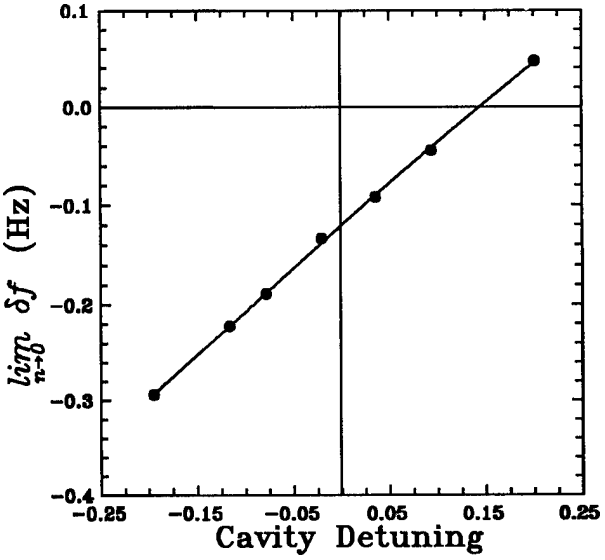


Figure 3: *The residual frequency shift δf taken in the limit of zero H atom density as described in figure 2, plotted as a function of the* absolute *resonator tuning. The solid curve represents a fit of equation 7 to the data. Both $\overline{\lambda}_0$ and Γ_o are determined from this fit. Note that the result for Γ_o depends primarily on* relative *changes in Δ and thus can be used in conjunction with data such as that shown in figure 2 to provide a measure of the* absolute *resonator tuning. The experimental uncertainty in each datum is too small to be indicated on this scale.*

Concluding Remarks

Our experimental result for $\overline{\lambda}_0$ has been plotted in figure 4 along with what is to date the most accurate calculation of $\overline{\lambda}_0$ (full curve) from the known H-H interatomic potentials [7, 22]. The theoretical cross section is roughly half the experimental value and lies well beyond the experimental uncertainty. This situation is in marked contrast to the temperature range 77 K to 363 K over which satisfactory agreement between theory and experiment is obtained [23, 24, 25]. Also shown in figure 4 (dashed curve) is an earlier calculation of $\overline{\lambda}_0$ [26] based on slightly less accurate potentials.[3] The difference between these two predictions gives some indication of the sensitivity of the calculation to the detailed form of the potentials which are used. Note that the difference between the theoretical values for $\overline{\lambda}_0$ as calculated in the DIS approximation and using the fully quantum mechanical theory of Verhaar *et al.* [6, 7] is negligible on this scale.

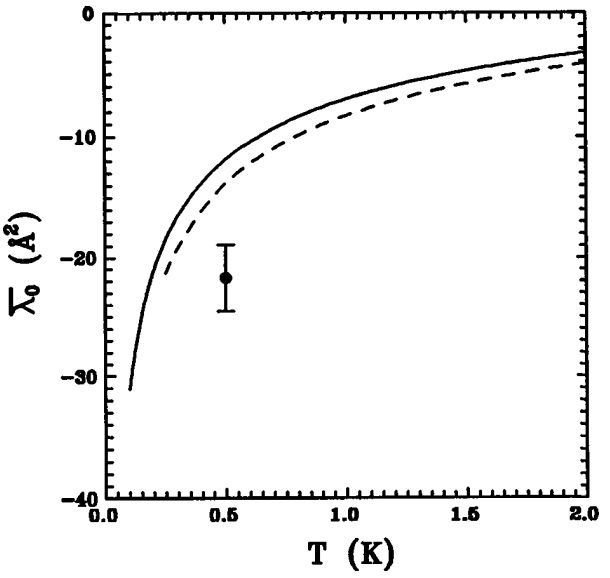


Figure 4: *A comparison of the spin-exchange frequency shift cross section $\overline{\lambda}_0$ measured as outlined in this article with the most sophisticated theoretical calculation of this parameter to date (full curve) [6, 7]. An earlier calculation of $\overline{\lambda}_0$ (dashed curve) [26] based on less accurate potentials is included to give the reader a feeling for the sensitivity of the calculated result to the details of the potentials.*

The principle uncertainty in our experimental determination of $\overline{\lambda}_0$ is due to the uncertainty in the absolute resonator tuning. Care was taken, however, to establish Δ_0 in a manner which makes it insensitive to standing waves in the external circuitry. Further measurements reported in [10, 13] indicate that our determination of Δ_0 is fully consistent with a measurement of the Δ dependence of the power output of the maser.

[3] Further improvements have been made to the relevant potentials since the calculations of Verhaar *et al.* [27].

The question remains as to whether the rather large discrepancy between theory and experiment is caused by inaccuracies in the interatomic potentials used for the theoretical calculations, or by some fundamental or systematic process which has not been accounted for in the analysis of the experimental data. From the latter perspective, a determination of $\overline{\lambda}_0$ (as well as Ω) by an alternate means is certainly desirable.

Acknowledgements

It is with pleasure that the authors acknowledge and thank S. B. Crampton and B. J. Verhaar for many useful and enlightening conversations regarding this work.

References

[1] S. B. Crampton, Hyperfine and Spin Exchange Experiments with the Atomic Hydrogen Maser, PhD. dissertation, Harvard University, 1964 (unpublished).

[2] S. B. Crampton, "Spin-Exchange Shifts in the Hydrogen Maser," Phys. Rev., vol.158, p. 57, 1967.

[3] D. W. Allan, "In search of the best clock - an update," in Frequency Standards and Metrology, A. De Marchi, ed. New York: Springer Verlag, 1989, p. 29.

[4] P. L. Bender, "Effect of Hydrogen-Hydrogen Exchange Collisions," Phys. Rev., vol. 132, p. 2154, 1963.

[5] L. C. Balling, R. J. Hanson and F. M. Pipkin, "Frequency Shifts in Spin-Exchange Optical Pumping Experiments," Phys. Rev., vol. 133, p. A607, 1964.

[6] B. J. Verhaar, J. M. V. A. Koelman, H. T. C. Stoof, O. J. T. Luiten, and S. B. Crampton, "Hyperfine Contribution to Spin-Exchange Frequency Shifts in the Hydrogen Maser," Phys. Rev. A, vol. 35, p. 3825, 1987.

[7] J. M. V. A. Koelman, S. B. Crampton, H. T. C. Stoof, O. J. T. Luiten, and B. J. Verhaar, "Spin-Exchange Frequency Shifts in Cryogenic and Room Temperature Hydrogen Masers," Phys. Rev A, vol. 38, p. 3535, 1988.

[8] Ronald L. Walsworth, Studies in Atomic Physics using Hydrogen Masers, PhD. dissertation, Harvard University, 1991 (unpublished).

[9] Ronald L. Walsworth, Isaac F. Silvera, Edward M. Mattison, and Robert C. Vessot, "Measurement of a Hyperfine-Induced Spin-Exchange Frequency Shift in Atomic Hydrogen," Phys. Rev. A, vol. 46, p. 2495, 1992.

[10] M. E. Hayden, Studies of Atomic Hydrogen Spin-Exchange Collisions at 1 K and Below, PhD. dissertation, University of British Columbia, 1991 (unpublished).

[11] Michael E. Hayden, Martin D. Hürlimann, and Walter N. Hardy, "Atomic Hydrogen Spin-Exchange Collisions in a Cryogenic Maser," IEEE Trans. on Instr. and Measurement, vol. 42. p. 314, 1993.

[12] S. B. Crampton and H. T. M. Wang, "Duration of Hydrogen-Atom Spin-Exchange Collisions," Phys. Rev. A, vol. 12, p. 1305, 1975.

[13] M. E. Hayden, M. D. Hürlimann and W. N. Hardy, "Measurement of Atomic Hydrogen Spin-Exchange Parameters at 0.5 K using a Cryogenic Hydrogen Maser," (to be published).

[14] Daniel Kleppner, H. Mark Goldenberg, and Norman F. Ramsey, "Theory of the Hydrogen Maser," Phys. Rev., vol. 126, p. 603, 1962.

[15] J. Vanier and C. Audoin, The Quantum Physics of Atomic Frequency Standards, Bristol, U.K.: Adam Higler, 1989.

[16] M. D. Hürlimann, W. N. Hardy, A. J. Berlinsky, and R. W. Cline, "Recirculating Cryogenic Hydrogen Maser," Phys. Rev. A, vol. 34, p. 1605, 1986.

[17] W. N. Hardy, M. D. Hürlimann, and R. W. Cline, "Application of Atomic Hydrogen at Low Temperatures: The Recirculating Cryogenic Hydrogen Maser," Jap. J. of Appl. Phys., vol. 26, p. 2065, 1987.

[18] M. D. Hürlimann, Cryogenic Hydrogen Maser, PhD. dissertation, University of British Columbia, 1989 (unpublished).

[19] W. N. Hardy, M. Morrow, R. Jochemsen, and A. J. Berlinsky, "Magnetic Resonance of Atomic Hydrogen at Low Temperatures," Physica, vol. 109 & 110B, p. 1964, 1982.

[20] P. J. Nacher, M. Cornut, and M. E. Hayden, "Compression of ^{3}He by Refluxing ^{4}He: A Model for Computing HEVAC Effects in ^{3}He-^{4}He Mixtures," (submitted to J. Low Temp. Phys.).

[21] M. E. Hayden and W. N. Hardy, "A Technique for Measuring Magnetic Filling Factors with Applications to Cryogenic Magnetic Resonance Experiments," (to be published).

[22] B. J. Verhaar: private communication.

[23] R. M. Mazo, "Linewidths in the Electron Paramagnetic Resonance Spectrum of Gaseous Atomic Hydrogen," J. Chem. Phys., vol. 34, p. 169, 1961.

[24] S. B. Crampton, J. A. Duvivier, G. S. Read, and E. R. Williams, "Frequency Shifts due to Hydrogen-Hydrogen Spin-Exchange Collisions," Phys. Rev A, vol. 5, p. 1752, 1972.

[25] M. Desaintfuscien, J. Viennet, C. Audouin, and J. Vanier, "Temperature Dependence of Spin-Exchange Frequency Shifts in H-H Collisions," J. Physique Lett., vol. 36, p. L-281, 1975.

[26] A. J. Berlinsky and B. Shizgal, "Spin-Exchange Scattering Cross Sections for Hydrogen Atoms at Low Temperatures," Can. J. Phys., vol. 58, p. 881, 1980.

[27] L. Wolniewicz, "Relativistic Energies of the Ground State of the Hydrogen Molecule," J. Chem Phys., vol. 99, p. 1851, 1993.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

CRYOGENIC HYDROGEN MASER AT 10 KELVIN

D. R. McAllaster, J. J. Krupczak, A. L. Cole, A. J. Kerman, and S. B. Crampton

Williams College, Department of Physics, Williamstown MA 01267-2565

Abstract

We have constructed and are evaluating an atomic hydrogen maser that oscillates at temperatures ranging from 8.5 to 11.5 kelvins. The principal scientific purpose for which the maser was constructed is to study the effects of hyperfine interactions during collisions on the radiative frequency shifts produced by the collisions. An equally important technological objective is to explore the reductions of short term frequency fluctuations postulated for low temperature operation. Our maser oscillates strongly using atomic beam intensities about 10^{12} state selected atoms s^{-1}, with output power (at temperatures above 10 K) essentially undiminished by collisions between atoms or with the storage surface. Preliminary frequency measurements indicate that the short term stability of our cryogenic maser is at least as good as the measurement limitations imposed by the short term frequency stability of the conventional hydrogen masers that we use as frequency references.

Motivation

Room temperature hydrogen masers are the most stable frequency source for periods from a few seconds to a few hours[1]. The engineering and design of the best of these masers is such that their stability over these periods is limited by the ratio of the maser power to thermal noise in the maser's microwave cavity and the receiver electronics. Since maser power is in turn limited by spin-exchange collisions between hydrogen atoms in the microwave cavity[2], the dramatically smaller spin-exchange cross sections at cryogenic temperatures[3] offer the possibility of greatly increased maser power, which would contribute to greater frequency stability via a larger signal-to-noise ratio. With other benefits of cryogenic operation, such as less thermal noise in the microwave cavity, decreased dependance of material properties on temperature, and the possibility of using ultra low-noise cryogenic preamps, cryogenic masers have been predicted[4] to be capable of Allan variances as low as 2×10^{-18} of the output frequency. This is two to three orders of magnitude more stable than the best room temperature masers[1].

However, work done by Verhaar and collaborators[3,5] at Eindhoven suggests that this performance may be very difficult to achieve. As temperature declines, the duration of atomic hydrogen collisions in the maser becomes greater, and the phase of the radiation emitted by an atom after a collision (relative to the radiation emitted before the encounter) is predicted to become increasingly dependent on the hyperfine state of the colliding atom. This process can be parametrized in terms of six cross sections, three of which lead to shifts in frequency, and three which broaden the atomic line. The effect of these cross sections is predicted to couple the maser frequency to the atomic density and hyperfine level population in the storage region in a non-linear manner, defeating the usual spin-exchange tuning procedure[6] which renders the maser frequency independent of density.

Thus, there are two major questions that we will address with this research. The technological question is whether the technical difficulties of low temperature operation can be surmounted, and the potential stability improvements of cryogenic masers can be realized. The scientific question is whether the Eindhoven group's theory accurately describes atomic hydrogen collisions and their effect on the maser frequency.

Apparatus

In order for any maser to oscillate, the interior of the storage region of the maser must be coated with some material that has a sufficiently low binding energy for hydrogen, as compared to the kinetic energy of the hydrogen atoms. Room temperature masers have for many years used copolymers of Teflon as a coating material. Masers made with these coatings cease to oscillate at

0-7803-1945-1/94 $3.00 © 1994 IEEE

about 50 K [7], as the atoms spend more and more time adsorbed upon the wall and relaxation broadens the atomic line. To proceed lower in temperature, some material with a smaller binding energy is needed. The forces responsible for the binding energy between hydrogen atoms and the coating are termed van der Waal's forces, and stem from dipole-dipole and higher multipole forces. Coatings with the smallest binding energies will have the smallest polarizabilities, and the least polarizable of atoms are the inert "gases": helium and neon have so far been used as coatings for low temperature storage regions.

An additional issue in cryogenic masers is the considerable vapor pressure of the coating material. Atoms that evaporate from the coating scatter hydrogen atoms out of the incoming beam and reduce the achievable density. Collisions between gas phase coating atoms also shift the frequency of the maser, in proportion to the rate of collision. In addition, the longer a hydrogen atom spends adsorbed on the surface, the greater the phase shift the radiating atom picks up due to perturbation of the hyperfine transition by the nearby surface. This leads to the "wall shift" of the maser frequency, which becomes greater as temperature decreases. Since the magnitude of the vapor shift increases with temperature (with the vapor pressure), if the wall and vapor shifts have the same sign, then there will be a temperature at which the sum of the wall and vapor shifts will not depend on temperature. This is the most desirable temperature regime to operate a cryogenic maser, where the maser frequency is independent of temperature to first order.

Cryogenic masers have achieved self-sustained oscillation with superfluid ^{4}He coatings around 0.5 K [8,9,10], and solid neon about 10 K [11]. Our maser uses a solid neon surface, and was based on a successful prototype version[11]. The first neon surface maser experienced self-sustained oscillation between 8.5 K and 11 K, but was designed for adjustability, rather than stability, and had poor thermal control and temperature uniformity across the atom storage region.

Design

The current version (see Figure 1) was designed to address the thermal problems in the prototype. The vacuum vessels containing the maser are immersed in a liquid nitrogen bath (at 77 K), and the microwave cavity and storage region are cooled to operating temperature by cold helium gas continuously transferred from a liquid helium storage dewar through a coldplate. Vacuum is maintained within the maser by charcoal cryopumps, which are also cooled by continuous helium gas transfer. Hydrogen gas enters a liquid nitrogen cooled RF discharge, which dissociates the molecules into atoms. The atoms effuse from a 2 mm diameter hole in the dissociator, and those atoms in the upper two hyperfine states are focussed into the storage region by a hexapole magnet. The storage region is formed by the interior of a cylindrical sapphire-loaded microwave cavity, whose TE_{011} mode is tuned to the hyperfine transition frequency. Assuming negligible loss on the surface of the storage region (and the 2 mm diameter of the storage region entrance), atoms entering the storage region bounce randomly inside the cavity for a little more than a second before encountering the orifice at the top and exiting.

Temperature Control and Uniformity

Temperature control of the storage region is important because of the variation of wall+vapor shift with temperature. For a solid neon surface, the wall+vapor shift is negative, and its absolute value reaches a minimum of about 1 Hz at a temperature near 11.5 K [12]; if the maser is not operated at this temperature, then changes in temperature are coupled directly to changes in maser frequency. For example, at 10 K, a change of temperature of 1 mK will change the maser frequency by about 1 mHz (7×10^{-13} of the maser frequency).

The temperature of the storage region is measured by four germanium resistance sensors (Lakeshore Cryotronics series GR-200) mounted on the exterior of the copper cavity enclosure. These sensors are repeatable to 1 mK, and as read by the ohmmeter in the LakeShore DRC-93C Controller, have a resolution of 0.2 mK. The temperature of the enclosure is regulated by a bifilar heater wound on the cavity enclosure, operating in a feedback loop with a sensor and the controller. The temperature stability of the enclosure is at worst 0.2 mK, limited by the resolution of the digital sensor readout.

Uniformity of the temperature inside the cavity is enhanced by the thermal conductivity of the sapphire dielectric, which with the neon coating forms the storage region. The thermal conductance of the sapphire is more than 1000 times that of the copper enclosure, ensuring that temperature gradients along the length of the sapphire will be more than 1000 times smaller than the observed gradient of 0.3 K along the copper enclosure. Such a gradient would add less than 0.1% to the atomic linewidth, even at the suboptimal operating temperature of 10 K.

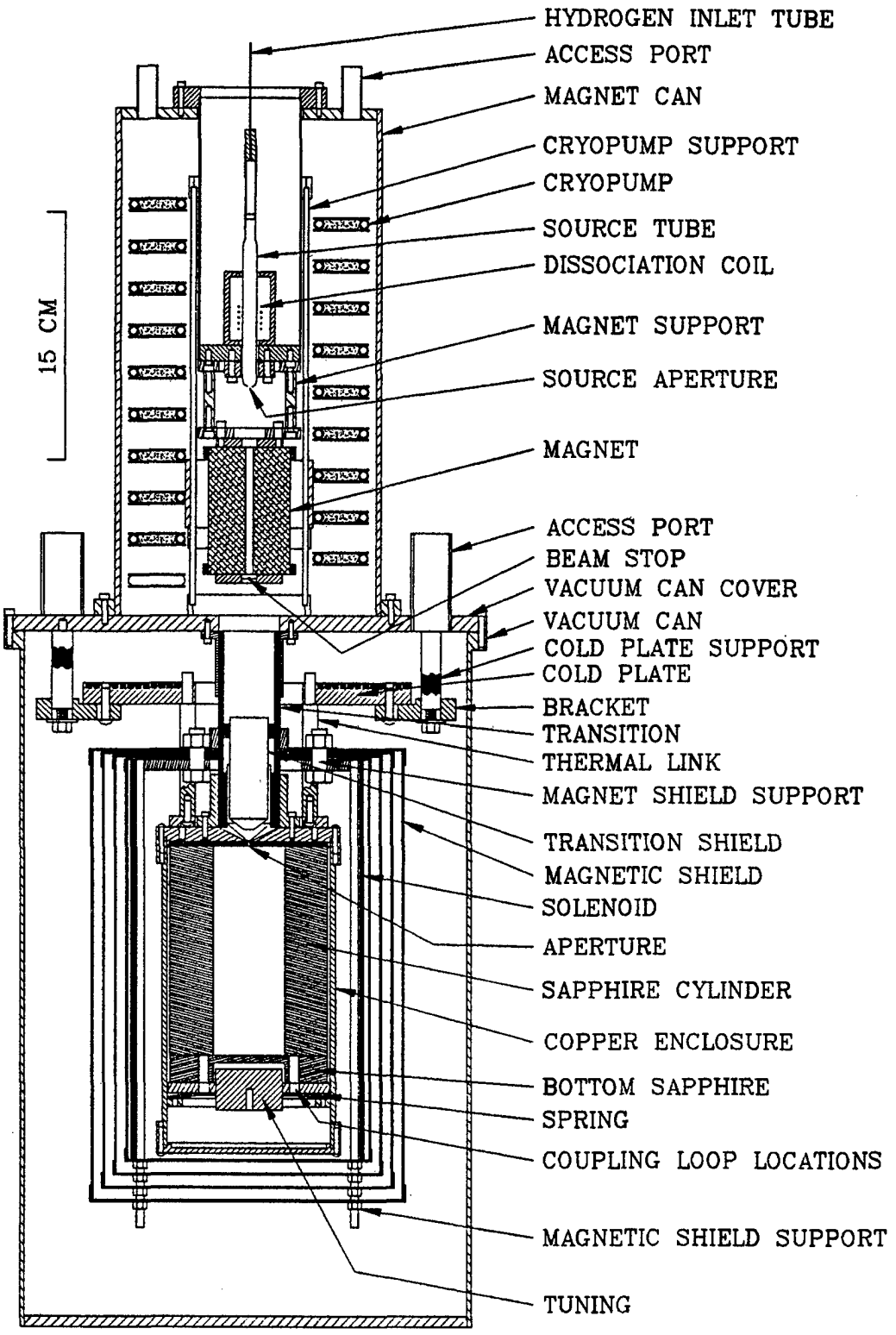


Figure 1. Cryogenic hydrogen maser.

Operation

The toughest technical challenge in running the neon surface maser may be in forming the neon surface coating. The goal is a smooth, uniform surface coating all surfaces accessible to the hydrogen atoms in the storage region. Investigations exploring the properties of two-dimensional electron gases on solid neon surfaces[13] reveal that atomically smooth neon films can be formed by cooling from the neon triple point (325 Torr, 24.56 K) to 10 K over a period of hours. Neon laid down faster than a few μm/hr has a rough, "hoarfrost" appearance, with chambers and crevices that can trap hydrogen atoms and cause them to lose radiative coherence with the rest of the atoms in the storage region. Also, a rough surface smoothes out over time, as atoms on sharp edges redistribute themselves to lower energy sites on flat areas. Since the magnitude of the wall shift is directly proportional to the storage region's surface area, changes in surface roughness are immediately coupled to the maser frequency. For example, at 10 K, a decrease of the coating surface area by 0.1% will increase the maser frequency by 1 mHz.

Unfortunately, design restrictions have led to a maser not ideally suited to forming atomically smooth neon surfaces in the manner of the 2-D electron gas experiments. Because the magnet can (see Figure 1) is directly connected to the storage region, the magnet can is filled with neon to the same pressure as the storage region. As the storage region is cooled, neon condenses onto the storage region surface, the vapor pressure of the neon drops, and neon gas flows from the magnet can, through the 2 mm orifice, and into the storage region. However, flowing neon gas can plug the orifice below 20 K, which precludes maser operation. To avoid this, we pump on the magnet can as the storage region is cooled, regulating the pumping rate so that the magnet can pressure matches the vapor pressure of neon at the temperature of the storage region. In this way, we avoid flow through the orifice, and the consequent plugging of the cavity. However, with this manual procedure it is difficult to avoid occasionally pumping too fast, which may degrade the uniformity of the surface (see below).

Another consideration in surface formation is uniformity. We measure the thickness of the neon coating by monitoring the shift in the microwave cavity's resonant frequency, about -500 Hz/μm. This method is sensitive only to the average neon thickness on the vertical curved surfaces of the storage region, because of the TE_{011} electric field line configuration, and so the cavity monitor cannot tell if there are thin spots in the neon coating. This is important because of the probable presence of paramagnetic impurities in our sapphire: while pure Al_2O_3 is water-clear; our sapphire is pink, indicating the presence of chromium. Early attempts to operate the maser exhibited excessive relaxation and failed to achieve self-sustained oscillation with thin (10 μm) neon surfaces, because (we believe) the hydrogen atoms got too close to the sapphire impurities. This experience parallels a similar problem seen with a bare sapphire cavity briefly used in the SAO (Smithsonian Astrophysical Observatory) helium surface maser[14]. To more strictly segregate the hydrogen atoms from the paramagnetic sapphire, we now deposit 70 μm coatings, which however may be rougher than the thinner surfaces[15]. If the neon coating is not uniformly thick, due to unanticipated thermal gradients, or the erosive action of the atomic beam, then atoms striking thin spots will experience additional relaxation.

Results

The primary predicted advantage of low temperature masers over room temperature versions is their potential for higher maser power before being limited by spin-exchange. Currently, the maximum atomic flux of 10^{12} atoms/sec (about 1% the flux attained by the prototype neon maser) limits the maser to a power output comparable to room temperature masers (-95 dBm). However, we can perform a "proof of concept," and show that higher beam intensities (once achieved) will produce more power.

We examine this issue by observing the amplitude transient after a perturbation of the oscillating maser. We can introduce this perturbation in one of several ways: for example, by instantaneously increasing the RF power to the hydrogen dissociator, or by suddenly turning off a signal sent to the cavity at the Zeeman frequency, which depopulates the upper masing state. After the perturbation, the maser output recovers to its new, higher level. If the fractional change in maser amplitude ε is small, then the recovery to the new level will be governed by[16]

$$\frac{d^2\varepsilon}{dt^2}+\frac{1}{T_1}\frac{d\varepsilon}{dt}+2b_s^2\varepsilon=0\,, \qquad (1)$$

where T_1 is the level population relaxation time, and b_s is the steady-state Rabi frequency, a measure of the microwave field amplitude in the cavity in angular frequency units:

$$b_s = 2\pi\mu_B \langle B_z \rangle_b / h. \quad (2)$$

Here $\langle B_z \rangle_b$ is the vertical component of the microwave field amplitude averaged over the storage region, μ_B is the Bohr magneton, and h is Planck's constant.

For high levels of oscillation, Equation 1 has a damped sinusoidal solution

$$\varepsilon = Ae^{-t/2T_1}\sin(\omega_{tr}t + \phi) \quad (3)$$

with

$$b_s^2 = \frac{\omega_{tr}^2}{2} + \frac{1}{8T_1^2}. \quad (4)$$

Thus, the transient dies away with time constant $2T_1$, and its frequency is related to the microwave amplitude in the cavity. An example of an amplitude transient in the cryogenic maser is shown in Figure 2.

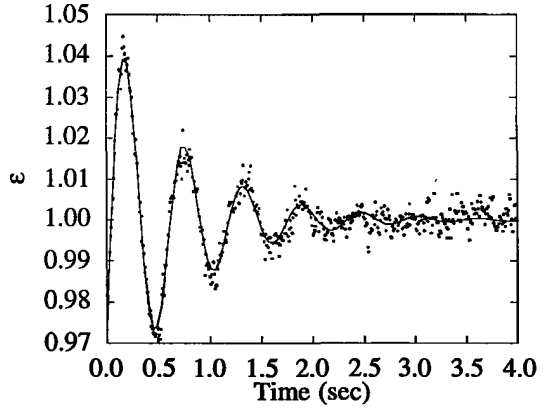


Figure 2. Normalized amplitude versus time for an amplitude transient at 11 K, along with the fit to an exponentially damped sine.

We may take as a figure of merit the quantity $(b_sT_1)^2$, which compares the level population relaxation rate T_1^{-1}(due to spin exchange and any other source) to the maser power output (proportional to the energy in the cavity and thus to b_s^2). A good $(b_sT_1)^2$ value for a high power room temperature maser might be 1; even with a relatively low flux of 10^{12} atoms/sec, the amplitude transient of Figure 2 gives $(b_sT_1)^2 = 8.3$. This shows that once higher flux levels are achieved in this maser, we will indeed be able to capitalize on the small low temperature spin-exchange relaxation rate to produce high maser power.

Another encouraging sign was the observation in an early maser run that at low atomic densities, the linewidth was close to that expected from the geometric lifetime (Figure 3). With an extrapolated zero-power linewidth of 0.4 Hz, and given the 1.3 second bottle residence time, the coherent lifetime of atoms (due to relaxation at the neon storage coating) must be about 2 seconds.

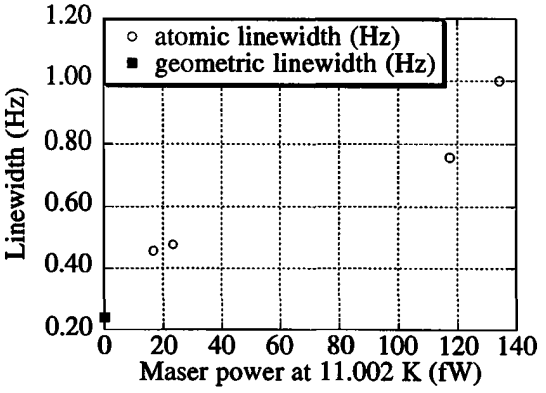


Figure 3. Maser power versus linewidth.

The chief problem experienced by the maser is a persistent frequency drift. This drift may be either positive or negative, and it ranges from 0.1 to 100 μHz/sec ($.35\times10^{-15}$/sec to 71×10^{-15}/sec of the maser frequency), more than ten thousand times faster than good room temperature masers without active cavity tuning. The problem has been worse in recent runs, and has been associated with a larger wall+vapor shift, along with increased binding energy of hydrogen atoms to the wall (manifested by a higher minimum temperature for self-sustained oscillation).

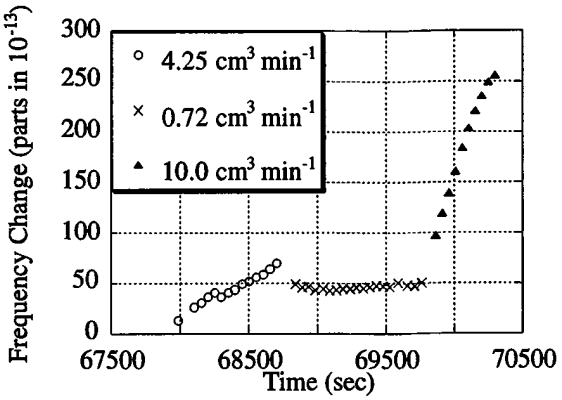


Figure 4. Relative frequency change in the maser for three different H_2 flow rates, as a function of time.

Figure 4 illustrates that the drift rate depends on the molecular hydrogen flow rate into the dissociator. Shown is the frequency of the cold maser (relative to a room temperature reference maser) versus time for three different flow rates. The increased drift at high flow rate is due either to contamination of the surface by impurities from the hydrogen discharge, or damage or thinning in the neon coating. Both would be expected to scale with flow rate.

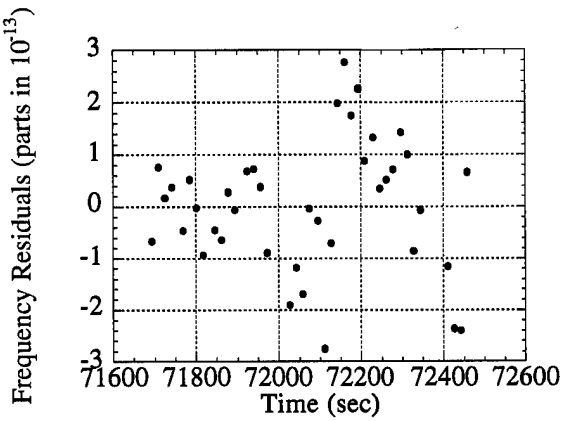


Figure 5. Frequency residuals after removal of linear drift.

Clearly, with this frequency drift problem, the frequency stability of this device does not constitute an improvement over current maser standards. However, by making the mild assumption that the frequency changes linearly with time, we can consider the frequency residuals about that line to gain some idea of the "intrinsic" stability of the maser. A plot of this kind (Figure 5) gives residuals comparable to those obtained using two of our room temperature masers.

Future Directions

We plan to make substantial changes in the maser soon. The first change will be to insert a quartz liner into the storage region, to keep the hydrogen atoms further away from the paramagnetic sapphire impurities, without requiring the use of a thick neon coating. When installing the liner, we will widen the orifice to reduce the probability of plugging the opening with neon; this will give us a great deal more latitude in how we form the surface. We may also cool and temperature control both the bottom of the copper cavity enclosure and the top; this would reduce the already small (predicted) temperature gradient along the length of the sapphire still further.

Assuming the modifications are successful, we will go on to measure the frequency shift cross-sections, and compare them to the predictions of the Eindhoven group.

1. Robert F. C. Vessot, "State of the art and future directions for the atomic hydrogen maser," in Proceedings of the 22nd Precise Time and Time Interval (PTTI) Applications and Planning Meeting 1990.

2. D. Kleppner, H. C. Berg, S. B. Crampton, N. F. Ramsey, R. F. C. Vessot, H. E. Peters, and J. Vanier, "Hydrogen-maser principles and techniques," Physical Review vol. 138, pp. A972-A983, 17 May 1965.

3. J. M. V. A. Koelman, H. T. C. Stoof, O. J. Luiten, B. J. Verhaar, and S. B. Crampton, "Spin-exchange frequency shifts in cryogenic and room-temperature hydrogen masers" Phys. Rev. , vol. A38, p. 3535, 1988.

4. W. N. Hardy and M. Morrow, "Prospects for low temperature H masers using liquid helium coated walls," J. de Physique, vol. 42, pp. C8-171 - C8-179, 1981.

5. B. J. Verhaar, J. M. V. A Koelman, H. T. C. Stoof, O. J. Luiten, and S. B. Crampton, "Hyperfine contribution to spin-exchange frequency shifts in the hydrogen maser," Phys. Rev. , vol A35, p. 3825, 1987.

6. S. B. Crampton, "Spin-exchange shifts in the hydrogen maser," Phys. Rev. , vol. 158, pp. 57-61, 5 June 1967.

7. R. F. C. Vessot, E. M. Mattison, E. A. Imbier, and Z. C. Zhai, "Atomic hydrogen maser measurements with wall surfaces of carbon tetrafluoride," in Proceedings of the Sixteenth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, November 27-29, 1984.

8. Ronald L. Walsworth, Jr., Isaac F. Silvera, H. P. Godfried, C. C. Agosta, Robert F. C. Vessot, and Edward M. Mattison, "Hydrogen maser at temperatures below 1 K," Phys. Rev. A, vol. 34, no. 3, pp. 2550-2553, September 1986.

9. H. F. Hess, G. P. Kochanski, J. M. Doyle, T. J. Greytak, and D. Kleppner, "Spin-polarized hydrogen maser," Phys. Rev. A, vol. 34, no. 2, pp. 1602-1604, August 1986.

10. M. D. Hürlimann, W. N. Hardy, A. J. Berlinsky, and R. W. Cline, "Recirculating cryogenic hydrogen maser," Phys. Rev. A, vol. 34, no. 2, pp. 1605-1608, August 1986.

11. S. B. Crampton, K. M. Jones, G. Nunes, and S. P. Souza, "Hydrogen maser oscillation at 10 K," in Proceedings of the Sixteenth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 339-349, November 27-29, 1984.

12. S. B Crampton and S. Steel, unpublished.

13. Koji Kajita, Yutaka Nishio, and Wataru Sasaki, "Electrical conductivity of highly correlated 2D electrons formed on solid neon," Surface Science, vol. 170, pp. 88-93, 1986.

14. Ronald L. Walsworth, Jr., personal communication.

15. Koji Kajita, personal communication.

16. Jacques Vanier and Claude Audoin, The Quantum Physics of Atomic Frequency Standards. Bristol and Philadelphia: Adam Hilger, 1989, vol. 2, p. 1019.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A TIME TRANSFER TECHNIQUE USING A SPACE-BORNE HYDROGEN MASER AND LASER PULSE TIMING

Edward M. Mattison, Laurence M. Coyle, Robert F.C. Vessot, and David Arnold
Harvard-Smithsonian Center for Astrophysics
60 Garden St., Cambridge, MA 02138
R. Clayton Smith
Los Alamos National Laboratory
Los Alamos, NM 87545

Abstract

We describe a system for worldwide precision time transfer between ground-based clocks and a satellite-borne clock, using laser pulse timing. The system is designed to compare time between earth and space clocks with a precision of 100 ps or better, and will be tested as part of the NASA/SAO Hydrogen Maser Clock experiment, which is currently under construction.

Time Transfer Technique

As part of the NASA/SAO Hydrogen Maser Clock (HMC) experiment, we have developed a system for high-precision time transfer between space-borne and earth-based clocks, using laser pulse timing. When used with a hydrogen maser in an earth-orbiting spacecraft, this technique promises time comparisons with a precision of 100 ps or better worldwide. The HMC experiment will test this system, together with an active hydrogen maser designed for use in space, for a period of approximately six months on a spacecraft with an altitude of about 500 km.

In this system, time kept by a clock located at an earth-based laser ranging station (LRS) is compared with the space clock's time by measuring the arrival time of a laser pulse at the spacecraft in terms of both the space and earth time scales. A laser pulse transmitted from the LRS is detected at the spacecraft and its arrival time t_s is determined in terms of the space clock's time by a high-speed electronic interpolation circuit called an event timer. The pulse is also reflected by a retroreflector array mounted on the spacecraft and is received at the LRS as a return pulse. Its transmission and return times, t_{1e} and t_{2e}, are measured in terms of the earth clock's time by an event timer located in the LRS. The time difference Δt between the earth and space clocks is then given by $\Delta t = t_s - (t_{1e} + t_{2e})/2$.

The main components of the time transfer system, which are shown schematically in Figure 1, are the retroreflector array; a fiber-optic light collector mounted on the retroreflector array; a photomultiplier tube and preamplifier; and the event timer, consisting of a constant-fraction discriminator and a high-resolution time interpolator. These components, as designed for the HMC experiment, are described below.

Retroreflector Array

The HMC retroreflector array consists of 20 fused silica cube corners, each 1 cm in diameter, mounted in a hemispherical base. This shape provides a hemispherical field of view, permitting reflections independent of spacecraft attitude. (Two reflector arrays mounted on opposite sides of the spacecraft provide complete spherical coverage.) The use of a relatively large number of small reflectors confers three advantages: (i) velocity aberration is accounted for by diffraction by the reflector's entrance pupil, allowing the reflecting surfaces to be manufactured with a dihedral angle of exactly 90°; (ii) the array can be made compact, approximately 6 cm in diameter, reducing the variation in centroid as a function of viewing angle; and (iii) the array's far-field intensity

U.S. Government work not protected by U.S. copyright

pattern is uniform with angle, minimizing return intensity variations measured at the LRS.

Fiber Optic Light Collector and Photodetector

Laser pulses impinging on the retroreflector array are brought to a photomultiplier tube by an omnidirectional fiber optic light collector. The collector consists of a 22-cm long bundle of 127 optical fibers, each 200 μm in diameter. At one end of the bundle the fibers are splayed out into a hemispherical pattern and inserted through holes drilled in a 1.5-cm diameter hemispherical shell, as shown schematically in Figure 2. The fibers are epoxied in place within the shell, which is mounted at the apex of the retroreflector array. The other end of the fiber bundle connects to a photomultiplier tube (PMT) that detects the laser pulses. A focussing lens and a neutral density filter, located between the bundle and the PMT focus and attenuate the light, and a narrow-band interference filter eliminates ambient light away from the laser's 532 nm wavelength. A PMT was chosen as the photodetector because it has a large cathode area, which simplifies coupling the fibers to the detector, and a high gain that can be varied over a wide range, allowing adjustment to varying pulse intensities. The uniformity of the retroreflector array's return pattern with angle means that the distance offset between the light collector's centroid and the array's centroid is a function only of the incoming laser pulse's polar angle, and is independent of azimuth; this property simplifies the correction for spacecraft attitude.

Event Timer

The time transfer system's ability to resolve sub-nanosecond intervals results from a high-precision space-qualified event timer developed at the Los Alamos National Laboratory that is capable of timing with a resolution of 10 ps. The output pulse from the PMT is amplified by a preamplifier and sent to a constant-fraction discriminator (CFD), which produces a pulse whose shape is to largely independent of the input pulse amplitude [1-3]. This property of the CFD reduces the systematic variation in triggering time, called time walk, that would otherwise result from the orders-of-magnitude variation in arriving laser pulse intensity that can result from changes in atmospheric conditions and in spacecraft attitude and altitude.

The CFD's output pulse goes to a time interpolator[4,5],which is a combined digital and analog circuit that interpolates the 100 MHz clock signal provided by the spacecraft's hydrogen maser. When a pulse triggers the time interpolator, a capacitor is charged by a constant current I_c until the next clock edge arrives. The capacitor is then discharged by a second constant current $I_d = -I_c/1000$, and the discharge time is measured in terms of clock periods. By this technique the interpolator divides the 10 ns clock period into 1000 "bins", providing 10 ps resolution.

We have built and tested brassboard versions of the photomultiplier and event timer circuits. The interpolator's integral linearity, which is a measure of the total timing error for any bin compared to a perfectly linear interpolator, is less than 10 ps, with test-to-test repeatability of less than 1 ps.

Acknowledgment

This work is supported by the George C. Marshall Space Flight Center, Huntsville Alabama.

References

1. R.C. Smith, "Optimal design of high speed analog APD receivers," *SPIE Proceedings* **987**, Sept 1988; Los Alamos National Laboratory report LA-UR-88-3731.
2. B.T. Turko and R.C. Smith, "A precision timing discriminator for high density detector systems," Lawrence Berkeley Laboratory report LBL-30602.
3. B.T. Turko, W.F. Kolbe and R.C. Smith, "Ultra-fast voltage comparators for transient waveform analysis," *IEEE Trans. on Nucl Sci.* **NS-37**, 424 (1990).
4. R. Nutt, "Digital time intervals meter," *Rev. Sci. Instr.* **39**, 1342 (1968).
5. B. Turko, "Multichannel interval timer," *IEEE Trans. Nucl Sci.* **NS-31**, 167 (1984).

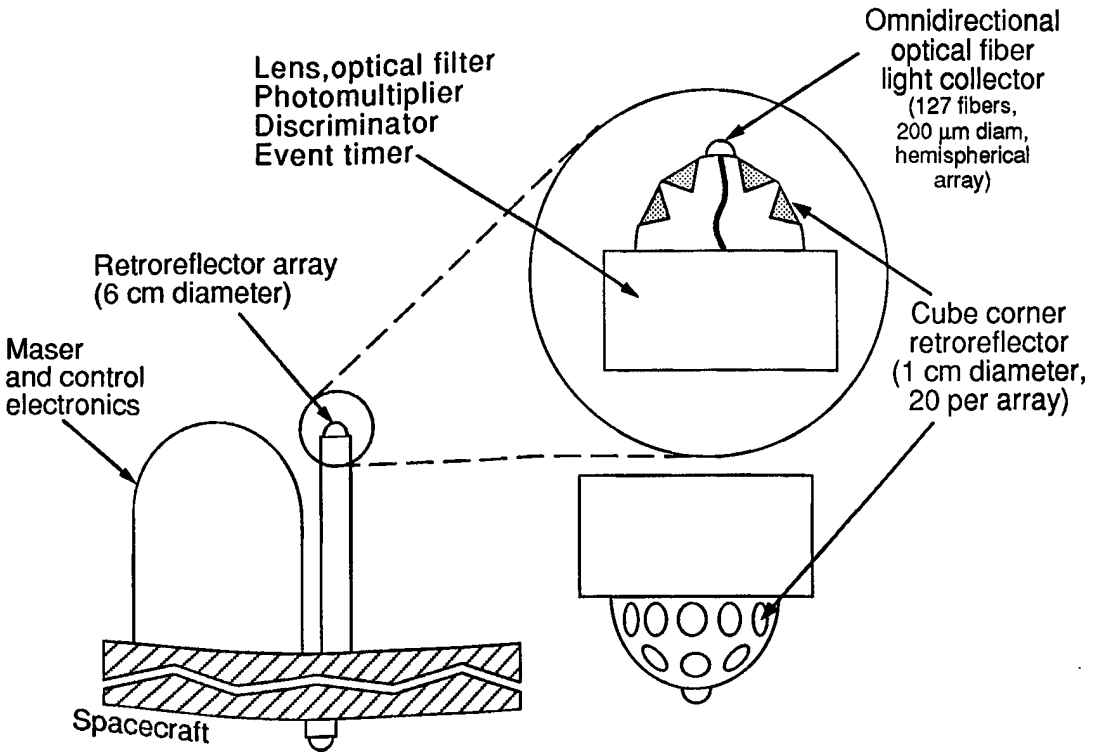


Figure 1. Schematic view of time transfer system components

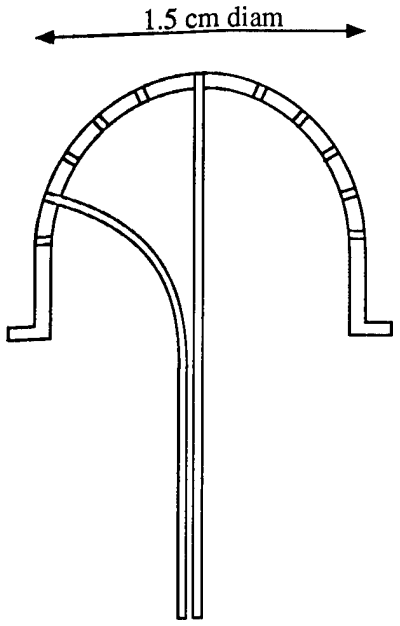


Fig 2. Schematic cutaway view of fiber-optic light collector shell. Two of 127 fibers shown.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

RADIATION RESISTANT AND REPRODUCIBLE FLUORINE-BASED COATINGS OF H-MASER STORAGE BOX FOR A POSSIBLE USE IN SPACE

A.Kalachev,
Institute of Petrolchemical Synthesis,
29, Leninsky pr., Moscow, 117912, Russia

V.Marihin, L.Myasnikova, V.Svetlov,
Joffe Phisico-Technical Institute of Russian Academy of Science,
26, Polyteknicheskaya str., St.-Petersburg, 194021, Russia

D.Muzhdaba, A.Myasnikov,
Russian Institute of Radionavigation and Time,
2, Rastrelly sq., St.-Peterburg, 193124, Russia

Abstract

An electron microscopical study of PTFE coatings of H-maser's storage box revealed a well pronounced morphology comprising from lathlike overlapped PTFE crystals that were suggested to act as the traps for active H-atoms. If was found that they were formed at any regime of sintering with subsequent annealing or quenching of the coatings irrespective of the change in their sizes. The latter was supposed to be one of the reason of the poor reproducibility of the H-maser frequency. The possible absorption of H-atoms and residual gases from H-maser vacuum system by a storage box surface, as well as the organic and inorganic impurities, surfactants, radicals and chain cilic on the surface can also lead to formation of the local active recombination centers of H-atoms. These result in reducing H-maser life time and widening its spectral line. The novel technique of the fluorine-based gaseous monomers is developed. The coatings comprise the crosslinked non-crystallizable polymer networks with the enhanced resistance to the deteriorating action of H-atom bombardment.

The coating produced by the proposed technique are better reproducible due to simplicity in controlling discharge regime by relevant equipment and possibility fast determining chemecal content with the help of XPS and FTIR. The use in an operating H-maser the coatings produced from glow-discharge in perfluorkerosene and C_4F_8 atmosphere demonstrated their promising properties in the working condition under the action of high irradiation closes.

Introduction

Accuracy parameters of GLONASS as well as its reliability and duration of self contained operation are directly dependent on stability of frequency standards installed space vehicles onboard. The topical problem is to reach the daily frequency instability 10^{-14} - 10^{-15} onboard. One of the way to solve this problem is developing H-maser onboard. There are, however, some difficulties, e.g. poor controlled change in shift frequency at quasy-adiabatic collisions of H-atoms with a wall storage box. It is one of factors limiting H-maser long term stability. The wall shift of H-maser frequency is a consequence of superfine structure perturbation of H-atoms at their collision with the surface of the fluorine-based coating. It is determined by pair interaction Lennard-Johnes potential (depth of potential well H->CF_2, concentration of CF_2 groups, sound velosity, absorption heat, thermal coefficient of accomodation and that of heterogeneous recombination) i.e. it is primarily determined by a surface state. The ideal coating of storage box should be homogeneous, stable to the action of chemical reagents, high radiation resistant and it should not change its properties during operational time of H-maser. The coating, however, are exposed to the deteriorating action of H-beam and irradiation which makes it difficult to obtain high characteristics of H-maser.

The aim of the presented study was the finding of a proper route for producing high quality, radiation-resistant coating of H-maser storage box and developing the techniques for control of its properties.

0-7803-1945-1/94 $3.00 © 1994 IEEE

Experimental

The objects of investigation have been fluorine-based coatings produced by conventional technique (FBCC) and those produced from a glow-discharge in an atmosphere of fluoride-based gaseous monomers, i.e. plasma-produced coating (PPC). The FBCC have been cast from polytetrafluorethylene or its copolymer water emulsion with subsequent sintering for removing the traces of moisture and surfactants. For production of PPC, 30 different freon-like gases have been used. The coatings have been deposited on the inner surface of H-maser storage box or on the Si-SiO_2 chips.

A complex study of the properties of the initial coatings as well as those after irradiation have been carried out with a help of different techniques such as electron microscopy (JSM-35), X-ray analysis in wide angles (DRON-1, CuK_α-irradiation, WAXS photo patterns), FTIR (IFS-113V) and XPS(Leybold-Hereaus surface analyser). Desorption of the residual gases and impurities on the investigated coatings have been studied in the experimental setup using mass-spectroscopy. The coefficient of H-atoms recombination (life-time of H-atoms) have been measured with the help of chemiluminiscent probe technique.

The coatings have been exposed to γ-irradiation using ^{60}Co-source (E = 1.28 MeV), X-ray in diffractomer (U = 36 kV, J = 20 μA CuK_α in DRON-1 in vacuum) electron (E_e = 5.7 MeV) and proton (E_p=10MeV) irradiation in Leybold-Hereaus surface analyzer or in pulse electronic accelerator ($E_{e,p}$=10MeV).

Results

It is seen in Fig.1 a well pronounced surface morphology of FBCC which comprises the lathlike PTFE crystallites piled on each other as in a felt. The latter form some kind of caves which probably works as the traps for the active H-atoms reducing their life-time (τ_s). X-ray diffraction pattern confirms semicrystalline structure of the coatings with isotrope distribution of small crystallites (Fig.2). Producing the coatings via proposed plasma technique [1] enabled one to receive amorphous films which is evidenced by structurless surface in scanning micrograph (Fig.3) and amorphous halo in WAXS pattern (Fig.4). A scheme of the experimental setup for PPC production is presented in Fig.5. The analysis of XPS view-spectra of coatings produced from different monomers (Fig.6) allowed to choose the monomers provided the highest fluorine content (44%) in the PPC (perfluorkerosene and C_4F_8). The optimal regime of glow-discharge is HF - 40 kHz, N = 0.3W/cm^3, P = 0.3Torr, film thickness is 10μm. The preliminary results obtained for H-maser with storage box coated from perfluorkerosene gave the results comparable with those for referent H-maser with FBCC ($N=8\times10^{-14}W$, $\delta\nu$ = 1.2Hz, $\sigma(2,\tau)=1\times10^{-12}\tau^{-1/2}$ (enhaced Q-resonator), "Wall" shift of frequency is 7×10^{-12}.

Studying the coatings FBCC and PPC in setup (Fig.7), the effect of nonequilibrium desorption of impurities, adsorbed molecules or adsorbed gases, organic molecules or surfactance (which are the active surface centers of recombination for H-atoms) have been discovered. The most probable channel of desorbtion is transmission of the energy of the oscillately exited molecules H_2^v->L to the adjacent neighbouring molecules in the adsorbed layers of impurities - with a susequent transition of these molecules in gaseos phase (H_2^v->L molecules are formed on the coating surface due to heterogeneous recombination of H-atoms). The treatment of coating surface via bombardment by diffuse intense stream of H-atoms (passivation) appeared to be an effective way for removing these active centers [2]. The quality of the coating surface was controlled by mass-spectroscopy analizator. Besides, the change in τ_s of H-atoms in a storage box was estimated using luminiscent probe technique [3], based on registration of the decay of the radical-recombination-chemiluminescense by interrupting H-atom beam (Fig.8). Note that the passivation can increase τ_s of coating by a factor of 3.

Radiation resistance of FBCC and PPC

At work in space, the onboard H-maser can be exposed to the action of ionizing radiation of different nature, which dose reaches 10^5 - 10^7 rad. The resistance of FBCC and PPC deposited on chips or storage boxes to irradiation has been investigated. FTIR spectra of γ- and e-irradiated samples are presented in Fig.9,10,11. From the comparison of FTIR spectra from γ-irradiated samples one can draw that in both FBCC and PPC depolymerization occurs, since the intensity of absorption bands attributed to CF_2 and CF groups drops by 25% and 5% respectively. This process in amorphous PPC, is accompanied by formation of double C=C bonds, probably, in the sites of connections of the ideal 15_7 helixes and the defect 10_3 and 4_1 ones. The latter does not lead to lowering mechanical properties of coating. At the same time in

semicrystalline FBCC the formation of C=C bonds does not observed. Increasing dose leads to the full deterioration of the irradiated coatings. This observation coincides with the data, obtained for irradiated PE [4] when crosslinking occurred in disordered regions and molecular scission prevealed in crysstalline ones. The e-irradiation of FBCC leads to a marked defluorination of a coating (by 25%) and conformational transformation, e.g. the change in a type of PTEE helix (Fig.11).

In the XPS spectra from γ-irradiated PPC samples no significant change is observed in the range of F_{1s} and F_{kll} peaks but one can see a widening of C_{1s} peak, which can be attributed to the overlapping of many different carbon-fluorine compounds with different fluorine countent (Fig.12,13), one can suggest that they are formed because of F-atom substitution. At the same time in XPS spectra of γ-irradiated FBCC samples one can see a decrease in the intensity of F_{1s}-peak (by 15%) and F_{kll} peak (by 5%).

E-irradiation of FBCC results in defluorination (by 20%) beginning from the dose of $1x10^7$rad which is followed from XPS spectra (Fig.14), while at D = 10^8rad the defluorination reaches 80%. PPC, however, is much more stable and lost only 20% of F-atoms because of crosslinking (Fig.15).

The effect of irradiation of H-maser storage box with FBCC had been also investigated and it was discovered that τ_s markedly increases with the increase of e-irradiation dose (about 10^6 rad) what can be attributed, in our opinion, to chemical purification of the storage box surface due to irradiation stimulated diffusion (Fig.16).

Conclusions

1. The new technique for production of the coating of H-maser storage box from the glow-discharge in an atmosphere of fluorine-based gaseous monomers (C_4F_8 and perfluorkerosene) is developed, which provides high reproducibility and enhanced properties in comparison with those of the standard coatings (deposited from water emulsion of polytetrafluorethylene and its copolymers) because of simplisity in controlling discharge regime, purity of initial monomers and the absence of well pronounced surface relief of plasma coatings. The novel coatings comprise crosslinked noncrystallizable polymer networks with enhaced resistance to the deteriorating action of H-atoms bombardment.

2. It is shown that the properties of standard coatings can be improved via the surface bombardment by diffuse stream of H-atoms (passivation technique). With time, however, the active centers of recombination diffuse from the volume to the surface again.

3. It is found that passivated coatings are stable under, γ- X-ray and e- and p-irradiation up to the doses 10^6 - 10^8 rad.

4. This dosage ruines the initial, untreated standard coatings.

5. Plasma-produced coatings appear to be more stable to irradiation than the standard ones.

Reference

[1] A.Kalachev et.al. "Plasma-chemical technique new route for producting fluorine-based covering of storage box of H-maser". Abstracts of the 7th ESCAMPIG, St.Petersburg, 1992, p.473.

[2] B.Grankin, D.Muzhdaba, A.Myasnikov, B.Romanyk, B.Styrov "Method of receiving of coating". a.s. N^{o} 1579573 USSR.Otkrytiya.Izobreteniya, N^{o}3, 1990.

[3] B.Styrov, U.Turin, I.Govorunov "Determination of the heterodyne coefficient of recombination of atoms (radicals) by luminescent technique". Dokladi AN USSR, 1976, V.228, N^{o} 3s, p.652-655.

[4] P.Klein, D.Woods and I.Werd. Journal of polymer science, 1987, V.25, p.1352-1371.



Fig.1. Scanning micrograph of the sample produced using FBCC technique.

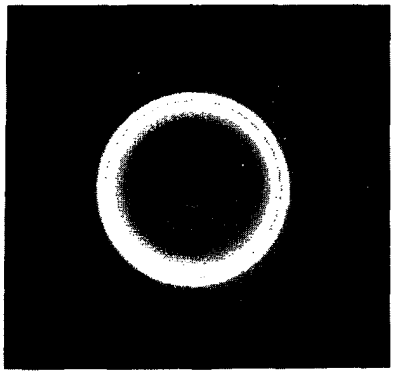


Fig.2. WAXS photo pattren of the sample produced using FBCC technique.

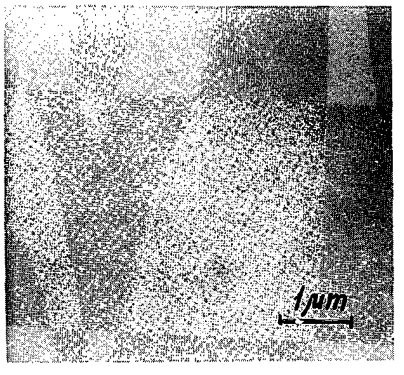


Fig.3. Scanning micrograph of the sample obtained from perfluorkerosene using PPC technique.

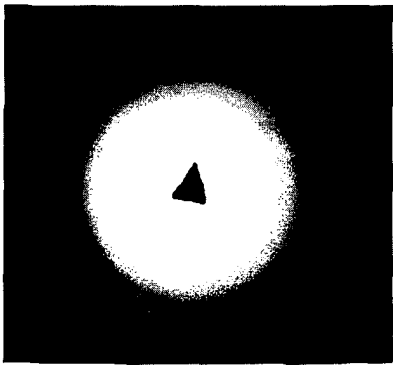


Fig.4. WAXS photo pattern of the sample demonstrated in Fig.3.

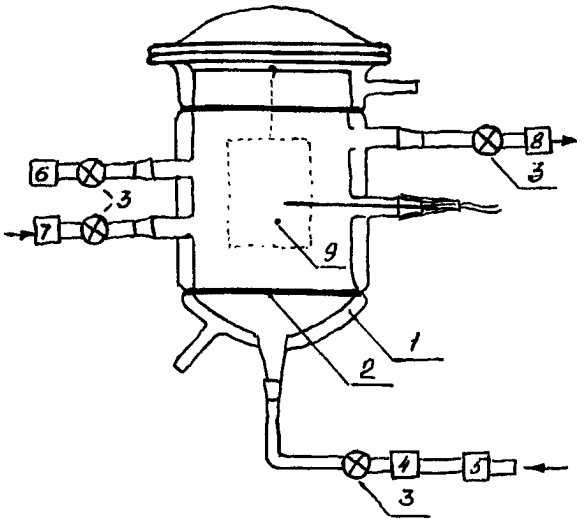


Fig.5. The experimental setup for plasma/chemical synthesis of fluorine-based coatings (PPC technique).
1 - reactor shell with ovenized case, 2 - external HF-electrodes, 3 - valve, 4,7 - in-leakage valves,
5 - mixer, 6 - absolute pressure gauge Baratron (USA), 8 - low-level vacuum pump,
9 - sample, 10 - thermocouple.

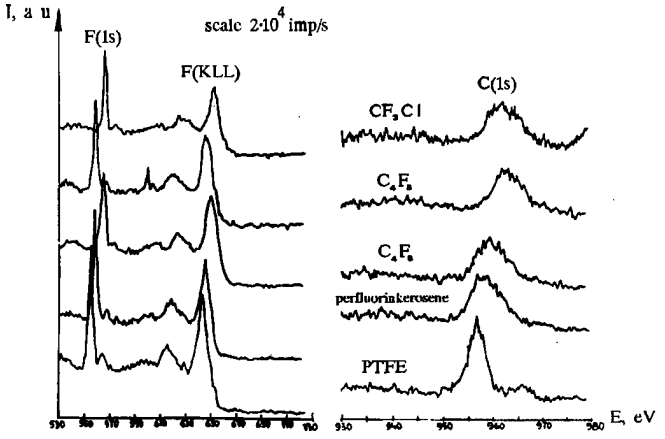


Fig.6. XPS-view spectra of fluorine-based coatings produced by FBCC and PPC technique.

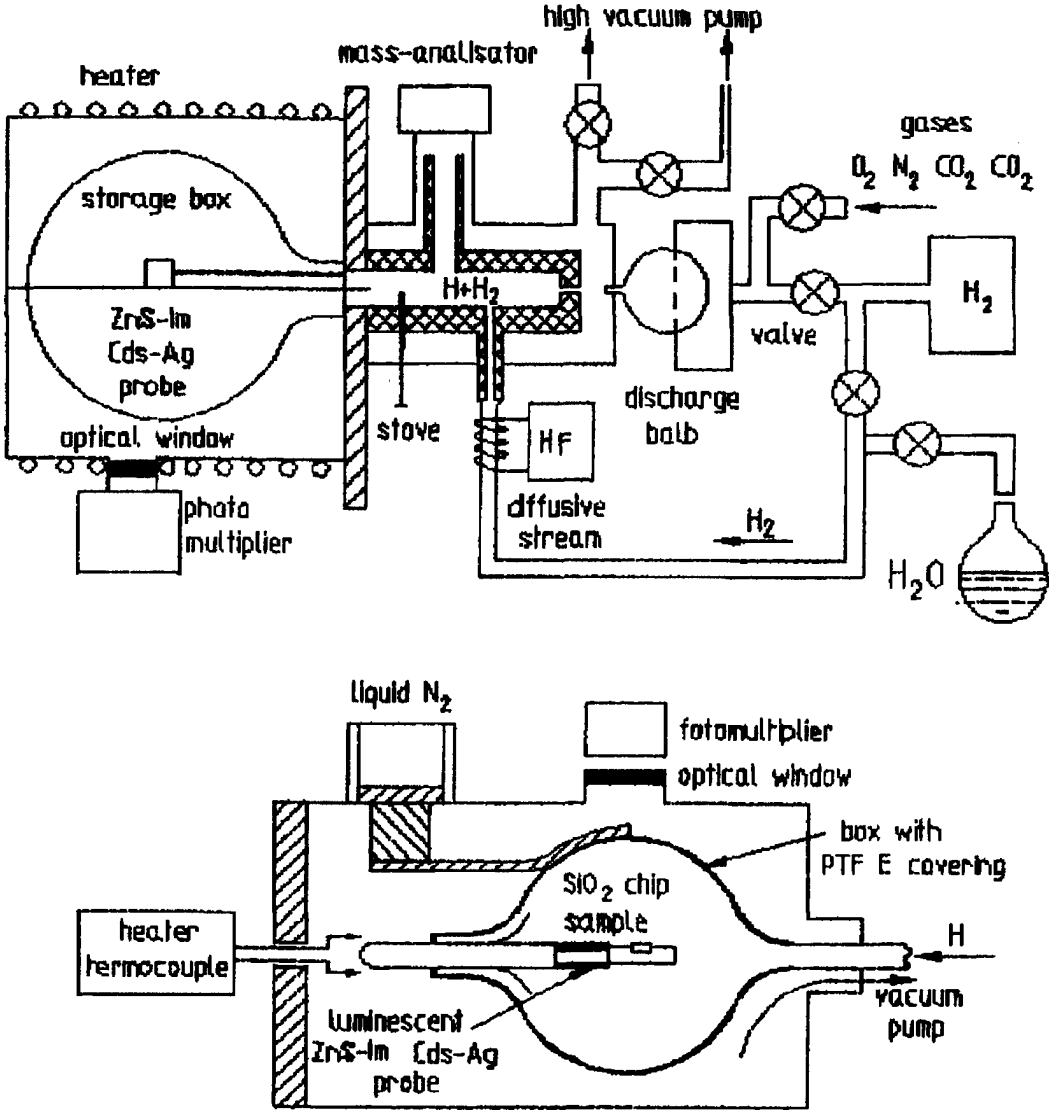


Fig.7. Experimental setup with an additional unit for investigating mass-analysis of desorption processes and measuring relaxation coefficients within H-maser storage boxes, as well as for Si-SiO_2 chip samples.

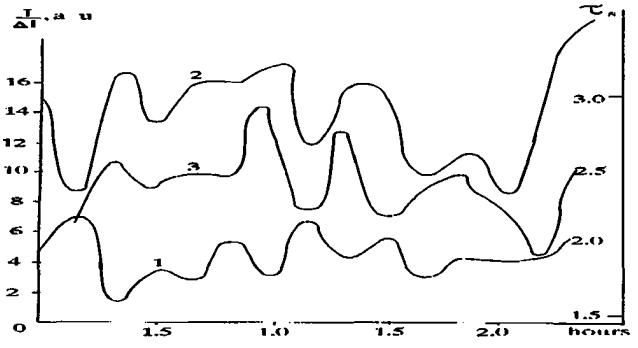


Fig.8. $I_{a.u.}$ for non-equilibrium desorption of H_2O-molecules (1) and OH (2) from PTFE surface and τ_S within a storage box (3) vs experimental time; T = 293K; P = 10^{-5}Torr; F = 2.5×10^{14} $cm^{-2}s^{-1}$; P_{H2O}= 20 Torr.

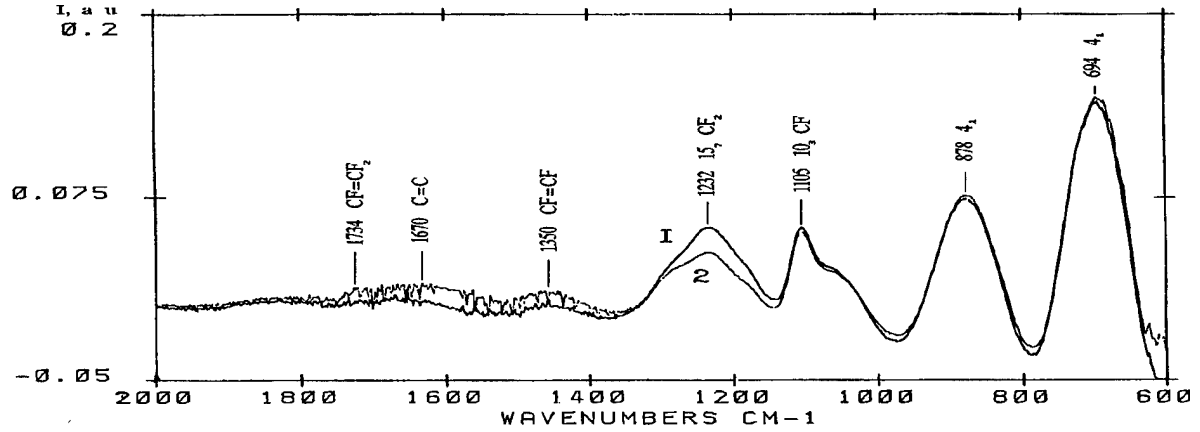


Fig.9. FTIR spectra for PPC from perfluorkerosene: initial (1) and after γ-irradiation (2), D = 10^7 rad.

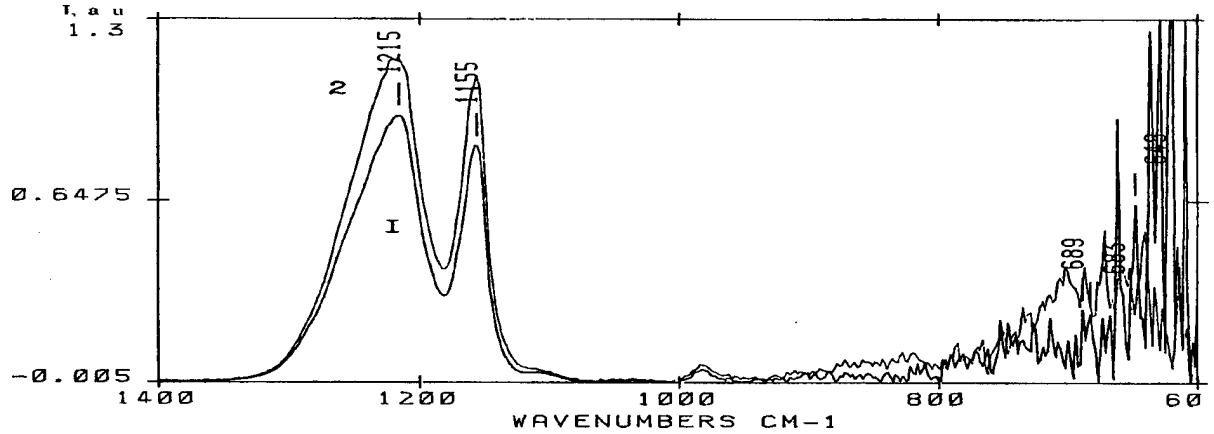


Fig.10. FTIR spectra for FBCC: initial (1) and after γ-irradiation (2), D = 10^7 rad.

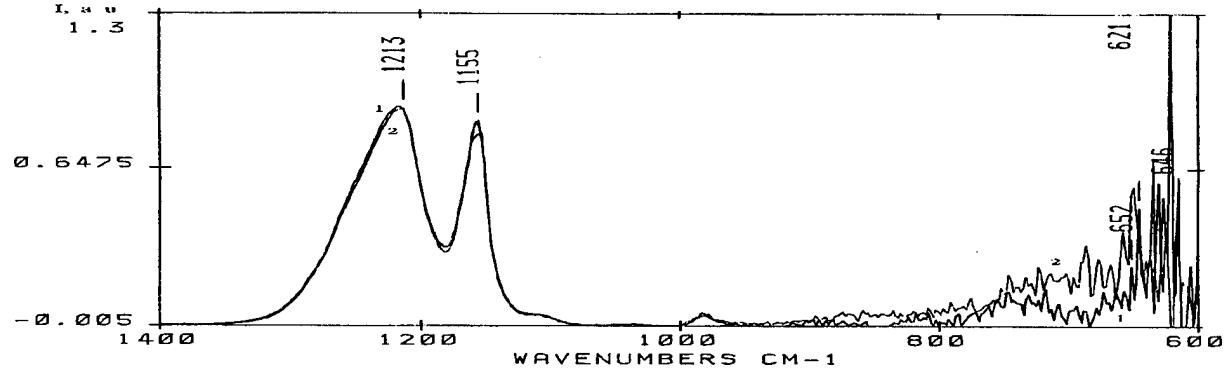


Fig.11. FTIR spectra for FBCC: initial (1), after treatment in a low temperature CF plasma (2), after e-irradiation in Leybold Hereaus surface analyzer at 10^{-8}Torr, U = 3kV, J = (1-10)μA, D = 10^6 - 10^8rad.

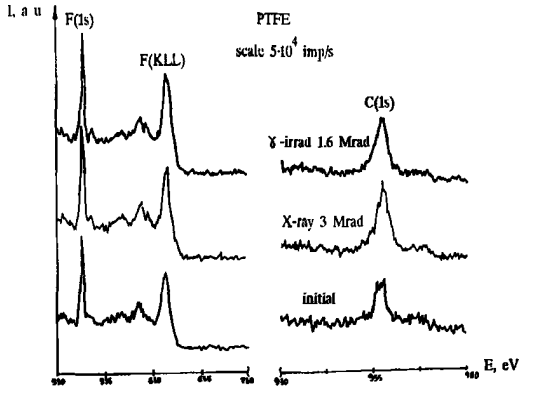


Fig.12. XPS spectra of γ-irradiated (D = 1.6 Mrad) and X-ray - irradiated (CuKα, 6 min, D = 3 Mrad) FBCC in a region of F and C peaks.

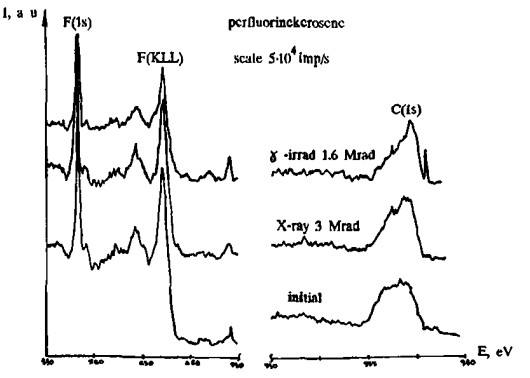


Fig.13. XPS spectra of PPC from perfluorkerosene irradiated in the same conditions as it is indicated in Fig.12.

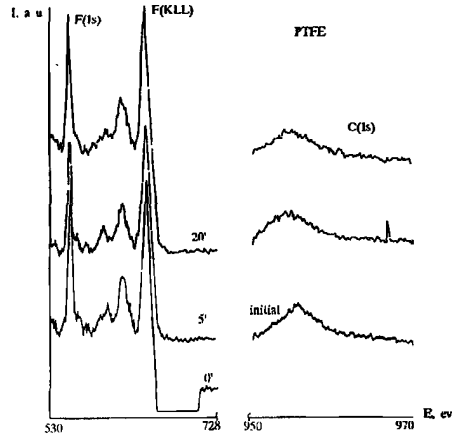


Fig.14. XPS spectra for FBCC e-irradiated in the same conditions as it is indicated in Fig.11.

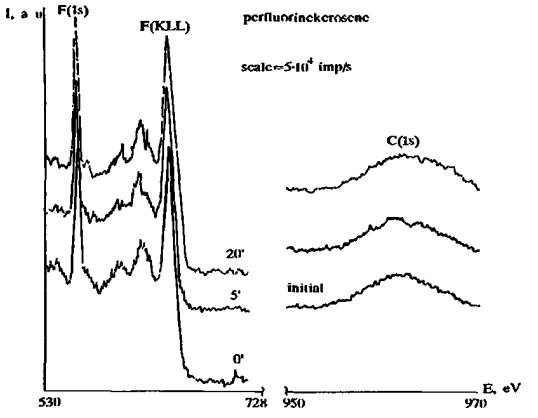


Fig.15. XPS spectra for PPC from perfluorkerosene e-irradiated in the same conditions as it is indicated in Fig.14.

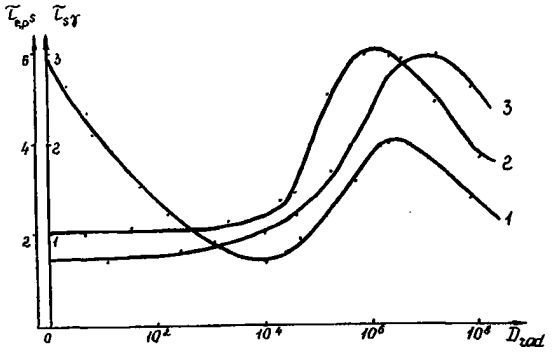


Fig.16. H-atoms life-time τ_S in H-maser storage box vs irradiation doses. 1 - E_γ = 1.28 MeV ^{60}Co, 2 - E_e = 5.7 MeV - pulse electronic accelerator, 3 - E_p = 10 MeV.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Frequency Control of Hydrogen Masers Using High Accuracy Calibrations

J.-S. Boulanger and R.J. Douglas
Institute for National Measurement Standards
National Research Council
Ottawa, Canada K1A 0R6
(613) 993-5698, Fax (613) 952-1394

Abstract

We apply a method for calculating standard uncertainty and optimize frequency control of a hydrogen maser which is calibrated from time to time against a cesium fountain primary frequency standard having a very small uncertainty. Realistic noise models for two types of maser are used. We compare the uncertainty resulting from different fitting procedures during the calibration runs, compare causal and acausal frequency control, and find optimum weights for a series of calibration runs.

Introduction

Hydrogen masers are very stable in the medium term (hours and days). They are commercially available and they have a proven record of reliable operation for generating time scales; but they also have a tendency to drift in the long term. Moreover, they do not realize the definition of the second, which is based on the cesium transition. On the other hand the promising new type of primary frequency standard, the cesium fountain [1] [2], is just in its infancy and likely will not offer continuous operation in the near future. It will likely achieve an accuracy approximating the best stability of hydrogen masers. The question then arises to the best way of combining the strengths of hydrogen masers and cesium fountains, and avoiding the weaknesses of each. This can be done by controlling the frequency of a hydrogen maser time scale, and the question devolves into the optimum choice of methods for calibrating the frequency of the time scale from a regular series of planned calibrations.

The most obvious strategy is simply to use the most recent complete frequency calibration of the maser by the fountain. A possibly better strategy would employ an algorithm that used previous calibrations in addition to the most recent one. Both of these techniques could be implemented in real-time, for example by synthesizer control. Better still might be a post-processed time scale which also uses frequency calibrations 'from the future'. We would like to be able to evaluate these strategies in a quantitative manner, and choose the best one. When actually using one of these strategies, we need rigorous method to evaluate the standard uncertainty [3] of the estimated average frequency of the maser, over a general time interval, offset by any time relative to the calibration period.

We developed [4] a formalism to calculate rigorously the local oscillator's contribution to uncertainty of a cesium fountain running in a quasi continuous fashion. This formalism can also be used to characterize many time or frequency correlations that were previously thought intractable [5] [6]. In particular, it makes possible the explicit calculation of standard uncertainties associated with various extrapolation or interpolation schemes, in terms of the noise characteristics of the oscillator under study. The standard uncertainty in aver-

0-7803-1945-1/94 $3.00 © 1994 Canadian Crown Copyright

age frequency of a general, weighted least-squares fit is derived explicitly in the Appendix. Instead of just trying to get the best fitting to the measurements against the frequency standard, we are trying to get the best fitting to minimize the noise transferred to the useful period of interest [6]. Our formalism allows this to be done by choosing the weights for general weighted least-squares fitting.

Two types of weight optimization will be considered. The first is optimizing the choice of weights on the phase comparisons during a single calibration run. Some general guidance on this subject has been available [7]. To the best of our knowledge, ours is the first quantitative calculation for general extrapolation intervals that considers all five power-law types of noise, and is the first consideration of weighting schemes intermediate between fits with equal weights and fits to end points only.

The second type of optimization is the choice of weights to best span a week with 5 weekday calibration runs. Both real-time and post-processed types of algorithms are considered, for the mixed noise models of two types of hydrogen masers. To the best of our knowledge, these represent the first optimizations of weights for a mixed noise model.

Cesium Fountains and Hydrogen Masers

Anticipating a cesium fountain with a standard uncertainty of $10^{-14}\tau^{-1/2}$[8], which might be able to operate for runs of an hour a day, the question of frequency transfer to the intervals between calibration runs is of the utmost importance. This might exploit the hydrogen maser, to carry frequency accuracy to the level of 10^{-15} or better. The weighting procedure for characterizing the hydrogen maser frequency during the calibration run against the cesium fountain will affect the accuracy of the frequency transfer. Envisaging multiple calibration runs per week, optimistically 5 per week, what is the best weighting procedure for using these calibrations in an algorithm to determine the frequency over a given interval?

In the course of evaluating such frequency transfers, we assume the frequency is very well known during the calibration procedure (with the primary standard's random uncertainty fully specified in the noise model). We assume no further knowledge of the frequency of the time scale between calibrations.

The only other parameters we need to know are those of the noise model of the H maser frequency transfer oscillator and the phase comparison processs of the calibrations, which add to the noise sources of the primary standard. In a time laboratory with an ensemble of masers, these parameters would generally be known from maser intercomparisons within the laboratory. Generally the only problematic coefficient would be the longest-term noise: the random-walk frequency noise. We prefer to avoid the drift removal process. Table 1 gives the power law noise coefficients h_α for two noise models we will use in this work: a free-running hydrogen maser (type 1), used in [5]; and a low flux auto-tuned hydrogen maser (type 2), used in [6] [8]. The maser noise is assumed to be dominant.

In this paper we apply the formalism, suggested in [6] and developed in the Appendix, to the case of a time scale derived from a hydrogen maser which is calibrated, by least-squares fitting to phase comparison data, from time to time against a primary frequency standard. From the knowledge of the power law noise model, we can try to optimize the weights in the least-squares fitting of the time scale, depending on the amounts of different types of noise in the noise model.

Frequency Transfer from One Run

Although some good generic advice is available concerning frequency transfer [7] techniques from an interval of dense phase comparison data to optimally estimate frequency, we felt that it was worthwhile to examine quantitatively the merits of different weighting schemes (the two extremes being

	Free-running	Auto-tuned
α	type 1	type 2
2	2.7×10^{-24}	6.7×10^{-23}
1	2.9×10^{-30}	2.9×10^{-30}
0	2.9×10^{-27}	2.9×10^{-27}
-1	2.6×10^{-31}	7.2×10^{-31}
-2	7.4×10^{-36}	4.9×10^{-37}
f_h	1 Hz	1 Hz

Table 1: Power law noise characteristics of the masers used in this study.

equal weights and end-points only fits), for the five noise types and for different extrapolation intervals, most particularly for the specific types of extrapolations to be used in post-processing.

We first study the extrapolation of average frequency from a centered interval of measurement of duration t_c to a period τ. This type of extrapolation is suited for post-processing algorithms. For the five types of noise we have calculated the variation of the standard uncertainty in the average frequency over the interval τ, for 31 equally spaced points centered within the larger interval τ. We calculated results for each of the different weighting distributions 'a'-'f' shown in Figure 1, for a linear least-squares fit: $x_p = a + bt$. We chose the uniform-weight case (maximally robust) as the reference fitting procedure, and used an upper-frequency cutoff $f_h = (\Delta t)^{-1}$, where Δt is the time interval between adjacent points of the fitting data ($t_c = 30\Delta t$).

The standard uncertainty in average frequency over $[t, t+\tau]$ was calculated using the method derived in the Appendix. Equation 12 was evaluated using our expressions for the general two-interval covariance $< [x(t_2) - x(t_1)][x(t_4) - x(t_3)] > = \mathcal{I}(t_4 - t_1) + \mathcal{I}(t_3 - t_2) - \mathcal{I}(t_4 - t_2) - \mathcal{I}(t_3 - t_1)$, where $\mathcal{I} = (2\pi)^{-1} \int_0^\infty S_y(\omega)\omega^{-2}\{1 - \cos\omega\tau\}d\omega$. Useful analytic expressions for $\mathcal{I}(\tau)$ exist [4] [5] [6] for the five power law noise types.

Figure 2 shows, for the case of pure white phase noise, the ratio of the standard uncertainty for symmetric extrapolation (using weight distributions 'a'-'e' of Figure 1) to the standard uncer-

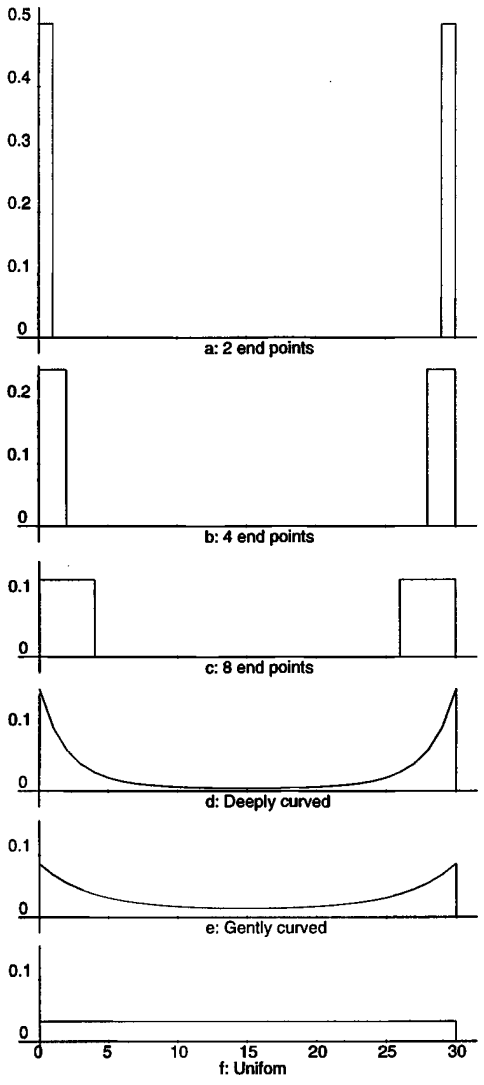


Figure 1: The different weights used for least-squares fitting and extrapolation in Figures 2-6. Weights 'a' (end points only) are the top curve, followed by weights 'b', 'c', 'd' and 'e'. Weights 'f' (bottom curve) are uniform (i.e. for unweighted least squares fitting) and are used for the reference standard uncertainty curve in each of Figures 2-6.

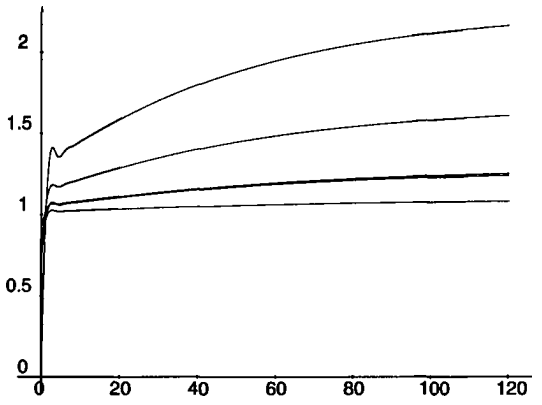


Figure 2: Average frequency uncertainty ratios for pure white phase noise ($\alpha = 2$). Weighted least-squares fits to 31 points are extrapolated symmetrically from t_c to τ. The standard uncertainty ratio to that for the unweighted fit (weights 'f') is plotted vs $[\tau - t_c]/\Delta t$. Weights 'a' (end points only) give the top curve (worst case), followed by weights 'b', 'c', 'd' and 'e'.

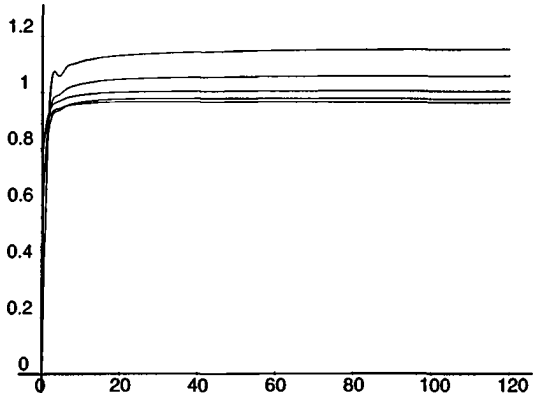


Figure 3: Average frequency uncertainty ratios for pure flicker $(1/f)$ phase noise ($\alpha = 1$). Weighted least-squares fits to 31 points are extrapolated symmetrically, and their standard uncertainty plotted as a ratio, with the unweighted fit (weights 'f') as the reference. Weights 'a' (end points only) give the top curve (worst case), followed by weights 'b', 'c', 'd' and 'e'.

tainty for the same extrapolation done with a linear unweighted least-squares fit to the data points (weight distribution 'f'). Similar uncertainty ratio curves are shown in Figure 3 for pure flicker phase noise, in Figure 4 for pure white frequency noise, in Figure 5 for pure flicker frequency noise and in Figure 6 for pure random walk frequency noise. All the curves are the uncertainties, $u_y(t_c, \tau)$, calculated for a given weighting procedure divided by the uncertainty of the uniform weighting procedure for the same conditions.

We have not computed the best weighting procedure for each noise type in this example. The structure of the equations to solve is too complex: for a linear fit of N points, it involves the simultaneous solution of $(N-1)$ third order equations. A different distribution of weights would be expected to be optimal for each interval τ and each noise type. With $N = 30$ this problem is beyond the scope of this paper, although it appears possible to linearize this problem in large measure.

Nonetheless, we find in our results some useful insights into the choice of fitting weights for different

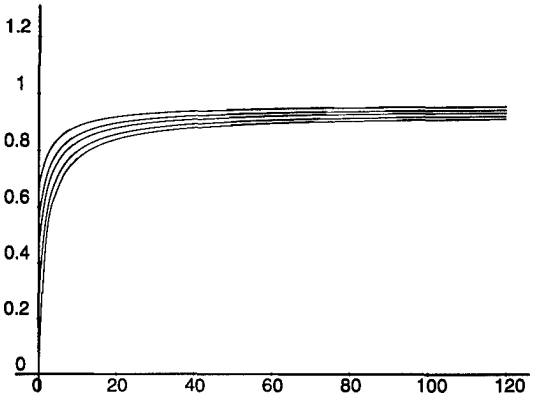


Figure 4: Average frequency uncertainty ratios for pure white frequency noise ($\alpha = 0$). Weighted least-squares fits to 31 points are extrapolated symmetrically, and their standard uncertainty plotted as a ratio, with the unweighted fit (weights 'f') as the reference. Weights 'a' (end points only) give the bottom curve (best case), followed by weights 'b', 'c', 'd' and 'e'.

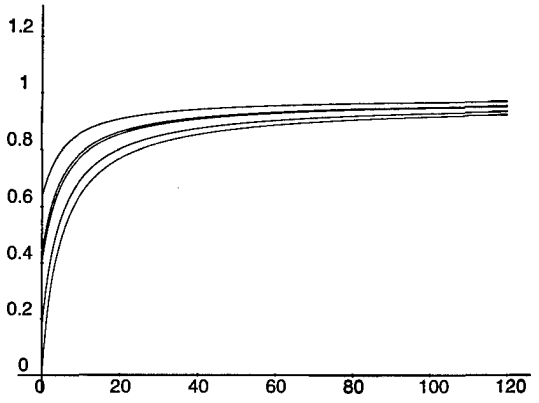


Figure 5: Average frequency uncertainty ratios for pure flicker frequency noise ($\alpha = -1$). Weighted least-squares fits to 31 points are extrapolated symmetrically, and their standard uncertainty plotted as a ratio, with the unweighted fit (weights 'f') as the reference. Weights 'a' (end points only) give the bottom curve (best case), followed by weights 'b', 'c', 'd' and 'e'.

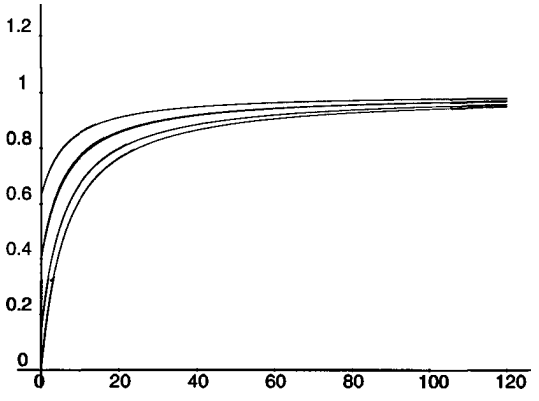


Figure 6: Average frequency uncertainty ratios for pure random walk frequency noise ($\alpha = -2$). Weighted least-squares fits to 31 points are extrapolated symmetrically, and their standard uncertainty plotted as a ratio, with the unweighted fit (weights 'f') as the reference. Weights 'a' (end points only) give the bottom curve (best case), followed by weights 'b', 'c', 'd' and 'e'.

types of noise and different ranges of extrapolation. In the case of white or flicker phase noise, it is seen that except for very short extrapolations the uniform weighting is the best procedure to use. It is interesting to note that the two end points procedure is always the way to get a best fit to the experimental data itself. It is only for the long term noise that this procedure is still the best, although marginally. We can already see that for calibration periods for which the Allan deviation is dominated by white or flicker phase noise, the best performance for extended two-sided extrapolation is an equal weighting for all the calibration points. In this example, the gain may be as much as a factor of 2 over the two end points procedure.

For our example, there is an advantage to using the end-point fitting for extracting average frequency in the presence of random noise dominated by white frequency noise, flicker frequency noise or random walk frequency noise. The advantage ranges from very large (for relatively small extrapolations) to about 10% for large symmetric interval extrapolations. This is smaller than the 20% advantage attributed to end point fitting for other cases [7], and may encourage those who value robustness over accuracy to use unweighted least-squares fitting. We value the slightly better accuracy of the end-point method, but we appreciate even more its computational simplifications, that allow us to do further optimizations and attempt to further reduce the standard uncertainty.

For the case of forward (one way only) extrapolation, the differences between various weighting procedures for other types of noise are negligible (less than 2%), except for the cases of white or flicker phase noise ($\alpha > 0$). For short one-way extrapolations (shorter than the calibration period) the equal weighting procedure is best for these cases. For longer extrapolations, and for $\alpha \leq 0$, the two end point procedure gives a marginal gain of less than 2% over the equal-weight least-squares fit.

Calibration runs of about one hour will be dominated by white frequency noise. We can then use the simple two end points procedure to establish the best frequency transfer accuracy for multiple calibration runs. For this procedure the standard

uncertainty for multiple calibration runs can be calculated more easily than in the general case.

Postprocessing Advantage

The ideas of frequency control are often aimed at the production of a real-time signal, and only use frequency error signals from the past. These might be classed as 'causal' servos, to distinguish them from the type of frequency control that is sometimes used in timescale work, which we might term 'acausal frequency control servos', to emphasize that the correction to the frequency can use information 'from the future'. In many important applications, the absence of a real-time output with all corrections applied is of little concern. For example, two laboratories wishing to intercompare cesium fountain frequencies could wait several weeks for postprocessing calculations to be done to bridge the breaks expected in the calibration schedule at either laboratory. We wish to examine quantitatively the advantage offered by postprocessing over real-time frequency control.

The reference real-time frequency control system requires some discussion on the degree of frequency control during the calibration run. There are two extremes. The first is a loose lock in frequency: after a calibration run (an hour in duration, in our example) is complete, the frequency of the maser is reset (through the synthesizer control, for example). The other extreme would have an output tightly locked in phase to the cesium fountain during the calibration run, followed by a frequency lock to the fitted frequency of the calibration run.

In Figure 7 we graph the advantage of the tight phase-lock frequency control in reducing the standard uncertainty in average frequency of a free-running (type 1) hydrogen maser. The advantage is relative to the method which only resets the synthesizer. The phase-lock type of frequency control removes the noise of the maser during the calibration run, giving it an advantage that remains significant for extrapolation intervals many times longer than the calibration interval. Therefore, in the one-run examples that follow, this type is used as the best 'real-time' reference.

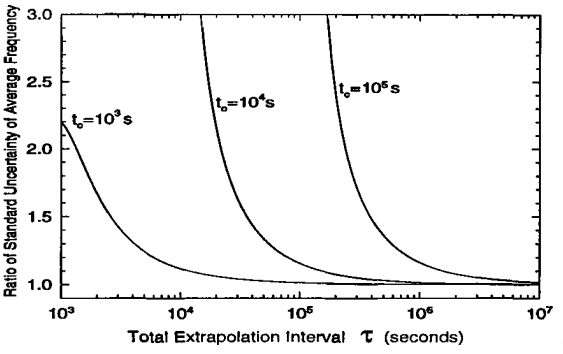


Figure 7: Comparison of two real-time frequency control methods for extrapolation. A tight phase-lock over the calibration interval t_c is added to synthesizer resetting, and the combined method compared with simple synthesizer resetting. τ starts at the end of t_c for synthesizer resetting, but at the beginning of t_c for the tight phase-lock method. The vertical axis is the ratio, for the two methods, of the standard uncertainty in average frequency over the interval τ for a free-running (type 1) maser noise model. In each case the synthesizer is reset to the end-point average frequency after $[t_c]$.

We use the two power law models for the maser noise given in Table 1, representing a free-running hydrogen maser (type 1) and an auto-tuned maser (type 2). The frequency fit chosen in these examples is the end-point method. We consider a single calibration interval t_c and calculate the ratio of the standard uncertainty of the average frequency over an interval τ for the best real-time frequency control to the symmetrically extrapolated time interval τ. The postprocessing advantage is up to a factor of two. The postprocessing advantage is plotted versus τ with different t_c's in Figure 8 for our free-running (type 1) maser model, and in Figure 9 for our auto-tuned (type 2) maser model.

A postprocessing advantage of two is really quite significant. To achieve the same improvement in the maser ensemble could be done - by increasing the maser ensemble size by four times! The postprocessing advantage of greatest interest to us is for τ representing extrapolation to the time inter-

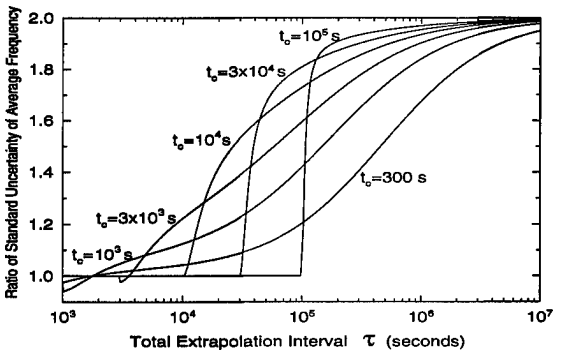


Figure 8: Postprocessing advantage over real-time frequency control for reducing the standard uncertainty of average frequency, using a free-running (type 1) maser model. The average frequency is extrapolated to an interval τ, from a calibration interval t_c. For postprocessing, t_c is centred within τ; for real-time control t_c is at the beginning of τ, although within τ in order to represent the 'best real-time' frequency control for a single calibration interval.

val between calibrations - which we expect would be between 1 day and 1 week. Initial interlaboratory frequency intercomparisons between cesium fountains, before regular calibration schedules can be set up, may require extrapolation times longer than 1 week for minimum uncertainty.

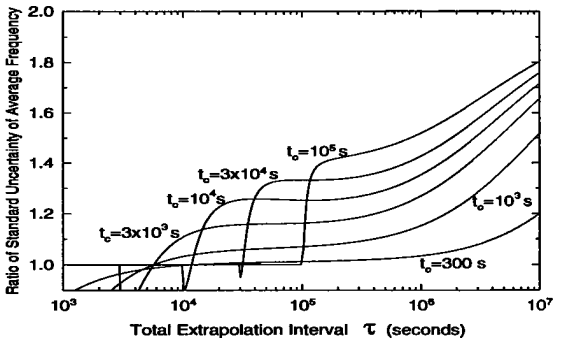


Figure 9: Postprocessing advantage over real-time frequency control for reducing the standard uncertainty of average frequency, using an auto-tuned (type 2) maser model.

Optimum Use of Multiple Calibrations

Multiple calibration runs can be used to servo the frequency of the hydrogen maser for a specific period, arbitrarily placed with respect to the calibration runs. Whatever algorithm is used to carry the frequency accuracy, as long as it can be reduced to a least-squares fit to a polynomial (or to even more general basis functions with linear coefficients), we can apply our technique to determine the best weighting procedure. Since for each calibration interval an independent arbitrary phase difference will have been fitted, only frequency information can be extracted from each calibration run. We develop here the simple weighted average of frequency as an example of weight optimization.

We consider calibration intervals long enough (an hour) to be in the regime where the two end point method is the optimum method of calibrating the maser frequency, and consider M such calibration intervals $[t_i, t_i + \tau]$. If a weighted average of the M calibrations is used, then the standard uncertainty in the average frequency over an interval $[t, t + \tau], u_y^2(t, \tau)$ is

$$u_y^2(t, t+\tau) = \left\langle \left[\frac{\{x(t+\tau) - x(t)\}}{\tau} - \sum_{i=1}^{M} w_i \frac{\{x(t_i+\tau_i) - x(t_i)\}}{\tau_i} \right]^2 \right\rangle. \quad (1)$$

Assigning a weight of -1 to the interval $[t, t + \tau]$, defining τ_0 as being equal to τ, Equation 1 can be rewritten as

$$u_y^2(t, t+\tau) = \left\langle \left[\sum_{i=0}^{M} \frac{w_i}{\tau_i} \{x(t_i + \tau_i) - x(t_i)\} \right]^2 \right\rangle. \quad (2)$$

A solution for the optimum weighting procedure is relatively easy to find since the minimum value for $u_y^2(t, t + \tau)$ is to be found for values of w_i satisfying $\frac{\partial}{\partial w_k}[u_y^2(t, t + \tau)] = 0$, so that after taking the derivative and separating out the $i = 0$ term

$$\sum_{i=1}^{M} \frac{w_i}{\tau_i \tau_k} \langle [x(t_i + \tau_i) - x(t_i)][x(t_k + \tau_k) - x(t_k)] \rangle$$
$$= \frac{1}{\tau_k} \langle [x(t + \tau) - x(t)][x(t_k + \tau_k) - x(t_k)] \rangle. \quad (3)$$

We use $M - 1$ of these equations, and for the M^{th} equation we use the normalization equation of the weights: $\sum_{i=1}^{M} w_i = 1$. This gives M simultaneous linear equations in the M unknown weights. The general interval covariance has analytic forms for our noise model, in terms of the $\mathcal{I}$-function [5]. If we define the $M \times M$ matrix $\mathbf{F}$: $F_{1,j} = 1$ for $j = 1..M$, $F_{i,j} = \frac{1}{\tau_i \tau_j}[\mathcal{I}(t_i + \tau_i - t_j) + \mathcal{I}(t_j + \tau_j - t_i)$ - $\mathcal{I}(t_i + \tau_i - t_j - \tau_j)$ - $\mathcal{I}(t_i - t_j)]$ for $i = 2..M$ and $j = 1..M$, and define $\vec{r}$: $r_1 = 1$ and $r_j = \frac{1}{\tau \tau_j}[\mathcal{I}(t + \tau - t_j)$ + $\mathcal{I}(t_j + \tau_j - t)$ - $\mathcal{I}(t + \tau - t_j - \tau_j)$ - $\mathcal{I}(t - t_j)]$ for $j = 2..M$. The M dimensional weights vector $\vec{w}$ is $\mathbf{F}^{-1} \cdot \vec{r}$.

We have solved for the optimum weights of the hydrogen maser calibrations to give the lowest standard uncertainty in average frequency over one week. The results are shown in Figure 10 for a free-running (type 1) maser. The week is best spanned by weighting the Monday and Friday runs more heavily, to account for the weekend gap in calibrations. If adjacent weeks calibration runs are also available, the optimum weights are shown by the lower curves of Figure 10, and are improved by a small admixture from the preceding Friday and the following Monday. In Figure 11, similar results are plotted for an autotuned (type 2) hydrogen maser noise model. The optimum weights have a slower variation through the weeks, and the three-week optimum has several % of the weight on points that are a full week from the calibration runs of the central week. In Figure 11, the lowest curves are the optimum weights for the case where five weeks of calibration runs are available to determine the average frequency over the central week.

For either maser model, the optimization of weights to apply to each run over multiple weeks gives about a 20% improvement in accuracy from the equal-weight case. It is not a large improvement, but it is almost free. Cascaded with the other advantages discussed earlier, and discussed in detail below, it results in a factor of 2.2 improvement in the accuracy transferrable with a free-running (type 1) hydrogen maser; and an improvement of 64% for the auto-tuned (type 2) maser.

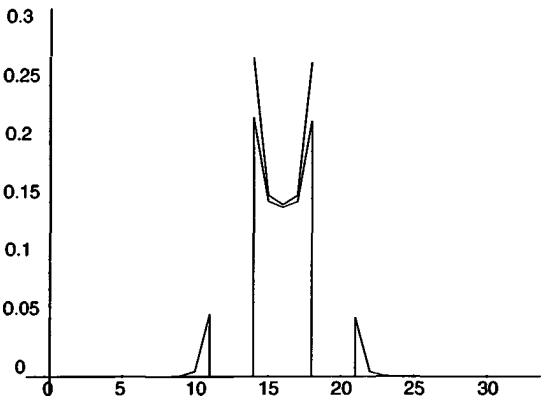


Figure 10: Weights vs date in days for optimum post-processing of 5 equal one-hour calibration runs per week, for a free-running (type 1) maser. The optimum is defined by the lowest standard uncertainty in average frequency over the week (days 13-19). The five runs are weekday runs, with none taken on weekends. The higher weights are optimum if only one week's calibrations (days 14-18) are used, the lower weights are optimum if adjacent weeks' runs are also used.

Other interesting strategies are beyond the scope of this work. Longer runs on Monday and Friday and/or early-Monday and late-Friday calibration runs could be invoked to further improve the performance. Our methods allow for weight optimization for any set of calibration runs, and for calculating the resulting standard uncertainty in average frequency.

Net Improvement in Accuracy

In Tables 2 and 3, the standard uncertainty in average frequency over a week, due to unfiltered maser noise, is presented for a series of seven frequency control strategies. Each uses five one-hour-long calibration runs each week (one at the same time each weekday) in a different way. Table 2 shows the results for a free-running maser (type 1 noise

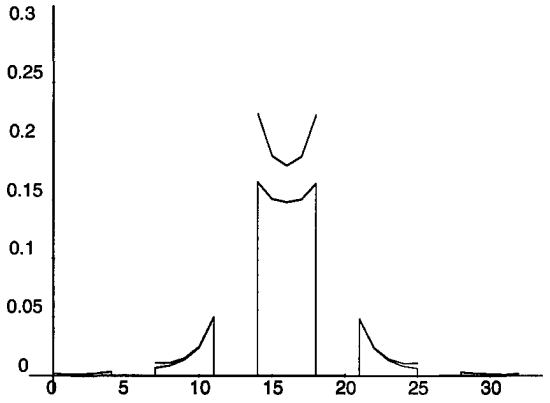


Figure 11: Weights vs date in days for optimum post-processing of 5 equal one-hour calibration runs per week, for an auto-tuned low-flux (type 2) maser. The highest weights are optimum if only one week's calibrations (days 14-18) are used, the middle weights are optimum if the middle three weeks' runs are used, and the lowest weights are optimum if all five weeks' runs are used.

method	$u_y(7d)$	Adv.	Cum. adv.
I	1.79×10^{-15}		1.00
II	1.77×10^{-15}	1.2%	1.01
III	1.73×10^{-15}	2.3%	1.04
IV	0.97×10^{-15}	77%	1.83
V	0.97×10^{-15}	1.1%	1.85
VI	0.81×10^{-15}	19%	2.21
VII	0.81×10^{-15}	0%	2.21
$\sigma_y(7d)$	5.48×10^{-15}		

Table 2: Reduction of standard uncertainty in average frequency at 7 days, for a free-running (type 1) maser, controlled by different methods from five 1-hour calibrations per week. The % advantage for each method is the accuracy improvement over the previous method (II over I, etc.). The last column gives each method's cumulative advantage over method I. The Allan deviation $\sigma_y(\tau = 7d)$ is also given.

method	$u_y(7d)$	Adv.	Cum. adv.
I	1.15×10^{-15}		1.00
II	1.15×10^{-15}	0%	1.00
III	1.12×10^{-15}	2.4%	1.02
IV	0.87×10^{-15}	29%	1.32
V	0.83×10^{-15}	4.7%	1.39
VI	0.70×10^{-15}	18%	1.64
VII	0.70×10^{-15}	.1%	1.64
$\sigma_y(7d)$	1.72×10^{-15}		

Table 3: Reduction of standard uncertainty in average frequency at 7 days, for an auto-tuned (type 2) maser, controlled by different methods from five 1-hour calibrations per week. The Allan deviation $\sigma_y(\tau = 7d)$ is also given.

model in Table 1), and Table 3 shows the results for an auto-tuned maser. Tables 2 and 3 also show the incremental improvement of one method over the previous one (Method II over method I, etc.), and the cumulative improvement in accuracy over method I. The seven methods I-VII are discussed, in order, below.

Method I: A real-time method. During each calibration run, no new frequency control is applied to the hydrogen maser. At the end of each calibration run the maser frequency is set to the frequency determined by a linear least-squares fit to the phase data over the past hour (in this specific calculation 121 points were used). Although this might appear to be a case of equal weights, the Friday calibration is most important since it is used to span a 72-hour interval instead of a 24-hour interval as do the Monday-Thursday runs. It is the reference method. The standard uncertainty due to the maser noise is significantly smaller than the Allan deviation of the maser at one week (by a factor of 3.06 for the free-running (type 1) maser, and 1.50 for the autotuned (type 2) maser), reflecting the noise components filtered out by the calibrations and the application of frequency control method I.

Method II: A real-time method. During each cal-

ibration run, no new frequency control is applied to the hydrogen maser. At the end of each calibration run the maser frequency is set to the frequency determined by an end-point (two-point) fit over the past hour. This method extrapolates 24 hours (four times) and 72 hours (once). It gives a 1.2% improvement over Method I for the free-running maser, and no significant improvement for the auto-tuned maser. (1.01 and 1.00 times better than method I).

Method III: A real-time method. During each calibration run, a tight phase-lock control is applied to the hydrogen maser, so that it follows the phase of the cesium fountain. At the end of each calibration run the maser frequency is set to the frequency determined by an end-point fit over the past hour. This method only extrapolates 23 hours (four times) and 71 hours (once). It gives a 2.3% improvement over Method II for the free-running maser, and 2.4% for the auto-tuned maser. (1.04 and 1.02 times better than method I).

Method IV: A post-processed (non real-time) method. This method extrapolates each calibration run independently, using the end-point fit. It is available with a maximum delay of 36.5 hours (and could be available in real-time for about half of each week). It extrapolates backwards and forwards in time by 11.5 hours (three times); backwards in time by 35.5 hours and forwards in time by 11.5 hours (once); and backwards in time by 11.5 hours and forwards in time by 35.5 hours (once). The weights for the Monday and Friday calibrations are 1.5 times the weights of the Tuesday through Thursday runs. It gives 77% improvement over Method III for the free-running maser, and 29% for the auto-tuned maser. (1.83 and 1.32 times better than method I).

Method V: A post-processed (non real-time) method. This method considers each week's calibration runs as a set, and applies to them the optimum weights, shown in the top curves of Figures 10 and 11. The maximum delay is 5.5 days, and the time scale could still be available in real-time 1.5 days/week. It gives a 1.1% improvement over Method IV, whose time-interval based weights are close to optimal for the free-running maser, and a 4.7% improvement for the auto-tuned maser. (1.85 and 1.39 times better than method I).

Method VI: A post-processed (non real-time) method. This method considers each group of three weeks of calibration runs as a set, and applies to them the optimum weights to obtain the best average frequency over the middle week. These weights are shown in the lower curve of Figure 10 and the middle curve of Figure 11. The maximum delay in availability of this time scale is 12.5 days. It gives 19.4% improvement over Method V for the free-running maser, and 18% improvement for the auto-tuned maser. (2.21 and 1.64 times better than method I).

Method VII: A post-processed (non real-time) method. This method considers each group of five weeks of calibration runs as a set, and applies to them the optimum weights to obtain the best average frequency over the middle week. These weights are shown in the lower curves of Figures 10 and 11. The maximum delay in availability of this time scale is 19.5 days. It gives negligible improvement over Method V for the free-running maser, and 0.1% improvement for the auto-tuned maser. (2.21 and 1.64 times better than method I).

Our method for calculating the standard uncertainty for realistic noise models has allowed us to compare a wide variety of algorithms for treating one particular calibration schedule. We have optimized the algorithm in several ways, and have identified ways to improve the accuracy of the maser frequency control by 2.2 and 1.64 times. We find that the expected weekly asymptote for a single auto-tuned (type 2) maser could reach $7 \times 10^{-16} n^{-1/2}$ for n weeks. Thus a flicker floor and accuracy of 10^{-16} for the cesium fountain is accessible for periods of a year with current masers carrying the time scale. The reliability of a cesium fountain which might do this seems to be a major challenge, perhaps comparable to the challenge of making a cesium fountain with a flicker floor and accuracy of 10^{-16} !

Conclusions

We have shown that using the powerful accuracy metrics defined in [6], we can evaluate rather general fitting procedures and get the best accuracy from a single calibration run. We find that in the cases of interest to us, the advantages are less than we had construed from the literature for the endpoint fitting over equal-weight least-squares fitting. We find 1% for extrapolation forward in time, and 10% for post-processing extrapolation forward and backwards in time. For other cases, depending on the type of noise present and on the time intervals under consideration, the best weighting procedure for a single interval of data can improve the accuracy substantially: two-fold or more.

For small extrapolations in time from calibration runs, the tight phase lock method of frequency control (see Method III above) can greatly improve average frequency accuracy, but in the cases of interest to us the advantage is small (less than 2.5%).

The accuracy advantage to be gained from post-processing is up to a factor of two for pure random walk frequency noise. For our two maser models the advantage is smaller, but still substantial: 77% and 29%.

Optimizing the weights of the calibration runs grouped, and identifying the optimum number of weeks to group for postprocessing, gives a further 21% and 25% improvement in the accuracy of the average frequency.

The overall set of small improvements combines to form a substantial improvement in the accuracy, of 2.2 and 1.64 times. In contrast to simulation methods, where 0.1% accuracy is difficult to achieve, our analytic methods has no difficulty in dealing with the high accuracy needed to keep track of small improvements. Our methods should be amenable to direct search methods for optimizing small numbers of weight parameters.

The standard uncertainty in average frequency, $u_y(t,\tau)$, can be calculated by our methods for all the most commonly used fitting procedures. Our methods allow anyone involved in frequency measurement or control to calculate and specify the uncertainty, in even intricately derived fits, in the recommended way [3] from the familiar coefficients of the power law noise model.

We have developed and used a formal structure for optimizing weights to be used in least-squares fitting of data. It is a second layer of optimization beyond the choice of parameters which minimize the sum-of-squares of residuals. In this sense the weight optimization is "meta-fitting". The meta-fitting optimization uses a specific noise model (with known autocorrelation functions), a specific metric (such as the standard uncertainty in average frequency over a week - at points which are not part of the original data set) and yields a set of weights which give the best answer to the specific quantity sought. To the best of our knowledge this is the first time that such a procedure has been used in, or proposed for, data processing.

References

[1] A. Clairon, C. Salomon, S. Guellati and W.D. Phillips, **Ramsey Resonance in a Zacharias Fountain**, *Europhys. Lett.*, **16**, pp. 165-170 (1991).

[2] K. Gibble and S. Chu, **Laser-cooled Cs Frequency Standard and a Measurement of the Frequency Shift due to Ultracold Collisions**, *Phys. Rev. Lett.*,**70** 12, pp. 1771-1774 (1993).

[3] **Guide to the Expression of Uncertainty in Measurement**, International Organization for Standardization, Geneva, 1993. First edition. (Published in the name of BIPM, IFCC, ISO, IUPAC, IUPAP and OIML.)

[4] R.J. Douglas and J.-S. Boulanger, **Local Oscillator Requirements for Timekeeping in the $10^{-14}\tau^{-1/2}$ era**, *Proceedings of the 1992 IEEE Frequency Control Symposium*, pp. 6-26 (1992).

[5] D. Morris, R.J. Douglas and J.-S. Boulanger, **The Role of the Hydrogen Maser Frequency transfer from Cesium Fountains**,

Japanese Journal of Applied Physics,**33**, pp. 1659-1668 (1993).

[6] R. J. Douglas., J.-S. Boulanger and C. Jacques, **Accuracy Metrics for Judging Time Scale Algorithms**, *Proc. 25th Annual PTTI Applications and Planning Meeting*, pp. 249-266 (1993).

[7] D.W. Allan, **Time and Frequency (Time-Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators**, *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* **UFFC-34**, pp. 647-654, (1987).

[8] J.-S. Boulanger, D. Morris, R. J. Douglas. and M.-C. Gagné, **Hydrogen Masers and Cesium Fountains at NRC**, in *Proc. 25th Annual PTTI Applications and Planning Meeting*, pp. 345-356 (1993).

[9] J.A. Barnes, A.R. Chi, L.S. Cutler, D.J. Healey, D.B. Leeson, T.E. McGunigal, J.A. Mullen Jr., W.L. Smith, R.L. Sydnor, R.C. Vessot and G.M.R. Winkler: **Characterization of Frequency Stability**, *IEEE Trans. Instrum. Meas.* **IM-20**, pp. 105-120 (1971).

Appendix: A Metric for Choosing Weights in Weighted Least-Squares Fits

We develop here analytic expressions for the standard uncertainty in the average frequency of the fit in a weighted least-squares fit to phase data. The standard uncertainty is now the internationally recommended [3] way of specifying uncertainty. With our procedures the standard uncertainty can be evaluated for a broad class of noise models long used in frequency standards work, for any set of fitting points, for any extrapolation or interpolation interval, for any set of basis functions with linear coefficients, and for any least-squares weighting.

The noise model $x_m(t)$ is the modeled phase difference between the calibrating frequency standard (cesium fountain) and the standard being calibrated (hydrogen maser). The noise model is taken as being the sum of a deterministic part (which could include a phase offset, frequency offset and frequency drift) and a random noise part, $x_0(t)$. The random noise includes the "full" noise model that is usually used in discussions of frequency standard stability [9]. Expressing the five terms in terms of the spectral density of the mean-square of the fluctuations in $\frac{dx_0(t)}{dt}$ (or $y_0(t)$) at a frequency f, $S_y(f)$, each noise term is described by an amplitude h_α which is taken to be independent of any time translations (stationarity and random phase approximations).

$$S_y(f) = \sum_{\alpha=-2}^{2} h_\alpha f^\alpha \qquad (4)$$

Weighted Least-Squares Fits

Weighted least-squares fitting chooses the n linear coefficients d_l of the n basis functions $g_l(t)$, to arrive at a function $x_p(t)$ which will be used for interpolation or extrapolation. In frequency standards work, we would usually fit a phase offset, a frequency offset, and sometimes a drift rate and higher terms such as daily or seasonal fluctuations.

$$x_p(t) = \vec{d} \cdot \vec{g}(t) = d_1 + d_2 t + d_3 t^2 + \sum_{l=4}^{n} d_l g_l(t). \quad (5)$$

The coefficient vector $\vec{d}$ is chosen to minimize the sum over the N fitting points with phase difference values of $x(t_i)$ at times t_i

$$L_2 = \sum_{i=1}^{N} W_i [x(t_i) - \vec{d} \cdot \vec{g}(t_i)]^2 \qquad (6)$$

where the weight W_i is applied to the square of the i^{th} residual. Setting the n derivatives of L_2 equal to zero gives a set of n linear equations which can be solved for the n fitting coefficients of $\vec{d}$: $\mathbf{G}\vec{d} = \vec{s}$, where $\mathbf{G}$ is an $n \times n$ matrix with elements $G_{qr} =$

$\sum_{i=1}^{N} W_i g_q(t_i) g_r(t_i)$, and $\vec{s}$ is an n-dimensional vector with elements $s_r = \sum_{i=1}^{N} W_i x_0(t_i) g_r(t_i)$. For the purposes of modelling the standard uncertainty, we use $x_0(t_i)$ to model $x(t_i)$, since it can be shown [6] that any general offset in phase, offset in frequency or a linear frequency drift is exactly absorbed by the fit.

We can now study the effects of the weighted least-squares fit reacting to the noise model, and we are not restricted to the studying the variance at the fitting points but we can do this at any time t. The expected variance in $x(t)$ from the fit $\vec{d} \cdot \vec{g}(t)$ can be calculated in terms of the autocorrelation function

$$\left\langle \left[x(t) - \vec{d} \cdot \vec{g}(t) \right]^2 \right\rangle = \sum_{i=0}^{N} \sum_{j=0}^{N} D_i D_j \left\langle x_0(t_i) x_0(t_j) \right\rangle , \tag{7}$$

where $D_i(t) = W_i \sum_{q=1}^{n} \sum_{r=1}^{n} (\mathbf{G}^{-1})_{qr} g_r(t_i) g_q(t)$ and $D_0(t) = 1$. For the standard noise model, the autocorrelation function $\langle x_0(t) x_0(t) \rangle$ can be evaluated analytically [6], although the resulting expressions can challenge the dynamic range of conventional computing.

Fortunately, for the purposes of frequency control, better behaved functions may be used [4], [5], [6]. We wish to evaluate the standard uncertainty in average frequency over an interval $[t, t+\tau]$, caused by the noise model as filtered by the weighted least-squares fitting procedure to the points $\{t_i\}$. Although the noise model is independent of time translations, clearly the standard uncertainty in average frequency, u_y, would be expected to depend on the offset of t from $\{t_i\}$, as well as the interval breadth τ. It is defined by

$$u_y^2(t,\tau) = \left\langle \left[\left\{ \frac{x_0(t+\tau) - x_0(t)}{\tau} \right\} - \left\{ \frac{\vec{d} \cdot (\vec{g}(t+\tau) - \vec{g}(t))}{\tau} \right\} \right]^2 \right\rangle . \tag{8}$$

We note that $\{x_0(t+\tau) - x_0(t)\} = \sum_{j=1}^{N+1} [x_0(t_j) - x_0(t_{j-1})]$, if we define $t_{j=0} = t$ and $t_{j=N+1} = t+\tau$. Although it might be convenient to envisage the set of $\{t_j\}$ as an ordered set with $t_j > t_{j-1}$, it is not necessary to do so. Ordering the fitting points does not detract from the generality in any way, but we do not wish to restrict the values of t or $t+\tau$. We would like to re-express the $\vec{d} \cdot \{g(t+\tau) - g(t)\}$ as a sum over only differences of the form $x_0(t_i) - x_0(t_j)$. We note that we can expand $x_0(t_i) = x_0(t_1) + \sum_{j=2}^{i} \{x_0(t_j) - x_0(t_{j-1})\}$, so that $\vec{d} \cdot \{\vec{g}(t+\tau) - \vec{g}(t)\}$ is equal to

$$\begin{aligned} & \mathbf{G}^{-1} \textstyle\sum_{i=1}^{N} W_i x_0(t_i) \cdot \{\vec{g}(t+\tau) - \vec{g}(t)\} \\ = \; & \textstyle\sum_{i=1}^{N} W_i \sum_{j=2}^{i} \{x_0(t_j) - x_0(t_{j-1})\} \\ \times \; & \textstyle\sum_{q=1}^{n} \sum_{r=1}^{n} (\mathbf{G}^{-1})_{qr} g_r(t_i) \{g_q(t+\tau) - g_q(t)\} \\ + \; & x_0(t_1) \textstyle\sum_{i=1}^{N} W_i \sum_{q=1}^{n} \sum_{r=1}^{n} (\mathbf{G}^{-1})_{qr} g_r(t_i) \\ \times \; & \{g_q(t+\tau) - g_q(t)\} \end{aligned} \tag{9}$$

and the last term, multiplying $x_0(t_1)$, can be shown to be equal to zero. To show this, it is sufficient to show that $\sum_{i=1}^{N} W_i \vec{g}(t_i) \mathbf{G}^{-1} \vec{g}(t)$ is independent of t, or that $\sum_{i=1}^{N} W_i \vec{g}(t_i) \mathbf{G}^{-1}$ is equal to the vector [1,0,0,...0]. We observe that, from our definition of $\mathbf{G}$ and since $g_1 = 1$, $\mathbf{G}[1,0,0,...0] = \sum_{i=1}^{N} W_i \vec{g}(t_i)$, and premultiplying by $\mathbf{G}^{-1}$ completes this proof, provided only that g_1 is a constant. Thus $u_y^2(t,\tau)\tau^2$ is equal to

$$\begin{aligned} & \left\langle \left[\textstyle\sum_{j=1}^{N+1} \{x_0(t_j) - x_0(t_{j-1})\} \right. \\ - \; & \textstyle\sum_{i=1}^{N} W_i \sum_{j=2}^{i} \{x_0(t_j) - x_0(t_{j-1})\} \\ \times \; & \textstyle\sum_{q=1}^{n} \sum_{r=1}^{n} (\mathbf{G}^{-1})_{qr} \\ \times \; & \left. g_r(t_i) \{g_q(t+\tau) - g_q(t)\} \right]^2 \right\rangle . \end{aligned} \tag{10}$$

Collecting the expressions with the same difference term $\{x_0(t_j) - x_0(t_{j-1})\}$ allows us to write a useful form, namely

$$u_y^2(t,\tau)\tau^2 = \left\langle \left[\sum_{j=1}^{N+1} \tilde{D}_j(t,\tau) \{x_0(t_j) - x_0(t_{j-1})\} \right]^2 \right\rangle , \tag{11}$$

where for the range $2 \le j \le N$, $\tilde{D}_j(t,\tau) = 1 - \sum_{i=j}^{N} W_i \sum_{q=1}^{n} \sum_{r=1}^{n} (\mathbf{G}^{-1})_{qr} g_r(t_i) \{g_q(t+\tau) - g_q(t)\}$; $\tilde{D}_{j=1}(t,\tau) = 1$ and $\tilde{D}_{j=N+1}(t,\tau) = 1$. Multiplying the terms explicitly gives a computable

form for the standard uncertainty in average frequency:

$$u_y^2(t,\tau)\tau^2 = \sum_{j=1}^{N+1}\sum_{k=1}^{N+1}\tilde{D}_j(t,\tau)\tilde{D}_k(t,\tau)\times \langle[x_0(t_j)-x_0(t_{j-1})][x_0(t_k)-x_0(t_{k-1})]\rangle. \quad (12)$$

The utility of this form lies in the fact that it is a sum over functions of the general form $< [x_0(t_2)-x_0(t_1)][x_0(t_4)x_0(t_3)] >$, which are easier to compute for our full noise model - they are less divergent than the autocorrelation function of $x_0(t)$. The analytic forms for these functions, and some comments on their numerical evaluation, are given in our earlier work [6], [5], [4].

This is the form used for calculating the standard uncertainties for the different distributions of weights, in Figures 1-6. It is also a form which suggests the possibility of largely linearizing the search for the optimum weights for rather large data sets.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A HYDROGEN MASER FOR LONG-TERM OPERATION IN SPACE

Robert F.C. Vessot, David A. Boyd, Laurence M. Coyle, Richard F. Jendrock, Edward M. Mattison, and George U. Nystrom
Harvard-Smithsonian Center for Astrophysics
60 Garden St., Cambridge, MA 02138
Thomas E. Hoffman
Hoffman Design and Development
Marblehead, MA 01945

Introduction

We describe the design of a space-worthy hydrogen maser capable of four or more years of continuous operation in space. This maser system can be adapted for several future space applications, such as very long baseline interferometry, high precision time transfer, and tests of relativistic gravitation. The maser clock system is described, including details on its design for surviving very high levels of launch stress, operating with frequency stability, $\sigma_y(\tau) = 1 \times 10^{-15}$ for $\tau = 24$ hours over ambient temperature variations from 14 to 25 C, and magnetic field variations of $\pm$ 0.5 gauss. These specifications are intended to allow it to be used on a wide range of launch vehicles and space platforms. Tests of this maser clock system for approximately 1 year duration will be made in a low orbiting spacecraft by making short pulsed-laser time comparisons with a maser monitored via a common view GPS technique located in a laser-equipped earth station. The timing system and the event timer, with 10 ps time resolution, is described in an accompanying paper at this conference.[1]

Operation of the space-borne maser system is controlled and monitored using an on-board microprocessor that implements a number of automated procedures for controlling the maser and testing its operational parameters while in orbit. The entire system, including two reflector-detector-event timer modules, will operate on less than 95 watts. The package, as configured for mounting on the European Space Agency's EURECA spacecraft, weighs less than 110 kg.

Objectives

The purpose of this project is to design, build, and test a spaceborne atomic hydrogen maser clock system and evaluate its performance in space. Three tasks will be performed while in space:

- Measure the maser's frequency relative to international time scales using laser pulses from a ground-based satellite laser ranging station.

- Observe the maser's response to controlled changes in the maser's functions and from variations of its environment. Control and monitoring will be done through the spacecraft's telemetry and telecommand systems.

- Determine from frequency measurements and from monitored data, the maser's frequency stability over the duration of the mission, and the effect of ambient conditions on the maser's long-term frequency performance.

Overview

The main component under test is a hydrogen maser atomic clock system located in a spacecraft in a low earth orbit (altitude of approximately 500 km.) Time kept by the atomic hydrogen maser (H maser) clock will be compared with earth-based clocks by controlling the times of laser pulses sent from a laser ranging station. Their arrival times will be recorded at the spacecraft and the pulses will be reflected back to the ranging station by cube corner retroreflectors mounted on the spacecraft. A photodetector mounted with the retroreflectors will detect the arrival of a laser pulse at the spacecraft and will trigger a timing circuit (an "event timer") that records the pulse's arrival time in terms of the spacecraft maser's time scale. The pulse arrival times at the space craft will be corrected for propagation by measuring the two-way time of propagation at the earth station. An overview of the system is shown in Figure 1. Figure 2 shows a block diagram of the major system components.

Systems

The experimental package consists of (i) an H maser physics unit, (ii) one or two laser retroreflector/detector

U.S. Government work not protected by U.S. copyright

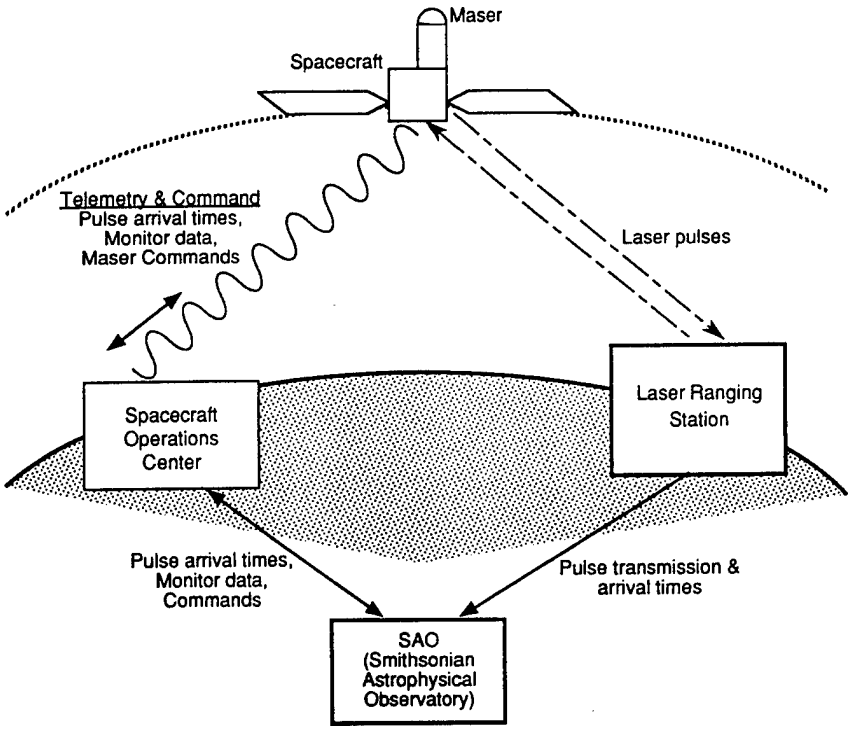


Figure 1

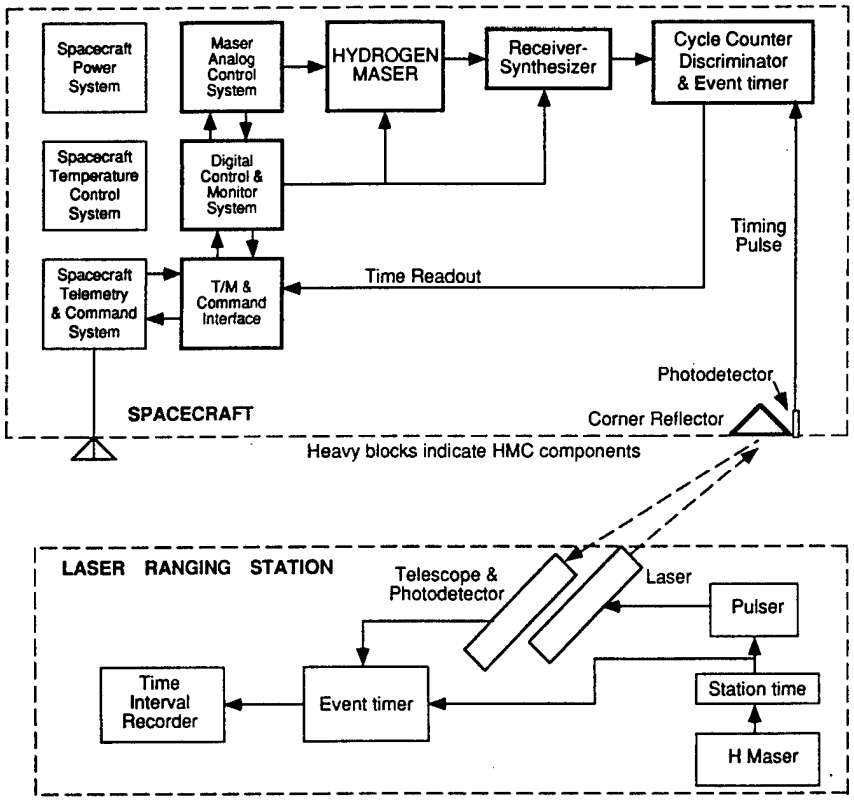
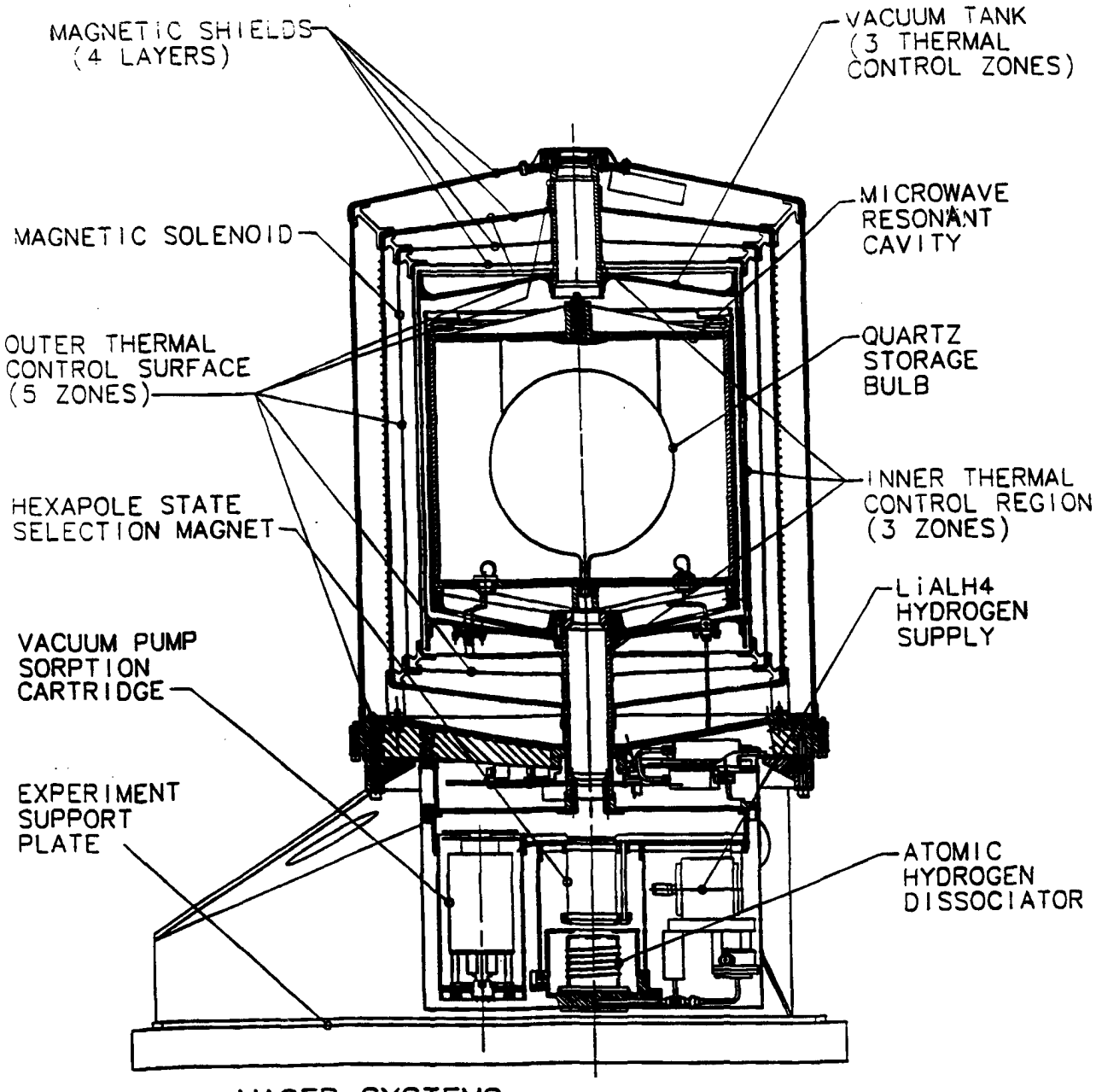


Fig. 2. Block diagram of major HMC experiment components



Cross section of maser and support structure on mounting plate, showing main maser systems.

Figure 3

units that detect the arrival of a laser pulse at the spacecraft and reflect the pulse back to the ranging station,and (iii) the electronics required to operate the maser, interface with the spacecraft and measure the arrival time of laser pulses.

The physics package

The H maser physics unit consists of a microwave resonant cavity assembly, an RF dissociator that creates a beam of hydrogen atoms, a permanent magnet state selector system, and vacuum pumps to remove expended hydrogen and other gases. The physics package is shown in cross-section in Figure 3.

The hydrogen beam that operates the maser is formed in a dissociator–state selector system mounted in the vacuum envelope opposite the entrance to the cavity resonator. The RF dissociator converts molecular hydrogen (H_2) into a beam of atomic hydrogen (H), and the state selection magnet focuses atoms that are in the desired quantum state into the quartz storage bulb. The vacuum manifold that contains the dissociator, state selection magnet, and sorption pumps is connected to the vacuum belljar by a titanium neck tube. Figure 3 shows the locations of the dissociator and hexapole state selection magnet.

The H_2 is supplied from (40 grams) of lithium aluminum hydride ($LiAlH_4$), contained in a stainless steel container whose temperature is controlled to maintain a constant hydrogen pressure. Hydrogen flow from the $LiAlH_4$ source into the maser is controlled by sensing the hydrogen pressure in the dissociator with a thermistor Pirani gauge, and regulating the temperature of a palladium silver diaphragm to keep the pressure constant.

In the dissociator, molecular hydrogen at a pressure of approximately 10^{-1} torr is led from a hydrogen source to a cylindrical glass bulb (~6 cm long,~ 3 cm diameter) mounted within the vacuum chamber. An RF coil that is connected through the vacuum envelope surrounds the dissociator bulb. Five watts of RF power at about 90 MHz is applied to produce a plasma discharge that dissociates the H_2. Atomic hydrogen is directed as a highly collimated beam of about 10^{14} H atoms per second into the hexapole state selection magnet which selectively focuses atoms in the desired atomic hyperfine state into the storage bulb. The hexapole magnet is approximately 8 cm long with a bore of about 3 mm and has a field between its pole tips of approximately 9 kilogauss.

Expended hydrogen is absorbed in two sorption cartridges operating at room temperature. These pumps scavenge only H_2. A small ion pump removes the usual non-hydrogen outgassing products from the metallic vacuum system and maintain a 10^{-6} torr vacuum.

The cylindrical vacuum tank is made of titanium alloy (6%Al-4%V) and contains a cylindrical microwave resonant cavity, within which is mounted the quartz hydrogen storage bulb. The three-part cavity (cylinder and two end plates) is made of Cer-Vit, a glass-ceramic material of a type often used for highly stable telescope mirrors. Its inner surface is coated with silver to create a conductive surface. A double Belleville spring clamps the cavity endplates to the cylinder with an axial force of approximately 480 lbs. The compression of the Belleville spring is adjusted so that the compressive force is nominally independent of the length of the holddown can. Under these conditions, the clamping force on the cavity, and thus the cavity's resonance frequency, is approximately independent of the thermal expansion of the hold-down can. The cavity mounting baseplate is attached near its center to the base of the vacuum tank, making the cavity structure essentially independent of dimensional changes in the outer vacuum envelope.

The spherical, 18-cm diameter, thin-walled fused silica (quartz) storage bulb with a mass of about 200 grams is epoxied to one end of the microwave cavity. Its inner surface is coated with a thin layer of Teflon. Hydrogen atoms enter and leave the bulb through a narrow tubular aperture. While in the storage bulb the atoms produce a very weak microwave signal at a frequency of 1.42 GHz that is coupled by a loop in the cavity and led, at a level of approximately –100 dBm, to a heterodyne receiver that phase-locks a 100 MHz VCXO. A second loop within the cavity incorporates a reverse-biased varactor tuning diode to adjust the cavity resonance frequency.

The magnetic field within the maser's resonant cavity is carefully controlled by an assembly of magnetic shields and a solenoid, shown in Figure 3. Four nested cylindrical molypermalloy magnetic shields surround the vacuum belljar. A magnetic field compensation system[2] is located within the outermost shield which consists of a small fluxgate magnetometer to operate a magnetic field servo system to null the axial field by controlling current to a coil wrapped around the next-to-outermost shield. This system attenuates ambient magnetic fields of ± 0.5 gauss to a level of a few microgauss within the belljar. A shielding factor, $S = \Delta H_{external} / \Delta H_{internal}$ in excess of 2×10^6 is

achieved by this system. A two-layer three section printed circuit solenoid closely fitted to the inside of the innermost shield is used to provide a uniform longitudinal magnetic field of about $5x10^{-4}$ gauss within the cavity.

The temperature of the resonant cavity and storage bulb is held constant by a 7-zone heat-added thermal control system. The vacuum tank constitutes the inner isothermal control surface and is controlled in three zones. The innermost magnetic shield surrounding the tank serves as the inner oven and two heaters are located at the joints near the neck. An outer oven consisting of three zones surrounds the tank and two additional zones control the temperature of the mounting plate and the uppermost end of the tank attachment structure. Figure 3 shows the locations of the heaters.

The inner magnetic shield is mounted to the outside of the titanium tank. The belljar is supported by a pair of tubular titanium necks, one connected to the main mounting flange and the other to an aluminum support can that surrounds the outer magnetic shield. The bottom of the support can is fixed to the main mounting flange. The three outer magnetic shields are separated by segments of open-celled foam that provide a semi-rigid support. This support scheme is designed to damp out shock and vibration in order to prevent work-hardening of the shields, which could lead to loss of magnetic shielding effectiveness. The outermost magnetic shield completely surrounds the maser physics package. The weight of the outer magnetic shields is transferred to the outer aluminum support can, and thence to the main mounting flange.

The main mounting plate is the primary point of attachment of the H maser. It supports the near belljar neck directly, and the far neck through the outer support can. It also supports the stainless steel vacuum manifold that houses the dissociator, the sorption pumps, and the state selector magnet. The main mounting plate is connected to the spacecraft by an aluminum structure that also houses the maser's electronics, the H_2 supply-flow controller and the ion pump. This assembly is shown conceptually in Figure 3.

Retroreflector and Photodetector Arrays

Time kept by the spaceborne hydrogen maser will be compared with time kept by clocks on the earth will be done using laser pulses. A description of the reflector-detector array and the 10 picosecond resolution event timer that encodes the arrival times of the pulses is given in reference 1.

Electronics

The electronics comprisise the following:

- Power conditioning circuits
- Computer and associated digital circuits for interfacing with the spacecraft's telecommand systems
- Circuits for controlling and monitoring the functions of the physics unit
- High-voltage (~2.5 kV) power supply for operating the maser's ion vacuum pumps
- RF oscillator and amplifier (~80 MHz, ~5 watts) for dissociating H_2 molecules into H atoms
- Heterodyne RF receiver and frequency synthesizer for phase-locking a crystal oscillator to the maser's output signal
- Photodetector and event timer for laser pulse timing

Power Conditioning

The experiment uses 28-volt DC power supplied by the spacecraft power systems. Isolated switching DC-DC converters produce ±15 volt and +5 volt power for operating the circuitry. Pulse-width modulated switching circuits are used to supply power to the physics unit's temperature control heaters. The power circuits are filtered to reduce radiated and conducted electromagnetic interference (EMI) to required levels. Average power, including electronic circuits as well as temperature-control heaters, is approximately 95 watts.

Computer and Spacecraft Interface

Standard digital circuitry provides the interface between the experiment and the spacecraft's telemetry and telecommand functions. A dedicated experiment microprocessor buffers signals between the experiment and the spacecraft, receiving telecommands and sending data to the spacecraft's mass storage system for telemetry to the ground station. The microprocessor also controls some of the physics unit's functions, acts upon telecommands, and preprocesses and formats information from the monitoring systems and event timer.

Three types of telecommand information are required by the experiment: (a) commands to adjust internally controlled parameters, such as magnetic field current or hydrogen beam flux; (b) commands that cause the computer to start an autonomous sequence of activities, such as measuring the maser's internal magnetic field; and (c) sets of data that update information in the computer's memory, such as opening gating circuits at the expected times of laser pulse arrivals. In addition, it

will be possible to upload revised programs to the experiment's computer if desired.

Control and Monitoring Circuits

Several functions of the HMC are controlled during the mission, as shown in Table 1.

The heater elements are bifilar-wound wires or flexible printed circuit heaters, depending upon location in the physics unit. Operation of the maser's circuits is monitored by converting controlled or measured parameters into voltages in a standard range and changing these voltages into digital form by a multiplexer and a digital-to-analog converter. Monitored voltages are read by the computer at preset intervals and transferred to the spacecraft's mass memory system for telemetry to the ground.

High-voltage Power Supply

The vacuum within the maser is maintained by a passive (non-powered) sorption cartridge pump that absorbs hydrogen, and by a small ion pump that removes residual gasses resulting from outgassing of internal surfaces. The ion pump is operated by a 2.5 kV, 100 microampere high-voltage power supply (HVPS) mounted integrally with the pump body. The HVPS provides a monitor voltage that is related to the current drawn by the ion pump. The HVPS is a sealed unit, and its high-voltage output is insulated to prevent corona discharge, EMI, and shock hazard.

Dissociator

The hydrogen dissociator uses RF power to make a plasma in low-pressure hydrogen gas that dissociates H_2 molecules into hydrogen (H) atoms. The frequency and amplitude of the dissociator oscillator are controlled to maintain a condition of optimum match. The dissociator oscillator and amplifier, which are located outside of the vacuum system are shielded to prevent EMI, while the matching circuit is shielded by the sealed metal vacuum envelope. The brightness of the red line in the atomic hyrrogen spectrum is measured by a photodetector and dielectric interference filter mounted outside of a small quartz window in the vacuum envelope. The dissociator bulb's support structure carries heat dissipated in the plasma to the outside of the vacuum envelope. The dissociator bulb is pressed by Belleville springs between two aluminum plates that are mounted to the vacuum manifold by aluminum rods with sufficient cross sectional area to carry the heat with minimum temperature gradient.

Receiver/Synthesizer

The output signal from the maser's resonant cavity is at a frequency of approximately 1.42 GHz and a power of approximately –90 to –100 dBm. A multiple-heterodyne receiver phase-locks a 100 MHz voltage-controlled crystal oscillator (VCXO) to the maser signal, producing a standard receiver output frequencies of 100 MHz at 0 dBm. An adjustable-frequency direct-digital synthesizer operating at approximately 405 kHz provides adjustment of the VCXO frequency.

Photodetector and Event Timer

The arrival of a laser pulse at a retroreflector is sensed by a photomultiplier tube (PMT) mounted within the array package. The PMT's output pulse is sent to a constant-fraction discriminator and event timer also located in the array package. Because the expected pulse length is considerably longer than the desired measurement precision of 10 picoseconds, and the pulse height can vary from pulse to pulse, the constant-fraction discriminator circuit is needed to produce a standardized logic pulse at a time that is independent of the laser pulse's amplitude.

The standardized pulses produced by the discriminator go to an event timer that records pulse arrival times in terms of a 100 MHz clock signal with a precision of 10 ps. The event timer consists of standard digital gates and registers, and a hybrid analog time interpolation circuit that provides 10 ps measurement resolution.

Summary of Size Weight and Power

The physics unit is a cylindrical structure 43 cm dia, 83 cm long weighing 74.5 kg. The electronics control box has dimensions 30 x 30 x 33 cm and weighs 11.4 kg. The maser receiver package has dimensions 16 x 15 x 12 cm and weighs 2.5 kg. The microprocessor has dimensions 26 x 16 x 14 cm. and weighs 3.8 kg.

The entire experiment will be located on a heat conducting support plate kept at about 20^0C; it will consume on average approximately 95 watts of 28 VDC power.

Acknowledgement

This work is supported by the George C. Marshall Space Flight Center, Huntsville Alabama.

Table 1 – Controlled Functions

Function	Sensor	Controlled parameter
Temperature (8 zones)	Thermistor	Heater current (pulse width modulated) (~50°C max)
Hydrogen flux	Thermistor	Heater current (pulse width modulated) (~50°C max)
Magnetic field in Cavity	None	Coil current (resistor ladder via relays) (~50 mA max)
Varactor diode volts	None	Diode voltager (resistor ladder via relays) (10 volts max)
Dissociator operation	RF diode	Frequency, power
Synthesizer frequency	None	Direct digital synthesizer setting

---

[1] E.M. Mattison et al., *A Time Transfer Technique using a Space-borne Hydrogen Maser and Laser Pulse Timing* this publication.

[2] We grqtefully acknowledge helpful discussions on active magnetic compensation systems with Mr. H.E. Peters of the Sigma Tau Corp.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Developments of Rubidium Frequency Standards at Neuchâtel Observatory

Pascal Rochat, Hartmut Schweda, Gaetano Mileti, and Giovanni Busca

Observatoire Cantonal de Neuchâtel, 58 Rue de l'Observatoire
CH-2000 Neuchâtel, Switzerland

Abstract *The paper describes developments of novel miniaturized rubidium frequency standards Two variants stem from the same principal design based upon a magnetron cavity.*
1) a very small industrial unit for standard commercial applications. The overall volume of this model is only 0.25l with a mass of 470 g and power consumption of 8 W. This unit is close to industrialization,
2) a rugged spaceborne oscillator designed for operation on various spacecrafts or missions. This model was developed in the framework of ESA's In-Orbit Technology Demonstration Program. Two units, outcome from the present development, will constitute the main time reference for the Russian RADIOASTRON I space VLBI satellite. For this particular application the short time stability and temperature sensitivity are very important.The measured short time stability is $\sigma_y = 4 \cdot 10^{-12} \tau^{-1/2}$ *with* $1.6 \cdot 10^{-10}$ *temperature sensitivity over the whole operating range -20°C to +65°C.*

Introduction

Worldwide rubidium frequency standards are the by far most widely used atomic oscillators. They have a an ever increasing range of applications, i.e. in TV networks, digital data network synchronizations, PCM integrated telephone networks, navigation systems, space communication like tracking of satellites and rockets, etc.

There is a general trend to further miniaturize these devices, minimize their power requirements and reduce manufacturing costs, i.e. sales prizes.

Their compactness, portability and continuous operation capability allows them compete with quartz crystal oscillators whereby better reproducability , higher long- and medium-term stability, short warm-up time and low g-sensitivity makes them superior to them in most applications.

Since 1988 the Observatory of Neuchâtel has performed studies and developed breadboards of physics packages for miniaturized rubidium frequency standards for ground and space applications.The principle of the physics package based upon a magnetron type resonator was already described at EFTF 1992. [1]

In February 1992, we started to redesign this industrial physics package in order to reduce significantly the material and production costs. At the same time the development of an associated miniaturized industrial-type electronics was started at the Observatory. This development has recently reached maturity and an industrialization of this unit is planned in the near future.

The development of a space compatible rubidium standard started in October 1992 in the framework of the "In-Orbit Technolgy Demonstration Programme" of the European Space Agency (ESA). Two units, outcome from the present development, will constitute the main time reference for the Russian RADIOASTRON I space VLBI satellite and will be used for calibration purposes and as back-up local oscillator in case of failure of the main local oscillator system (which is based upon a quartz oscillator locked to ground H-masers via appropriate microwave links). In such a case the Rb clock will be used for observation of strong radio sources.

This paper presents the development activities in the field of lamp-pumped Rubidium Ultra-Stable Oscillators (RUSO) during last two years at the Observatory of Neuchâtel starting with a description of the industrial version. Aspects of diode laser pumped rubidium standards also being studied have recently been published elsewhere. [2]

0-7803-1945-1/94 $3.00 © 1994 IEEE

2. Industrial Miniaturized Rubidium Oscillator

This unit was developed as a general purpose oscillator for industrial application, i.e. navigation, telecommunication, network synchronization etc.

The principal characteristics of this unit are its very small volume (0.25 l), small mass (0.47 kg), low power consumption and a built-in data interface usable for monitoring and control.

The basic features and typical operation performances of this rubidium standard are

- $<1 \cdot 10^{-11} \cdot \tau^{-1/2}$ short term stability
- 3 to $4 \cdot 10^{-10}$ over - 20° to 65°C op. temp. range
- 10 MHz output, adjustable by steps of 2×10^{-10} with internal synthesizer
- fine adjustment by analog or RS232 command in steps of $1 \cdot 10^{-11}$
- <8W@25°C steady state power consumption
- 15W (25W optional) consumption during warm-up
- < 5 min. warm-time to reach a frequency stability of< 5×10^{-10}
- working parameters stored in EEPROM memory

Fig. 1 is a photograph of the unit half opened displaying the electronic boards surrounding the physics package tube. The outside dimensions are 50x57x90mm.

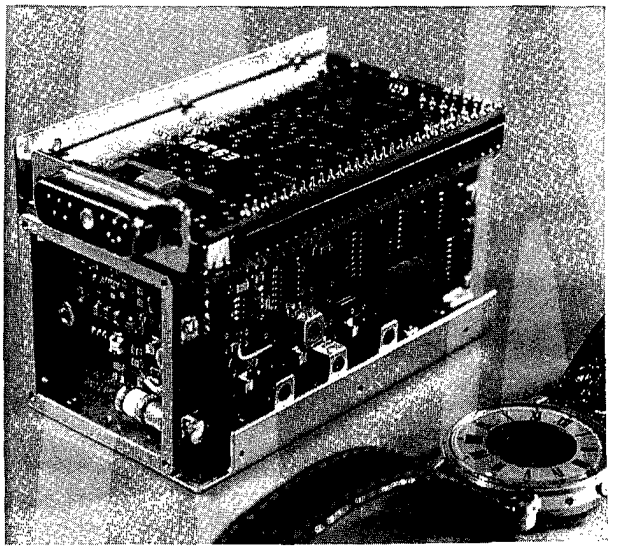


Fig.1 Industrial Rubidium Oscillator

2.1 Physics Package

The key elements of the optical package, the Rb discharge lamp assembly and absorption cell with the microwave cavity are arranged at either end of a small cylindrical tube

Other design features which contribute to the compact design are:

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

The integrated filter technique, which combines the optical filtering and pumping in one cell, contributes also to the reliability since the configuration is simplified and the number of components reduced. The thermal capacitance of the cell assembly is relatively low. As a consequence the necessary power during warm-up is greatly reduced.

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 field at the right region of the cell.

The physics package tube is nested into two magnetic shields whereby the outer shield serves as the housing of the the unit.

Two principal interface functions to the physics package, the lamp exciter RF oscillator and the step-recovery diode driver circuit, are directly mechanically attached to the physics package for stable impedance matching.

2.2 Electronics package

The electronics package is implemented on 3 different boards, which are arranged around the physics package, between the inner and outer magnetic shield:

The *interrogation board* generates the microwave signal corresponding to the hyperfine rubidium atomic transition.
Signal generation starts with a fundamental mode 20 MHz crystal multiplied by 9 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 is normalized to $2 \cdot 10^{-10}$/step at 10 MHz level.

The *controller board* contains the servo loop circuits and a microcomputer unit which is associated with multichannels D/A converters. The microcomputer also controls the interrogation and synchronous detection sequence as well as the Rb absorption "dip" search. The multichannel D/A converter is used for parameter setting of the instrument.
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 the working parameters can be stored in a built-in EEPROM memory. In addition an 8 channels A/D converter is used for monitoring of principal internal signals like: light level, signal level, thermostats voltages, RF power etc.
This feature was intensively used also for the characterization and tests of the space physics package.

The *power supply board* includes the lamp and the cell oven control circuits and the necessary sub-voltages generation.

2.3 Physics parameters optimisation

To find the optimal working points for the Rb lamp and absorption cell three main parameters have to be adjusted in an iterative manner: the lamp temperature and oscillator power, and the absorption cell temperature.

Due to the the light-shift effect which depends on the intensity *and* the spectrum of the pump light all parameters change the frequency in a non-linear way. Furthermore, the pressure shift and inhomogeneity effects taking place in the absorption cell also have to be taken into account in order to understand the observed dependence of the clock frequency on these critical parameters. [3][4]

Note, that the often observed linear light shifts are obtained in laboratory arrangements in which the pump light intensitiy is varied by means of neutral density filters. This is not the situation in assembled commercial units where light changes are usually caused by variations of operational parameters which change both the intensity and the spectral profile of the light.

Figure 2 is a sample plot of the frequency shift due to lamp oscillator current and cell temperature. For a given cell temperature, the behaviour corresponds to the lamp characteristic. The parabolic behaviour is a combined effect of light-shift connected to the light absorption and N_2 buffer gas pressure shift.

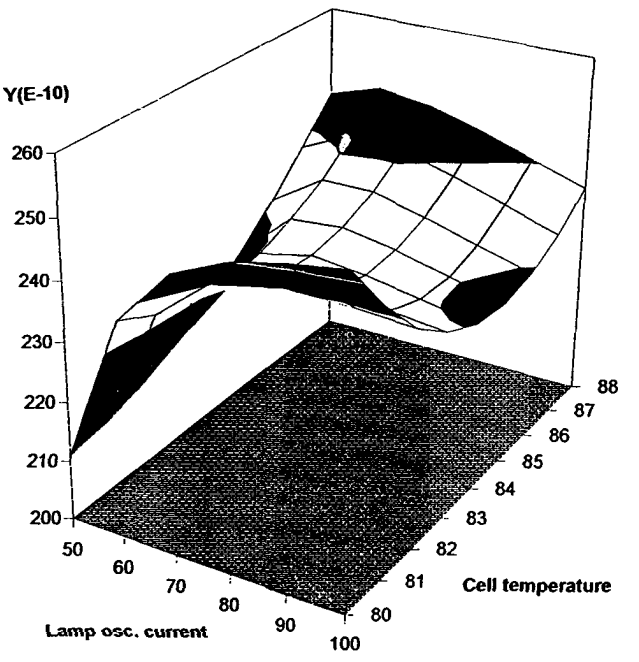


Fig. 2 Frequency vs. lamp oscillator current and absorption cell oven temperature

The optimal operating point in this particular case corresponds to about 70 mA lamp oscillator current and 83.5°C oven cell temperature. In Fig. 3 the corresponding frequency shifts due to lamp oscillator current and lamp temperature changes are displayed.

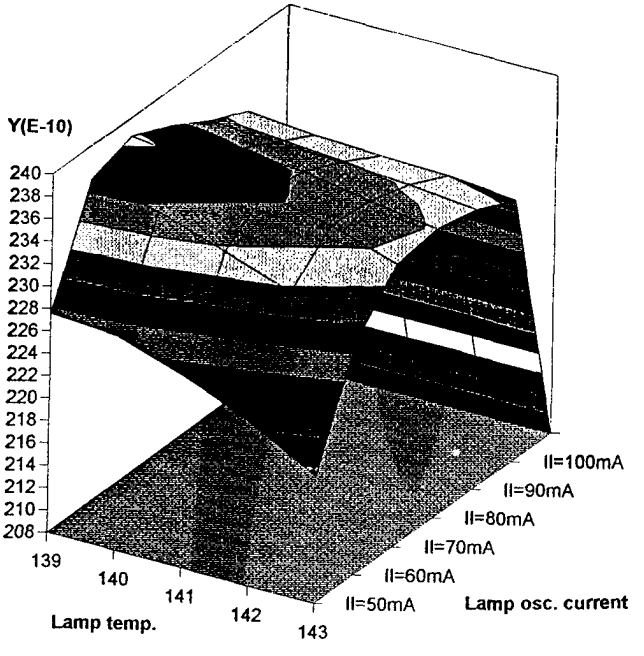


Fig.3 Frequency vs. lamp oscillator current and lamp oven temperature

This physical parameters optimisation can be made in an automatic manner by using the previously described electronics package. This provides a very good flexibility and a cost effective solution for the adjustment of the critical physics parameters.

2.4 Typical test results

Intensive tests are presently performed on 10 prototype units, starting with long-term stability measurements in order to define the specifications of the final product. Fig. 4 shows the units on the test bench.

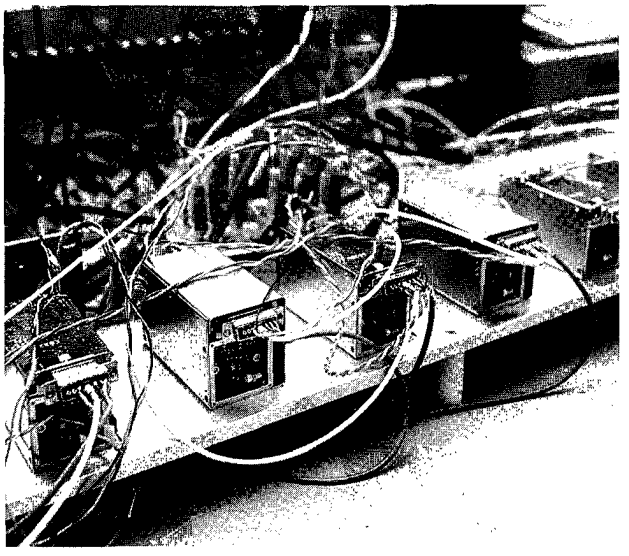


Fig.4 Rubidium oscillators on long-term test bench

The general long-term trend is a positive frequency drift, decreasing versus time ,associated with an increase of the light level transmitted through the absorption cell.
For a continuous operation, all the tested units reached the drift specification goal of 4×10^{-11}/month after at least 150 days of operation. The measured values of most units is significantly better. Figure 6 represents the long-term frequency behaviour under normal lababoratory conditions of unit No.8.

The future plan is to accelerate the preaging procedure by pre-baking the critical elements (lamp and cell) in order to shorten the burn-in period.

Fig. 5 shows typical results obtained for the phase noise at 10 MHz [L(f)], short-term stability and temperature sensitivity.

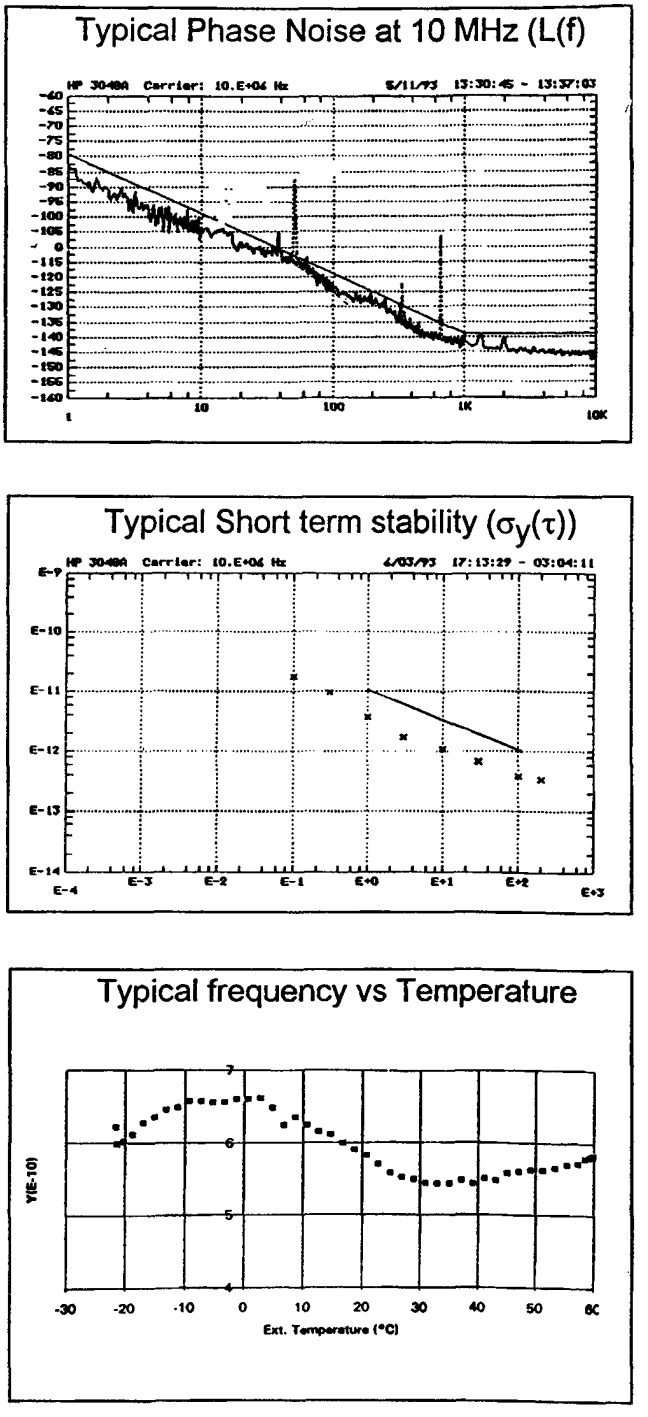


Fig. 5 Typical Test Results for Miniaturized Industrial Rubidium Oscillator

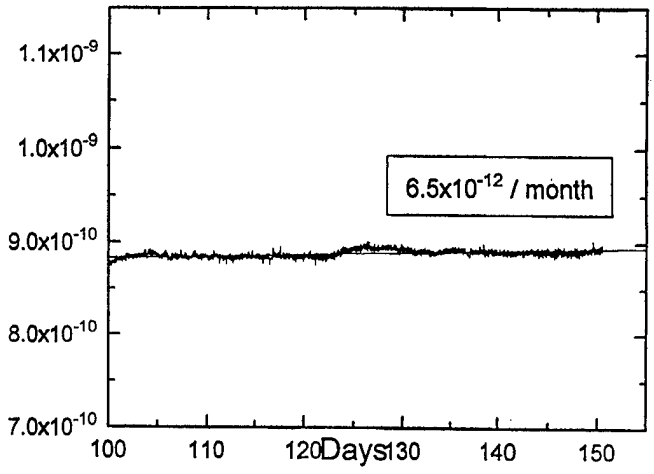


Fig.6 Long-term frequency drift measurement of miniaturized industrial rubidium oscillator

3. Spaceborne Rubidium Oscillator

The spaceborne design had to regard the more stringent levels of vibration, static acceleration, thermal vaccuum, radiation and eletromagnetic compatibilities in the spacecraft environment.

Since the unit is supposed to be capable to operate on various spacecrafts or missions it had to be equipped with an interface board dedicated to a selected spacecraft. The present development contains an interface which complies with the special requirements of the Russian RADIOASTRON 1 spacecraft.

A tested pre-engineering model of this spaceborne oscillator, named RUSO-S, has been delivered to Astro Space Center in Moscow in February 1994. An engineering model and several physics packages have been successively assembled and are presently under test.

3.1 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.

The basic configuration follows closely that of the miniaturized industrial unit. 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 temper-atures. 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.

Fig. 7 show the physics package tube before and after assembly. Connection between the lamp and cell units to the titanium tube is by electron beam welding to form a rigid structure. The attached copper tube for pumping and filling is pinched off in the final sealing process.

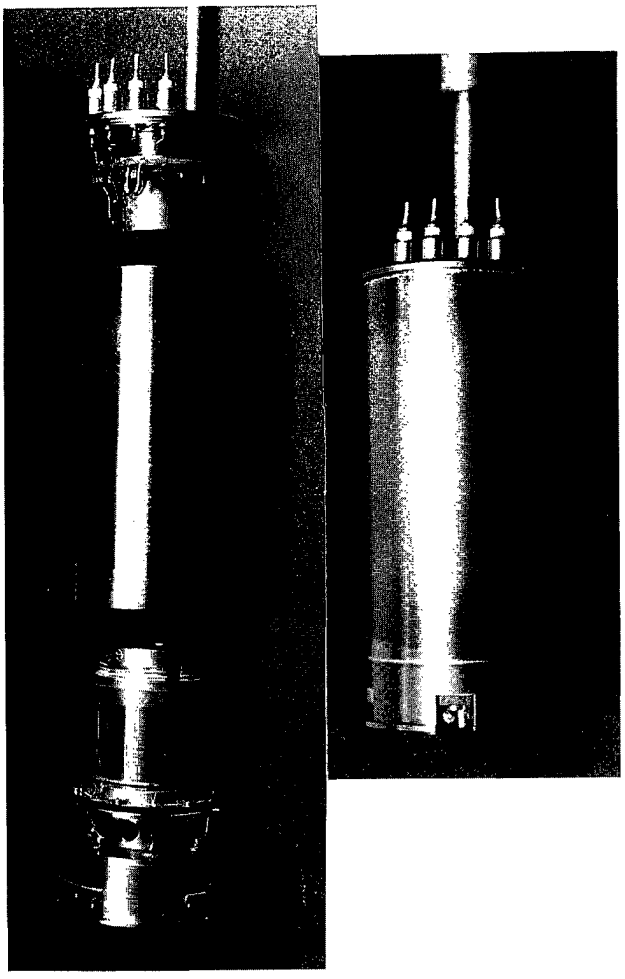


Fig. 7 RUSO-S physics package before and after integration

3.2 Space Electronics:

The space compatible design had to regard the radiation sensitivity of the components, operation capability in vacuum, insulation requirements, reliabilty considerations and interface requirements. The main differences to the ground electronics are :

- the use of high REL components
- the use of a field-programmable gate array for the control electronics synthesizer section
- full galvanic insulation of the power supply
- dedicated spacecraft interface for Radioastron.1

The complete space compatible electronics package including DC/DC converters.is integrated onto a rigid-flex-rigid PCB to avoid wiring. The board is folded over three levels within the housing box. The dedicated interface board with the connectror bracket sits on top of the arrangement. The overall configuration of the physics package and the arrangement of electronics boards is seen in Fig. 8. Fig.9 is a photograph of the complete RUSO and Fig. 10 shows the open housing displaying the main and interface board.

A block diagram of the spaceborne RUSO is given in Fig. 10

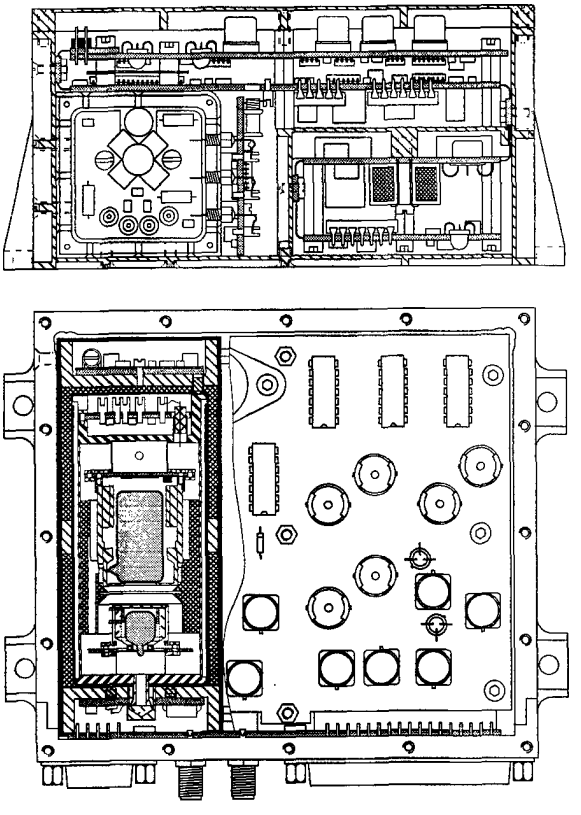


Fig. 8 RUSO-S overall configuration

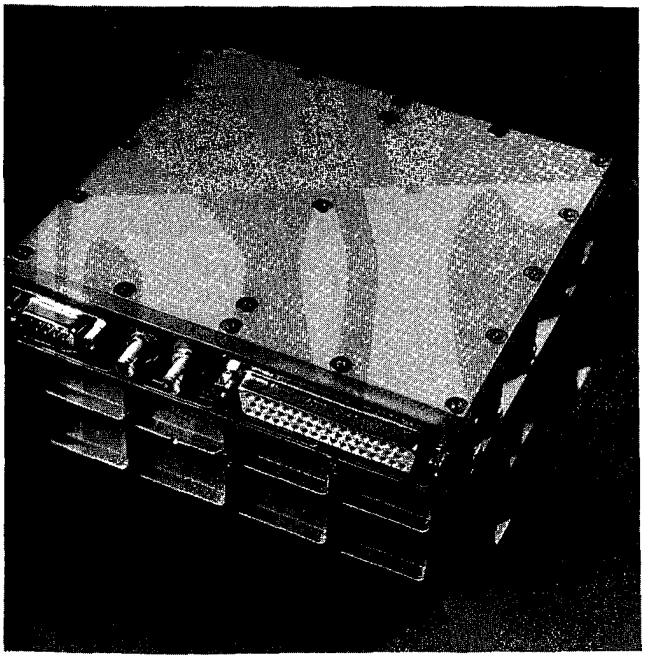


Fig.8 Fully assembled RUSO-S

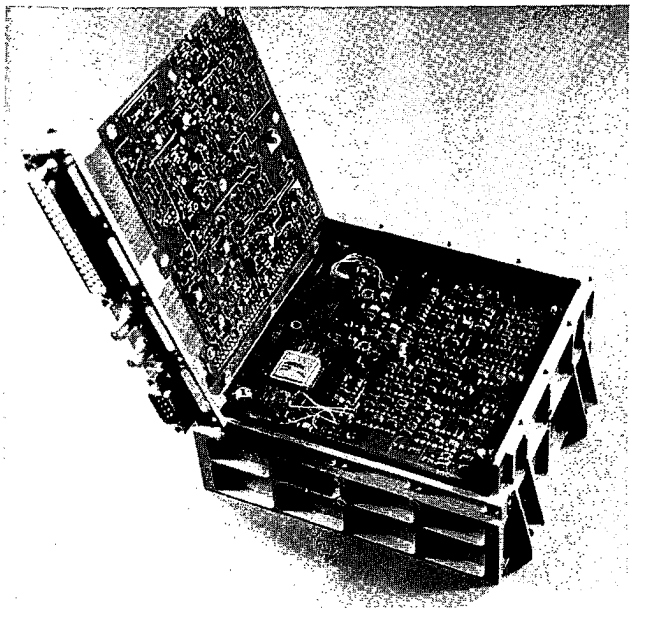


Fig. 9 RUSO-S showing main and interface board

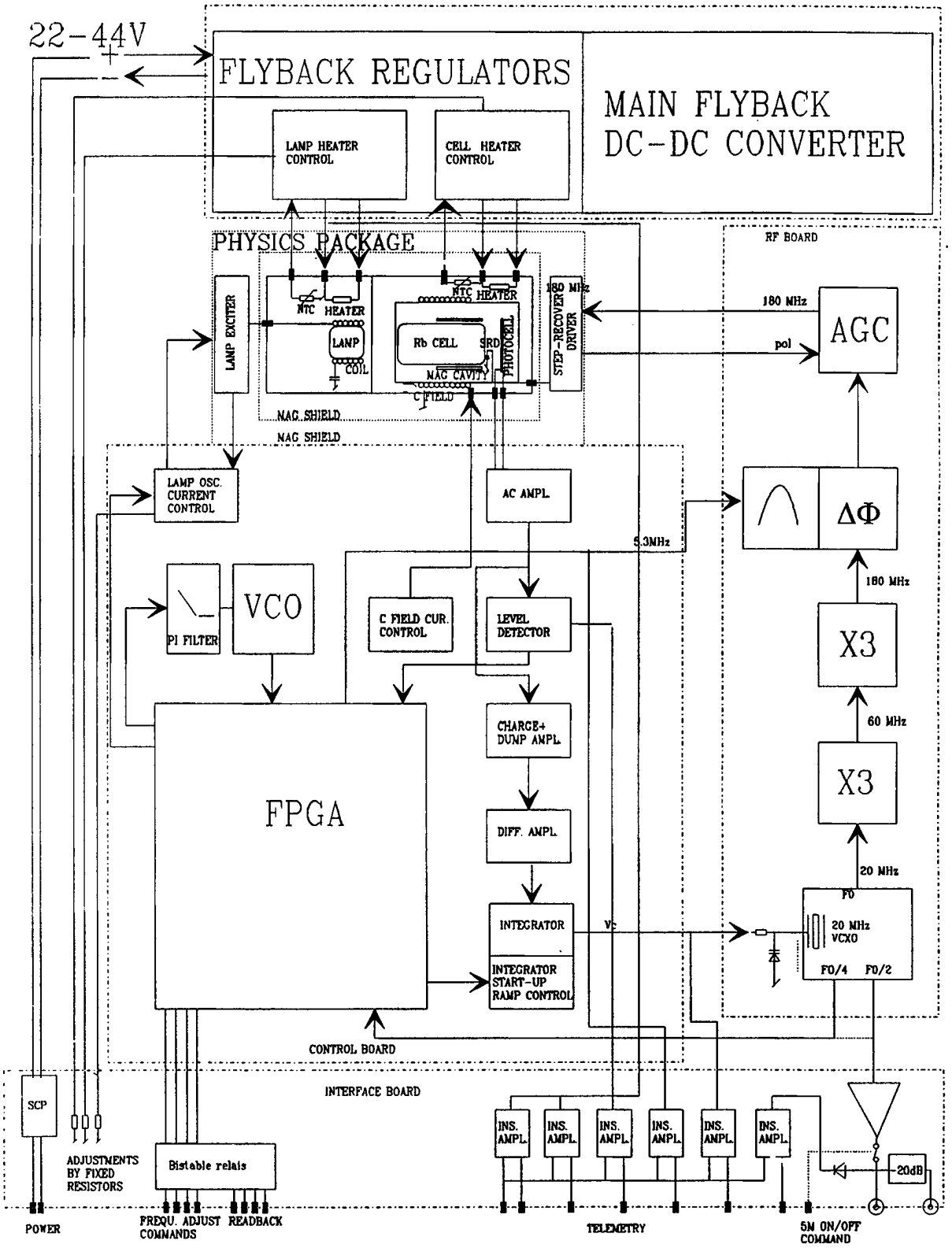


Fig. 10 Block diagram of spaceborn rubidium standard

3.3 Test results

The signal parameters measured on the RUSO-S engineering model are the following :

Absorption dip amplitude:	0.8mA
Photo cell current:	144μA
Shot noise:	8 pA(Hz)$^{-1/2}$
Total noise:	16 pA (Hz)$^{-1/2}$
Line quality factor:	3.6·10-6
Discriminator slope	0.84 nA/Hz
Predicted stability (with total noise and 50% dead time):	2.8·10-12·$\tau^{-1/2}$
Measured stability (pre-engineering model)	4·10-12·$\tau^{-1/2}$

The following table is a condensation of the specifications and test results measured on the PEM physics package associated with an industrial electronics package.

PARAMETER	ESA spec.	TEST results
Long term frequency stability (Df/f_o) per month	< 4e-11/m	<1e-11/m
Short term stability $\sigma(\tau^{-1/2})$ for $1<\tau<100$s	<1e-11 $\tau^{-1/2}$	4e-12 $\tau^{-1/2}$
Temperature stability (Df/f_o) in - 25 to + 60°C range	< 3e-10	1.6e-10
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_o = 10 MHz) in dB		
f_m= 1 Hz	-80	-84
= 10 Hz	-100	-104
= 100 Hz	-120	-124
= 1.000 Hz	-140	-142
= 10.000 Hz	-140	-145
Warm-up time to reach stability < $5x10^{-10}$	<6 min	<3e-10 / 6min
Retrace (Df/f)	<2e-11	<1.2e-11
Power consumption during warm-up	< 15W	14.5W
Power consumption during normal operation	< 8W	7.5W
Volume	<1liter	1.2 liter
Mass	<1kg	1.3kg
DC power supply voltage range	22 to 43V	16 to 45V

Acknowledgements:

This work was supported in part by ESA under The Agency's In-Orbit Technology Demonstration Program (TDP-II). We would like to thank ESTEC experts for their efficient and continuing support during this on-going program as well as our colleges P. Thomann, M. Frosio, C. Couplet, M. Dürrenberger, P. Scherler and J.-C. Sapin.

References:

[1] G. Mileti, I. Ruedi and H. Schweda, Proc. 6th EFTF, 515, (1992)

[2] G. Mileti and P.Thomann Proc. 8th EFTF (1994) under print

[3] P. Willemin, G. Busca, J.T. Liu and P. Thomann, Proc. 5th EFTF, 18, (1991)

[4] J. Rabian, P. Rochat, Proc. 2th EFTF, 461, (1988)

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THE DOUBLE BULB RUBIDIUM MASER

W.M. Golding and A. Frank
SFA Inc. Landover, MD 20785
R. Beard, J. White, F. Danzy and E. Powers
Naval Research Laboratory Washington, D.C. 20375

Abstract

The proposed design of a new ^{87}Rb maser frequency standard is presented. This device is expected to perform similarly to a hydrogen maser at short and moderate averaging times. An operating maser of this design would much be more compact than a full size hydrogen maser (the cavity need only be $\sim$ 6 cm in length and diameter). It should also be much less expensive to build. Optical pumping techniques are used to produce a nearly complete population inversion in an evacuated wall coated cell. The population inversion is produced in a cell separated from the microwave interaction region to eliminate problems with light shifts. Since no buffer gas is used the atoms are free to travel through an exchange tube from the optical pumping region to the maser interaction region where oscillation can take place. The system will be completely closed-cycle since atoms will return from the maser region to the optical pumping region where they will be repumped. This eliminates the need for a vacuum system, greatly simplifying operation. The principles of operation and the basis of the stability estimates will be discussed.

Introduction

In this paper, we describe the design of a new optically pumped, active ^{87}Rb maser which uses a double storage bulb technique to separate the optical pumping region from the region of maser oscillation. Relatively large frequency shifts known as light shifts occur if an atom is interrogated while an optical pumping source is on [1,2,3,4]. If a diode laser is used as the pumping source, these shifts force one to deal with laser stabilization techniques which do not in principle eliminate the problem. Techniques which separate the region of interrogation from the region of optical pumping are useful in that they, in principle, eliminate the effect of the light shift altogether. The double bulb design of this maser accomplishes this separation and allows the rubidium atoms to evolve in the dark, thereby, eliminating the problem of light shifts. This Rb maser design is substantially different in operation than the standard Rb maser [5,6,7] which uses a buffer gas to reduce Doppler broadening.

The development of this new maser is made possible by the use of a tetracontane coated, Pyrex cell that can be used to efficiently store rubidium atoms without the use of a buffer gas [8,9,10]. Using such a cell, it is possible to store rubidium atoms in a way that is similar to the standard technique used to store hydrogen atoms in the Teflon coated storage bulb of a hydrogen maser. Robinson has called such a cell an evacuated wall-coated cell (EWCC). Using standard optical pumping techniques in an EWCC of 25 cm^3, the wall relaxation contribution to the line width was measured as 8 Hz [9,11,12]. This produced an atomic line Q of 0.85×10^9 at 6.835 GHz.

We have designed a system which will use a pair of such EWCC's connected by a small transfer tube in a double storage bulb arrangement. By using optical pumping techniques to produce a population inversion

U.S. Government work not protected by U.S. copyright

in one cell and allowing the population inversion to be transferred to a dark cell located in a tuned microwave cavity, it is possible to build an oscillating rubidium maser with excellent medium term stability ($\sigma_y(\tau) \approx 2 \times 10^{-14}/\sqrt{\tau}$). This new design is free from many of the systematic frequency shifts that plague other rubidium standards such as the light shift, the buffer gas frequency shift, and the RF power shift. The frequency shifts which are expected to significantly affect the performance of this active rubidium maser design are; the wall shift, the cavity pulling effect and the spin exchange frequency shift. These have all been dealt with adequately in well-understood hydrogen maser designs, and we have therefore based the interaction region of this rubidium maser design on the hydrogen maser [13,14,15].

The linear dimensions of the fully developed oscillating maser physics package will be at least 5 times smaller than a standard size hydrogen maser due to the higher hyperfine frequency of ^{87}Rb. The cylindrical microwave cavity which will enclose the rubidium storage bulb will be 5.8 cm long and have a 5.8 cm diameter. Relative to a standard size hydrogen maser, there will be a substantial savings in size and complexity, because the state selecting magnets, the dissociator and the vacuum system necessary in the hydrogen maser design are not necessary in this maser.

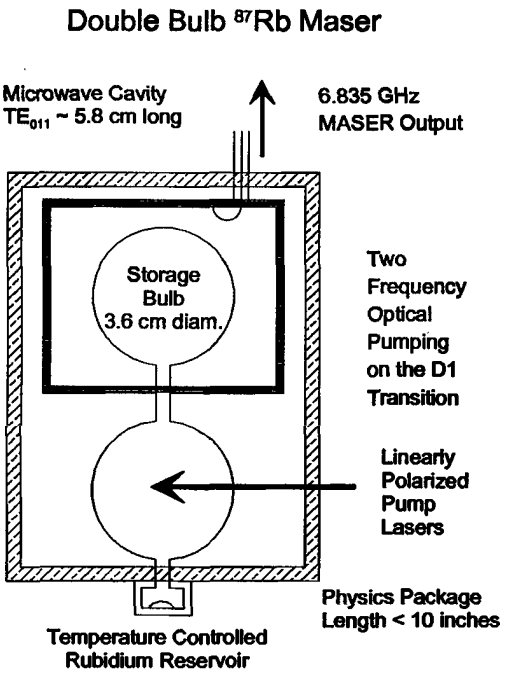


Figure 1: The basic design of the double bulb maser.

Description of the Design

The design of the proposed maser is shown in Figure 1. A spherical storage bulb resides within a standard cylindrical microwave cavity operated in the TE_{011} mode. It is connected by a thin exchange tube to a pumping bulb which lies outside the cavity volume. The inner surfaces of both bulbs and the exchange tube are coated with n-tetracontane. An axially directed magnetic field (the C-field) is applied to the whole structure. A reservoir of ^{87}Rb is attached by a narrow opening to the pumping bulb. Only rubidium vapor is present in the double bulb. No buffer gases are used in this design. The operating density of the rubidium vapor is fixed by the vapor pressure at the operating temperature. Within the pumping bulb, the vapor is optically pumped into the single upper hyperfine state $|F = 2, m_F = 0\rangle$ at a high rate [12]. These atoms then effuse through the exchange tube into the storage bulb where the maser oscillation occurs. The spent atoms return to the pumping bulb where they are again pumped to the upper hyperfine state, completing the cycle. This cycle is illustrated in Figure 2.

The designs of the storage bulb and cavity are obtained by scaling a standard hydrogen maser design to an operating frequency of 6.835 GHz. The maximum unloaded cavity Q obtainable in this cavity design is about 37000 assuming a high quality silvered surface. We assume an unloaded Q of 25000 as a conservative estimate for our calculations. The diameter and length of the full size cavity are both 5.8 cm. The radius of the

storage bulb should be 1.74 cm giving a bulb volume of 22.1 cm^3 and a filling factor, η' of 0.39. We have not yet considered the loading effect of the storage bulb on cavity frequency and Q. These will be considered in the final cavity and bulb design.

The actual shape of the pumping bulb will be designed to optimize the effectiveness of the laser diodes used for optical pumping. Both mounting considerations and interaction volumes need to be considered. The length and diameter of the exchange tube will be used to set the geometric storage time of the atoms in the storage bulb. The appropriate magnetic shielding, C-field coils and ovens for thermal control have not been included in Figure 1.

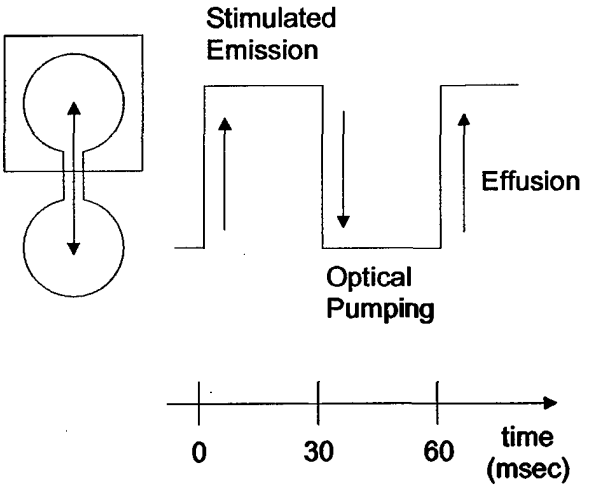


Figure 2: This figure represents schematically the path of a typical atom in the optical pumping → effusion → maser oscillation → effusion cycle.

A systematic study of the optical pumping process in a single bulb tetracontane coated ^{87}Rb vapor cell has been performed [12]. That analysis applies directly to the pumping bulb of the design presented here. The single difference being the presence of the exchange tube and the associated escape and return of atoms. As shown by [12], it should be possible to efficiently pump a large percentage of the Rb atoms in the pumping bulb into the single state $|F = 2, m_F = 0\rangle$ using readily available laser diodes. The exchange tube time constant is assumed long relative to the time scale of the optical pumping processes. Under these conditions, a large population inversion can be maintained in the pump bulb providing a source of atoms to maintain the maser oscillation in the storage bulb.

Stability Estimates

As a reasonable estimate of the performance of a completed Rb maser design, we calculate the mid-term ($\tau \approx 10$ to 10000 seconds) frequency stability of the oscillating maser. This stability analysis follows the analysis of hydrogen maser performance given in [13]. The stability for this range of averaging times is determined by three parameters: cavity temperature, operating line Q, and oscillation power, P. The theoretical mid-term stability of the oscillation is then given by,

$$\sigma_y(\tau) = \sqrt{\frac{kT}{2P\tau}}\frac{1}{Q}.$$

An important parameter for maser performance is the oscillation threshold parameter α. It is defined by

$$\alpha \equiv KQ_cT_1T_2I,$$

where we have $K \equiv \frac{\mu_0\mu_B^2\eta'}{\hbar V_b}$, Q_c is the loaded cavity Quality factor, T_1 and T_2 are respectively the longitudinal and transverse time constants for stored atoms, and I is the flux of the population difference entering the storage bulb. Also μ_0 is the permeability of free space, μ_B is the Bohr magneton, $\eta'$is the cavity filling factor, $\hbar$ is Planck's constant and V_b is the volume of the storage bulb. For a particular maser design to oscillate, we must have

$$\alpha \geq 1.$$

It is expected that atoms will be pumped at a high rate into the

$$|F = 2, m_F = 0\rangle$$

state and therefore most of the atoms effusing from the pumping bulb to the storage bulb will be in this state. The density, n, of Rb atoms in the storage bulb is therefore given by $T_b I$, where T_b is the geometric storage time of the storage bulb. Knowing the density and the bulb time constant, we can determine the effective atomic flux. For a storage bulb connected to a thin exchange tube,

$$T_b = \frac{4V_b}{\bar{v}A_a}\left(1+\frac{3L}{8r}\right),$$

where L and r are the length and radius of the exchange tube, A_a is the cross sectional area of the tube and $\bar{v}$ is the mean velocity of the stored atoms [14].

There are three major relaxation mechanisms which will contribute to $\gamma_1 = 1/T_1$ and $\gamma_2 = 1/T_2$. They are wall relaxation, spin exchange relaxation $\gamma_e = n\sigma\bar{v}_r$, and the geometric escape rate $\gamma_b = 1/T_b$. The total density of rubidium atoms in the storage bulb is n. The spin exchange cross section, σ, for rubidium is about 2.5×10^{-14} cm^2 and the mean relative velocity is

$$\bar{v}_r = 4\left(\frac{kT}{\pi m}\right)^{1/2},$$

where m is the rubidium atomic mass. We do not presently have direct measurements of γ_{1w}. We assume that it is approximately the same as the measured value for γ_{2w} of about 30 sec^{-1} although, relaxation rates measured for Paraflint indicate that γ_{1w} is likely to be at least 5 times smaller than γ_{2w} [8]. For a conservative estimate, we set $\gamma_{1w} = \gamma_{2w} = \gamma_w$ in what follows. The total relaxation rates are then,

$$\gamma_1 = \gamma_b + \gamma_w + \gamma_e$$
$$\gamma_2 = \gamma_b + \gamma_w + \tfrac{5}{8}\gamma_e$$

and the oscillation power level is given by

$$P = \frac{1}{2}\hbar\omega_0 I\left(1-\frac{1}{\alpha}\right).$$

A near optimal choice of operating density and escape rate is obtained if all the relaxation mechanisms

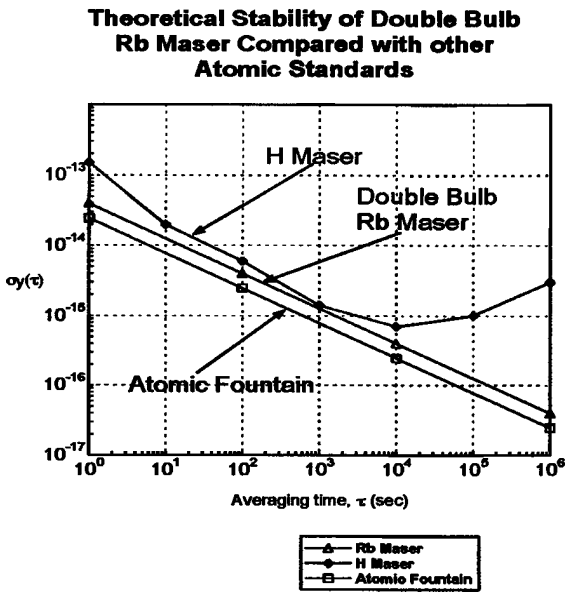


Figure 3: This plot compares the theoretical stability of the ^{87}Rb maser to the stability of an actual H maser. The theoretical stability of a proposed atomic fountain is included as an example of one of the new breed of passive atomic standard that may require H maser type performance from a slaved local oscillator.

contribute about equally to the total relaxation. This gives good mid-term stability and does not degrade the cavity pulling factor too severely. For

$$\gamma_e = \gamma_b = \gamma_w = 31 \text{ sec}^{-1}$$

at anoperating temperature of $35°C$, we calculate that

$$\sigma_y(\tau) = 4.1 \times 10^{-14}/\sqrt{\tau}.$$

By operating at $50°C$, a somewhat better mid-term performance can be obtained at the expense of a reduced line Q producing a larger cavity pulling factor that would degrade the long-term performance.

We find that by using $\gamma_e = \gamma_b = 128$ sec^{-1} and $\gamma_w = 31$ sec^{-1} we can obtain

$$\sigma_y(\tau) = 2.1 \times 10^{-14}/\sqrt{\tau}.$$

In the Table 1, we summarize the results of the system calculations at two different operating temperatures. In this device, the operating temperature has the effect of setting the Rb density through the vapor pressure of the Rb reservoir. One sees that if systematic fluctuations can be stabilized for timescales of the order of one hundred seconds, that this device would be capable of achieving a stability of

$$\sigma_y(\tau = 100 \text{ secs}) = 2 \times 10^{-15}.$$

A plot of the theoretical Allan variance is shown in Figure 3 along with the actual stability achieved by a hydrogen maser.

Wall Shift

It must be realized that the previous stability estimate assumes that the systematic frequency shifts are well under control for the time scale of interest. For averaging times less than about 10000 seconds, this is a reasonably good assumption for frequency shifts like the spin exchange shift, cavity pulling effect and second order Doppler for example. However, the wall shift appears to be rather large and also to have a large temperature sensitivity. We have compiled data from various sources on the wall shift for ^{87}Rb on n-tetracontane [16,9,11] and ^{87}Rb on Paraflint [17, 18]. The original frequency shift data were all taken in different size bulbs. The bulbs were prepared by different investigators by similar techniques of vacuum evaporation. Paraflint is described as a mixture of hydrocarbon chains or paraffins with the number of carbon atoms per molecule ranging from 40 to 80. n-tetracontane, $(CH_3(CH_2)_{38}CH_3)$, is a pure linear hydrocarbon chain with 40 carbon atoms. We expect there to be some similarity between the frequency shifts and wall relaxation rates measured on all of these surfaces. As far as we know, the measured differences in γ_2 type relaxation rates have not yet been adequately explained.

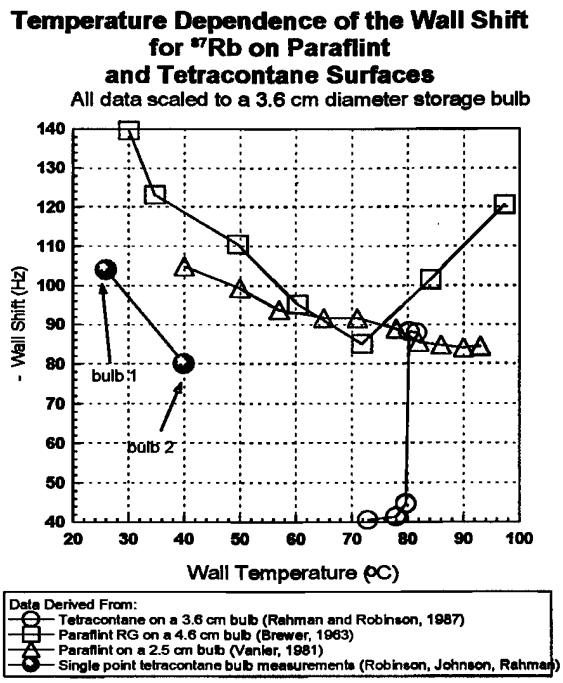


Figure 4: This figure presents ^{87}Rb wall shift data for n-tetracontane and for Paraflint as measured by different investigators over the last 30 years.

We have converted all measured data by scaling on the diameter to an equivalent 3.6 cm diameter bulb (nominally the maser storage bulb design size) for comparison. This data is plotted in Figure 4. From this data, we expect that the temperature coefficient of the wall shift is a few parts in 10^{13} per millikelvin. This is very large by hydrogen maser standards and seems to call for extreme temperature control or operation at a temperature extrema. In the n-tetracontane data it does appear that a temperature minima exists; however, it is extremely near to the melting point of tetracontane (81 to 83°C). The sharp melting point of tetracontane is related to the fact that it is a pure substance. Paraflint melts or softens [17] over a much

^{87}Rb Maser Performance Parameters

Rb temperature	$35°C$	$50°C$
Oscillation threshold α	1.57	2.89
Total T_2 (msec)	12.2	4.2
Output power (Watts)	1.5×10^{-11}	5.4×10^{-10}
Rb density cm^{-3}	3.2×10^{10}	1.2×10^{11}
Cavity pulling factor	1.7×10^{-14} per Hz	4.9×10^{-14} per Hz
Bulb relaxation rate	$31 sec^{-1}$	$128 sec^{-1}$
Wall relaxation	$31 sec^{-1}$	$31 sec^{-1}$
Spin exchange relaxation	$31 sec^{-1}$	$128 sec^{-1}$
Allan variance	$4.1 \times 10^{-14}\sqrt{\tau}$	$2.2 \times 10^{-14}\sqrt{\tau}$

Table 1: Basic performance analysis of the double bulb rubidium maser at two possible operating temperatures.

broader range of temperatures. Ideally, one would like to combine the broad temperature minimum of Paraflint combined with the excellent linewidths obtained using tetracontane. This suggests that controlled mixtures of pure paraffins be tested as wall coatings to try and obtain good relaxation properties and a broad temperature minimum in the frequency shift. If a useful temperature minimum can be found, then the ^{87}Rb maser could be designed to approach the theoretical frequency stability shown in Figure 3.

Conclusion

The basic design of our proposed ^{87}Rb maser was described. It was pointed out that the temperature sensitivity of the wall shift is expected to be a difficult problem to solve. However, if it can be solved the ^{87}Rb maser would be capable of approaching its theoretical stability out to 10000 seconds. The theoretical stability was shown to be about what one would expect for an active hydrogen maser yet the ^{87}Rb maser would be significantly smaller, lighter and cheaper to build. Thus, we expect that it would be a practical alternative to a hydrogen maser in situations where the cost and weight of a hydrogen maser are prohibitive.

References

[1] Camparo, J. C. and Frueholz R. P., "Fundamental stability limits for the diode laser-pumped rubidium atomic frequency standard," *J. Appl. Phys.*, vol. 59, pp. 3313-3317, (1986).

[2] Chantry, P. J., et al., "Towards a miniature Laser-Pumped Cesium Cell Frequency Standard," *Proceedings of the 1992 IEEE Frequency Control Symposium*, 114-122, (1992).

[3] Drullinger, R. E., et al., "Diode-Laser-Pumped, Gas-Cell Atomic Clocks," 1992 *IEEE Frequency Control Symposium*, 104- 107, (1992).

[4] Lowe, J. P., et al., "Ultra-High Stability Synthesizer for Diode Laser Pumped Rubidium," *1992 IEEE Frequency Control Symposium*, 183-187, (1992).

[5] Davidovits, P., Novick R., "The Optically Pumped Rubidium Maser," *Proc. IEEE* 54, No. 2, 155-170, (1966).

[6] Vanier, J., "Relaxation in Rubidium-87 and the Rubidium Maser," Phys. Rev. 168, No. 1, 129-149, (1968).

[7] Tetu, M., et al., "Experimental Results on a Frequency Standard Based on a Rubidium Maser," *Proc. 39th Annual Symposium on Frequency Control*, 64, (1985).

[8] Bouchiat, M. A., Brossel J., "Relaxation of Optically Pumped Rb Atoms on Paraffin-Coated Walls," *Phys. Rev.*, 147 No. 1, 41-54, (1966).

[9] Robinson, H. G., Johnson, C. E., "Narrow Rb87 hyperfine structures in an evacuated wall-coated cell," *Appl. Phys. Lett.* 40(9), 771 (1982).

[10] Robinson, H. G., "Methods to Recover the Narrow Dicke Sub- Doppler Feature in Evacuated Wall-Coated Cells wihout restriction on Cell Size," *Proc. 16th Annual PTTI*, 81-89, (1984).

[11] Newton, C. L. J., Robinson, H. G., "Study of Multiple Optical Transitions in 87Rb Using Laser Diodes," *Proc. 14th PTTI*, 133 (1982).

[12] Lee W., Robinson, H. G., Johnson, C. E., "Diode Laser 87Rb Pumping in an evacuated Wall-Coated Cell," *Proc. 16th PTTI*, 91 (1984).

[13] Kleppner, D., Goldenberg, H. M., Ramsey, N. F., "Theory of the Hydrogen Maser," *Phys. Rev.*, 126, No. 2, 603-615, (1962).

[14] Kleppner, D., et al., "Hydrogen Maser Principles and Techniques," *Phys. Rev.*, 138, No. 4A, A972-A983, (1965).

[15] Vanier, J., Audoin, C. *The Quantum Physics of Atomic Frequency Standards.* Bristol and Philadelphia: Adam Hilger, (1989).

[16] Rahman, C., Robinson H. G.,"Rb 0-0 Hyperfine Transition in Evacuated Wall-Coated Cell at Melting Temperature," *IEEE Journal of Quantum Electronics*, Vol. QE-23, 452-454, (1987).

[17] Brewer, R. G., "Study of Atom-Wall Collisions by Optical Pumping," *Journal of Chemical Physics*, Vol. 38, No. 12, 3015-3020, (1963).

[18] Vanier, J., et al., "Progress and Prospects in Rubidium Frequency Standards," *Journal de Physique*, Colloque C8, supplement au $n°12$, Tome 42, 139-150, (1981).

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

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

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

Anthony G. Mann, David G. Blair
Department of Physics
University of Western Australia
Nedlands WA 6009
Perth, Australia

Abstract

A microwave frequency standard based on buffer-gas cooled $^{171}Yb^+$ ions confined in a linear Paul trap has been demonstrated in prototype form. The standard exhibits a fractional Allan deviation $\sigma_y(\tau)=3.7 \times 10^{-13}\,\tau^{-1/2}$ for $\tau < 3000$ s.

Introduction

The application of trapped ions as a microwave frequency standard has been the subject of substantial recent interest [1,2,3,4,5]. The maximum Q factor of the resonance between a hyperfine transition and applied microwave radiation increases with the interaction time t_i according to

$$Q = 2t_i\nu, \qquad (1)$$

where ν is the frequency of the microwave resonance. The limiting Allan deviation, $\sigma(\tau)$, of a frequency source stabilised by reference to such a microwave resonance is proportional to

$$\sigma(\tau) \propto \frac{1}{Q}\frac{1}{SNR}\sqrt{\frac{t_i}{\tau}}, \qquad (2)$$

where SNR is the signal-to-noise ratio of the measurement. Interaction times of many seconds are achievable using trapped ions, potentially yielding much improved stability over more traditional standards.

The 12.6 GHz ground state hyperfine transition in $^{171}Yb^+$ ions (Fig.1) is attractive as a practical microwave frequency standard because the 369.5 nm ultraviolet resonance transition used for detecting the state of the ions is potentially accessible using relatively low-cost and compact frequency-doubled semiconductor lasers [6]. Furthermore, it is likely that frequency shifts due to the second order Doppler effect can be greatly reduced by laser-cooling the trapped ions.

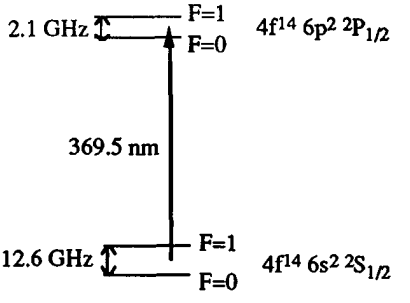


Figure 1: Partial energy level diagram of $^{171}Yb^+$ ions.

In this paper we report progress towards a 12.6 GHz microwave frequency standard based on buffer gas-cooled $^{171}Yb^+$ ions confined in a linear Paul trap.

Experimental

The linear Paul trap (Fig. 2) has been described in detail elsewhere [7,8,9].

For the experiments described in this paper the trap was operated with an RF excitation of 500 kHz at 250 Vp-p applied to the linear electrodes, and DC voltages of between 2 V and 26 V applied to the end electrodes.

0-7803-1945-1/94 $3.00 © 1994 IEEE

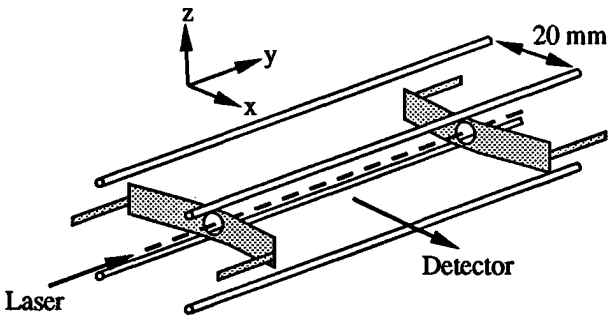


Figure 2: The linear Paul trap. The four rods are cut from 2 mm molybdenum rod, and the two end electrodes are folded from 0.25 mm OFHC copper sheet. The end electrodes are 60 mm apart.

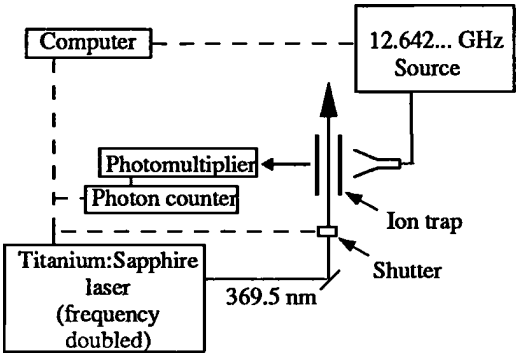


Figure 3: Simplified experimental arrangement. The Helmholtz coils and double magnetic shields surrounding the ion trap are not shown.

The experimental arrangement is shown schematically in Fig. 3.

The stainless steel vacuum chamber containing the trap is surrounded by a set of 3 pairs of square Helmholtz coils with an edge length of 300 mm for magnetic field control. Both the Helmholtz coil assembly and the vacuum chamber containing the trap are enclosed within a magnetic shield. The shield, which is approximately cubic with edge lengths of 600 mm, consists of two layers of Co-Netic magnetic shielding material. A hole in the base of the magnetic shield allows for the vacuum connection between the chamber containing the trap and the vacuum pumps. Smaller holes in the sides of the shield allow the passage of laser light to the trap, and the fluorescence from the ion cloud to reach the detector.

To improve the homogeneity of the magnetic field within the ion trap, care was taken to avoid the use of magnetic materials within the magnetic shield. In particular, electrical feedthroughs providing connections to the trap electrodes and other items are located away from the trap and outside the magnetic shield.

The primary source of inhomogeneity is thought to be from the stainless steel vacuum chamber, which is necessarily stressed by the tightening of flange bolts, and from the glass-metal seals of the chamber windows (MDC VP-400N). The current in the Helmholtz coils was adjusted to provide a magnetic field of 5.3 μT oriented approximately 20° above the y axis (Fig. 2) in the y-z plane. The magnetic shielding presently attenuates ambient laboratory magnetic field fluctuations by a somewhat disappointing factor of 20.

The ion cloud is located on the major axis of the trap (y direction) and is approximately 10 mm long and 1 mm wide. The number of ions in the trap has not yet been reliably measured.

The 369.5 nm laser beam propagates along the major axis of the ion trap (y direction) and is polarised at 45° to the z direction. To minimise scattered laser light the laser beam enters the vacuum chamber through a window tilted 45° to the beam axis, and exits through a Wood's horn. The diameter of the laser beam within the ion cloud is approximately 300 μm, and the laser power is approximately 20 μW.

Fluorescence from the trapped ions is monitored by a photomultiplier (Hammamatsu R1332) through cylindrical optics located along the x axis. Noise due to stray room light is largely eliminated by a Newport UG11 filter. The signal from the photomultiplier is monitored by a Stanford Research SR400 photon counting system.

The 12.6 GHz microwave field is radiated from a horn located within the magnetic shields, but outside the vacuum chamber. The microwave field propagates in the x direction into the vacuum chamber through a glass window and its H vector is oriented in the z direction.

The fluorescence signal from the trapped ions obtained by scanning the laser through resonance with the 369.5 nm optical transition is shown in Fig. 4. The ions were irradiated with resonant 12.6 GHz microwave radiation. Spectroscopic grade helium from a glass flask was continuously admitted to the vacuum chamber through a sapphire leak valve, and a pressure of 7.4×10^{-5} Pa was maintained by a turbo-molecular pump. The helium pressure is not yet actively stabilised, but generally remains constant within ±10%. The temperature measurements reported in this paper were made by fitting Gaussian profiles (Fig. 5) to three components of the excitation spectra obtained in the same way as the data shown in Fig. 4. The temperature measured in this way therefore corresponds to the secular motion along the major axis of the trap. At a He pressure of 7.4×10^{-5} Pa an ion temperature of 100 ± 15 °C was obtained.

The fluorescence signal is almost entirely absent without the microwave field, due to optical pumping between the ground state hyperfine levels.

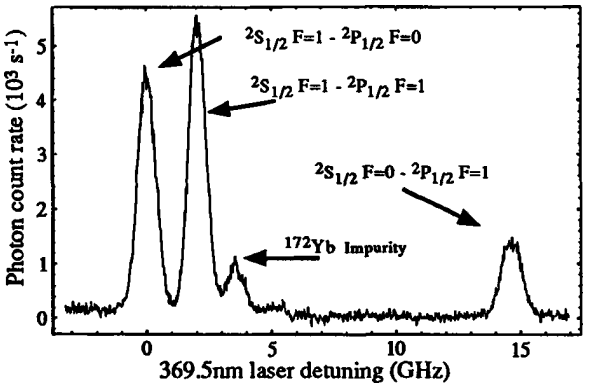


Figure 4: Excitation spectrum of isotopically enriched $^{171}Yb^+$, obtained by scanning the frequency of the 369.5 nm laser with the 12.6 GHz microwave field illuminating the ion cloud.

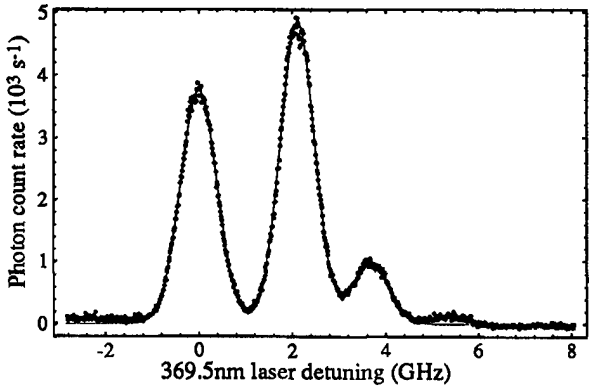


Figure 5: Partial excitation spectrum of isotopically enriched $^{171}Yb^+$ (dots), with Gaussian profiles (solid line) fitted to the F=1→0 and F=1→1 $^{171}Yb^+$ peaks, and to the $^{172}Yb^+$ impurity peak. The He pressure was 7.4 10^{-5} Pa.

Figure 6 shows the spectrum obtained by tuning the laser to the $^2S_{1/2}$ F=1 → $^2P_{1/2}$ F=1 optical transition, and scanning the frequency of the microwave field through resonance with the 12.6 GHz ground state hyperfine transition. Weak sidebands at 2.3 kHz and 12 kHz, attributed to the longitudinal and transverse secular motions respectively, are visible on the M_F=0↔0 'clock' transition, as well as on the two M_F=0↔±1 transitions.

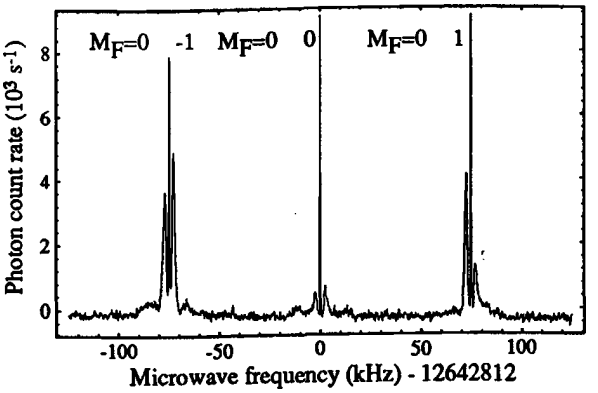


Figure 6: Zeeman sidebands at ± 74 kHz from the centre 'clock' transition.

The 12.6 GHz microwave source

The phase stability of the local oscillator over the time required for the interrogation of the clock transition is critical to the performance of an atomic frequency standard. For this reason the 12.6 GHz microwave source is based on a prototype sapphire-loaded superconducting cryogenic resonator oscillator operating close to 9.73 GHz [10]. This oscillator has a fractional Allan deviation of ≤1 x 10^{-14} between 1 and 1000 s. The microwave system (Fig. 7) was carefully engineered to minimise the addition of phase noise and drift. The 12.6 GHz signal was tuned with 1 mHz resolution using a HP3335 synthesizer interfaced to the computer controlling the experiment. Long term drift of the local oscillator was monitored by reference to a H maser, and the ion trap frequency stability data was corrected during post-processing.

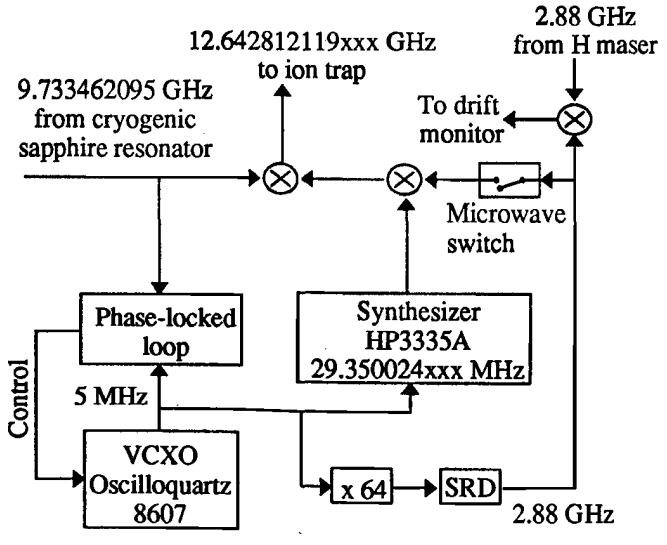


Figure 7: 12.6 GHz microwave source based on a 9.73 GHz sapphire-loaded superconducting resonator. The microwave switch is controlled by a computer to generate the Ramsey pulses. SRD = step-recovery diode.

Microwave interrogation

The 12.6 GHz ground state hyperfine resonance is coherently interrogated using Ramsey's method of separated oscillatory fields. The measurement sequence is shown in Fig. 8.

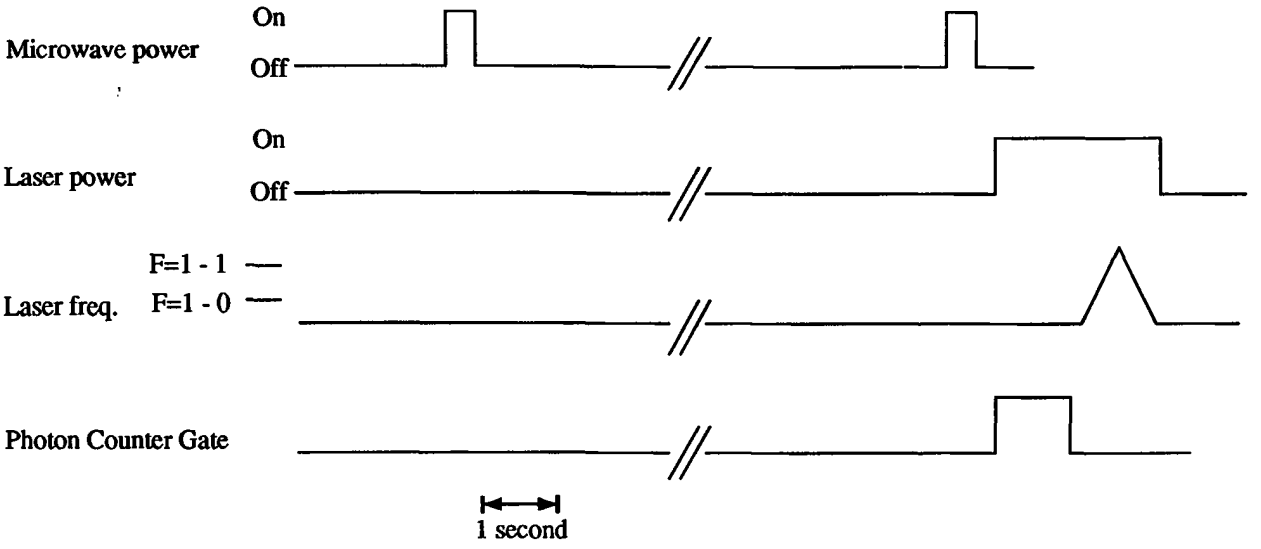


Figure 8: Sequence for microwave interrogation of the 12.6 GHz $^{171}Yb^{+}$ 'clock' transition.

The ions are initially prepared in the $^2S_{1/2}$ F=0 ground state hyperfine component by scanning the laser through resonance with the $^2S_{1/2}$ F=1 $\rightarrow$ $^2P_{1/2}$ F=1 optical transition. The laser is then tuned back to the $^2S_{1/2}$ F=1 $\rightarrow$ $^2P_{1/2}$ F=0 optical transition, and the laser shutter is closed. A pair of 400 ms microwave $\pi/2$ pulses are then applied, with the pulse centres separated by the microwave interaction time t_i. After the second $\pi/2$ microwave pulse, the laser shutter is then opened, the photon counter is enabled for 1 s, and the photon count rate, which is proportional to the population of the $^2S_{1/2}$ F=0 ground state hyperfine level, is recorded. The laser is then once again scanned through resonance with the $^2S_{1/2}$ F=1 $\rightarrow$ $^2P_{1/2}$ F=1 optical transition to reset the ions into the $^2S_{1/2}$ F=0 level in preparation for the next pulse sequence.

Using the $^2S_{1/2}$ F=1 $\rightarrow$ $^2P_{1/2}$ F=0 optical transition rather than the $^2S_{1/2}$ F=1 $\rightarrow$ $^2P_{1/2}$ F=1 transition for interrogating the population of the $^2S_{1/2}$ F=0 ground state hyperfine level results in an increase in signal of a factor of 18 under the conditions of our experiment. This increase results from the fact that the $^2P_{1/2}$ F=0 $\rightarrow$ $^2S_{1/2}$ F=0 transition is electric-dipole forbidden, so that each ion which is left in the $^2S_{1/2}$ F=1 ground state hyperfine level after interaction with the microwave field can emit several optical photons before decay to the $^2S_{1/2}$ F=0 ground state hyperfine level occurs. This decay may occur either via off-resonant excitation to the $^2P_{1/2}$ F=1 level, followed by allowed spontaneous emission to the $^2S_{1/2}$ F=0 ground state hyperfine level, or via the metastable $^2D_{3/2}$ level [11,12].

The data shown in Fig. 9 was obtained using pairs of 400ms $\pi/2$ microwave pulses separated by an interaction time t_i of 10 s, resulting in Ramsey fringes with a line Q of 2.5×10^{11}.

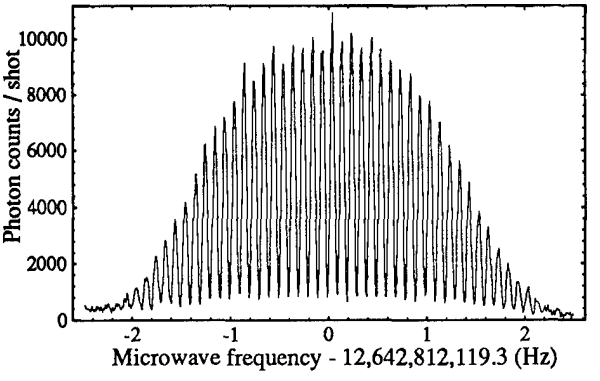


Figure 9: Ramsey fringes recorded using 10 s between microwave $\pi/2$ pulses. Each data point corresponds to a single pulse sequence. A background of 1200 counts/shot due to stray laser scatter has been subtracted.

Operation as a frequency standard

Operation as a frequency standard is achieved by servo-locking the frequency of the 12.6 GHz microwave signal to

a particular Ramsey fringe. The stability of the standard is evaluated by comparison with a hydrogen maser.

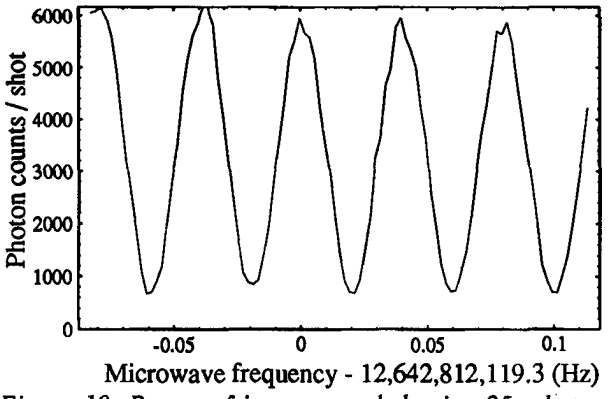


Figure 10: Ramsey fringes recorded using 25 s between microwave π/2 pulses. Each data point corresponds to a single pulse sequence. A background of 1500 counts / shot due to stray laser scatter has been subtracted. The noise and amplitude drift is primarily due to laser power fluctuations.

The Ramsey fringes were generated using the sequence shown in Fig. 8, with the microwave pulses separated by 25 s resulting in a Ramsey fringe separation $\Delta f_R=(25\ s)^{-1}=0.04$ Hz (Fig. 10). The measurement sequence was as follows: Two signals, S_- and S_+, in units of photon counts, corresponding to two microwave frequencies f_- and f_+ separated by half the Ramsey fringe separation were measured. The offset Δf between the mean frequency $f_m = (f_-+f_+)/2$ and a peak or trough of the Ramsey fringe is given (for $\Delta f << \Delta f_R$) by

$$\Delta f = \frac{S_+ - S_-}{A_R}\ \frac{\Delta f_R}{4}, \qquad (3)$$

where A_R is the amplitude of the Ramsey fringes in units of photon counts. Corrections are incorporated for laser power fluctuations and the 1 mHz resolution of the HP3335 synthesizer used to vary the microwave frequency (Fig. 7).

The mean frequency f_m was servo-locked to the centre of a Ramsey fringe using a three-term digital integrator [13], such that after the (i-1) th measurement of Δf, the mean frequency $f_{m[i]}$ is given by

$$f_{m[i]} = f_{m[i-1]} + g\left(0.42\Delta f_{[i]} + 0.67\Delta f_{[i-1]} - 0.08\Delta f_{[i-2]}\right), \qquad (4)$$

where g is a gain parameter, set to 0.5. Thus, with the servo loop closed, the mean frequency $f_{m[i]}$ is the standard frequency.

Figure 11 (a) shows the variation in the mean frequency f_m with respect to a H maser over a period of continuous operation lasting 15 hours. The trap was loaded at the beginning of the measurement, and was not reloaded during the measurement. Figure 11 (b) shows the fluorescence signal $S_- + S_+$, corrected for laser power fluctuations, over the same period.

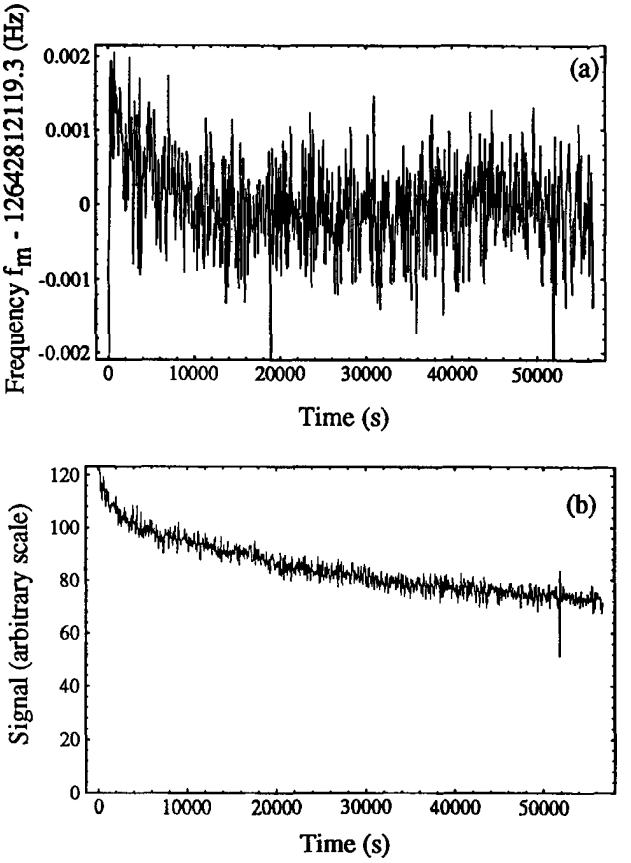


Figure 11: a) Measurement of the standard frequency f_m as a function of time. b) Variation of the signal from the ion trap (compensated for laser power drift) during the same measurement. The end electrode potential was 20 V.

The signal from the ion trap decays due to loss of ions from the trap, and due to population of metastable levels of $^{171}Yb^+$ [11,12]. Because of the low duty cycle of the laser, the decay of the signal in the present experiment is thought to be primarily due to loss of ions from the trap. This conclusion is indicated by the fact that only a small fraction of the drop in signal with time is recovered by operating the electron source (used for loading the trap), which has been shown to depopulate the metastable levels [14]. Direct electrical measurement of the number of ions in the trap (under development) will provide quantitative information on this matter.

During the first 3 hours after loading the trap, the frequency f_m decreases by approximately 1.5 mHz, and drifts much more slowly thereafter. This decrease in frequency is not obviously associated with a feature in the fluorescence signal. Measurements of the frequency of the Zeeman sidebands over this period have indicated a decrease in the

effective magnetic field, averaged over the trajectory of the ions, sufficient to explain the frequency drift. There is strong evidence that the drift is due to several insulating surfaces within the vacuum chamber becoming charged by electrons emitted from the electron source, and the ion cloud consequently having a spatial displacement, into a region of slightly higher magnetic field, which gradually reduces as the charge dissipates.

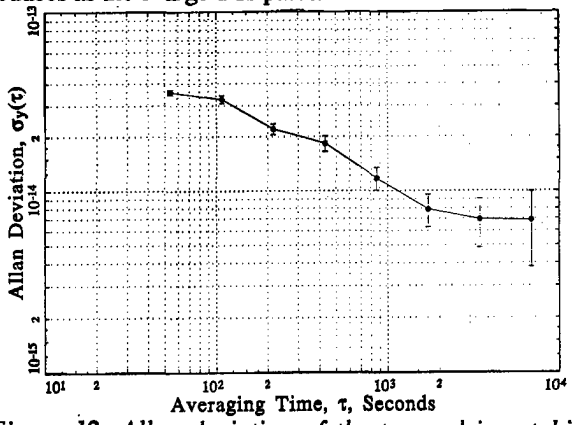


Figure 12: Allan deviation of the trapped ion-stabilised sapphire resonator system, measured with respect to a hydrogen maser.

The fractional Allan deviation $\sigma_y(\tau)$ of the data of fig 11 (with the first 10000 s dropped), is shown in Fig. 12, and indicates a performance of $\sigma_y(\tau)=3.7 \times 10^{-13}\,\tau^{-1/2}$ for $\tau < 3000$ s.

Sensitivity of the standard to trap parameters

In this section we present a phenomenological survey of the sensitivity of the standard frequency to some of the operating parameters of the ion trap.

The dependence of the standard frequency f_m on the voltage V_e applied to the end electrodes is shown in Fig. 13. At 20 V the slope of the fractional frequency dependence is

$$\frac{1}{f_m}\left(\frac{\partial f_m}{\partial V_e}\right)_{20\ V} = -5.7\times10^{-14}V^{-1}. \qquad (5)$$

The end electrode voltage presently fluctuates less than ±0.01 V, and consequently is not yet a source of significant instability in the standard. The decrease in the standard frequency with increasing end electrode voltage has been, in the case of a linear ion-trap $^{199}Hg^+$ standard [1], attributed to the increased second-order Doppler shift arising from RF heating and increased micro-motion due to the ions being squeezed away from the centre of the ion trap. In the present case, however, measurements of the ion cloud temperature show that the temperature of the ion cloud increases to approximately 100 °C above room temperature as the end electrode voltage is increased (Fig. 14). This temperature change can account for a second order Doppler shift of only 2 mHz, which is much less than the shift (~50 mHz) observed as the end electrode voltage is increased. The remainder of the shift would normally be attributed to to increased micro-motion amplitude, which in a linear trap is expected to result in a second order Doppler shift approximately proportional to the number of ions per unit length in the trap [7]. However, reducing the number of ions in the trap by 30% produced a frequency shift of less that 0.2 mHz. The dependence of the standard frequency on the end electrode voltage is therefore not yet fully understood.

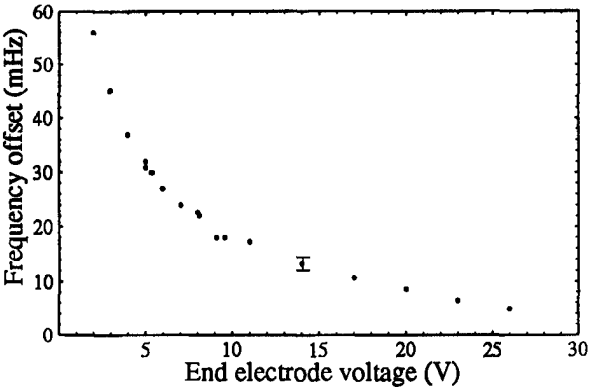


Figure 13: Dependence of the standard frequency on the voltage applied to the end electrodes.

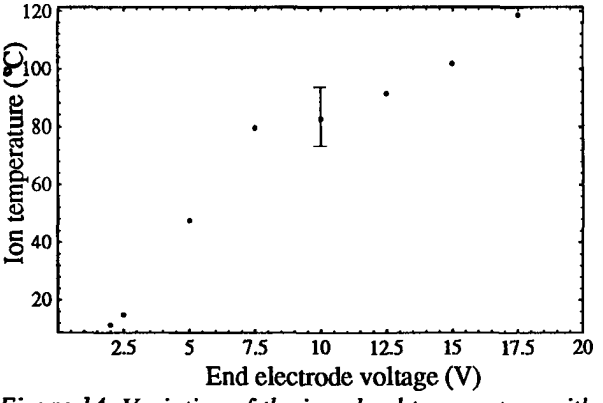


Figure 14: Variation of the ion cloud temperature with the voltage applied to the end electrodes.

Figure 15 shows the dependence of the standard frequency and the ion cloud temperature on He buffer gas pressure. There are two distinct regimes evident: Below 2×10^{-4} Pa the pressure dependence of the standard frequency is dominated by the second order Doppler shift, as the ion cloud temperature increases with decreasing He pressure. Above 2×10^{-4} Pa the ion cloud temperature remains essentially constant with increasing He pressure, and the standard frequency increases linearly due to a pressure shift. In this region the slope of the fractional frequency dependence is

$$\frac{1}{f_m}\left(\frac{\partial f_m}{\partial P_{He}}\right)_{P_{He}>2\times10^{-4}\,Pa} = 2.67 \pm 0.03 \times 10^{-10} Pa^{-1}. \quad (6)$$

It should be noted that this value is calculated using the ion gauge manufacturer's published relative sensitivity to He, which has not been checked on our system.

At the operating pressure of 7.4 x 10^{-5} Pa the slope of the fractional frequency pressure dependence is 1.0 ± 0.5 x 10^{-9} Pa^{-1}. Fluctuation in the unstabilised He pressure is therefore likely to be a significant source of instability in the standard in its present form. Operation at higher pressures may yield better stability, at the possible expense of faster loss of signal due to collisional population of metastable levels.

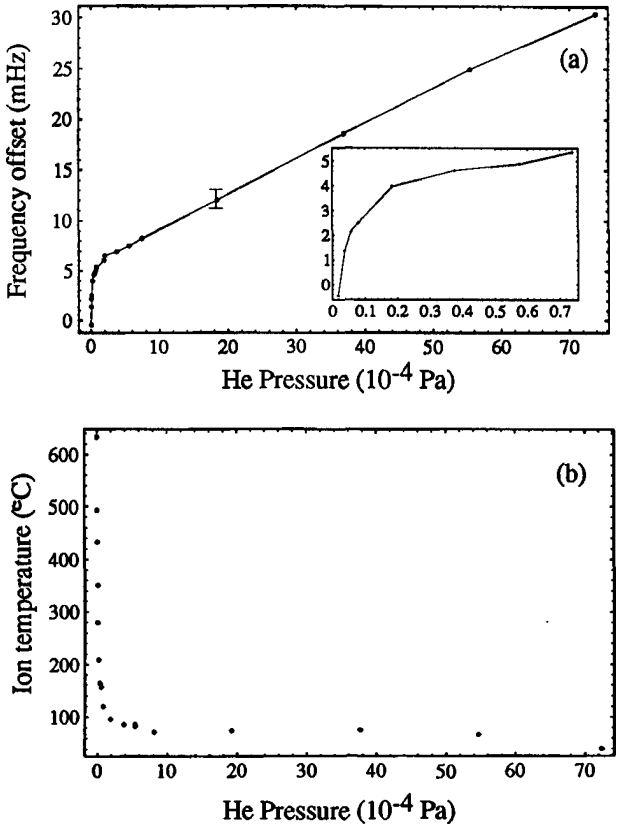


Figure 15: a) Dependence of the standard frequency on He pressure. b) Dependence of the ion cloud temperature on He pressure.

Variation of the frequency and amplitude of the RF trapping fields over intervals several times larger than the present fluctuations of these parameters produced frequency shifts of < 0.2 mHz (the present limit of resolution of the measurements made). A systematic study of these effects will be carried out after active stabilisation of the He pressure is implemented.

Conclusion

A microwave frequency standard based on buffer-gas cooled $^{171}Yb^+$ ions confined in a linear Paul trap has been demonstrated in prototype form. The standard exhibits a fractional Allan deviation $\sigma_y(\tau)=3.7 \times 10^{-13}\,\tau^{-1/2}$ for $\tau < 3000$ s, and the stability is currently limited by a slow drift which may be related to charging of insulated surfaces within the vacuum system, and by instability in the He buffer gas pressure.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] J.J. Bollinger, D.J. Heinzen, W.M. Itano, S.L. Gilbert, and D.J. Wineland, "A 303 MHz frequency standard based on trapped Be^+ ions," *IEEE Trans. Inst. Meas.*, vol. 40, no. 2, 126, April 1991.

[6] C. Tamm, "A tunable light source in the 370 nm range based on an optically stabilized frequency-doubled semiconductor laser," *App. Phys. B*, vol. B56, no. 5, 295, May 1993.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] H. Lehmitz, J. Hattendorf-Ledwoch, R. Blatt, and H. Harde, "Population trapping in excited Yb ions," *Phys. Rev. Lett.*, vol. 62, no. 18, 2108, May 1989.

[12] A. Bauch, D. Schnier, and C. Tamm, "Collisional population trapping and optical deexcitation of ytterbium ions in a radiofrequency trap," *J. Mod. Opt.*, vol. 39, no. 2, 389, 1992.

[13] G.H. Hostetter, *Digital control system design.* Holt, Rinehart and Winston, 1988.

[14] H.A. Klein, A.S. Bell, P. Barwood, P. Gill, and W.R.C. Rowley, "Studies of laser-cooled trapped Yb^{+}," *IEEE Trans. Inst. Meas.*, vol. 40, no. 2, 129, April 1991.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

RECENT STABILITY COMPARISONS WITH THE JPL LINEAR TRAPPED ION FREQUENCY STANDARDS[1]

R. L. Tjoelker, J. D. Prestage, G. J. Dick, L. Maleki

California Institute of Technology, Jet Propulsion Laboratory
4800 Oak Grove Drive, Bldg 298
Pasadena, California 91109

Abstract

The $^{199}Hg^{+}$ research frequency standards LITS-1 and LITS-2 were developed to provide continuous, reliable, high stability performance. For simplicity, a ^{202}Hg lamp is used for state selection and a helium buffer gas for ion cooling. In a preliminary 9 day comparison between the trapped ion standards, the Allan deviation was $\sigma_y(\tau)= 1x10^{-13}/\tau^{1/2}$ and a fractional frequency stability of $6x10^{-16}$ measured for averaging times greater than 10^5 seconds. A 40 day comparison of LITS-2 against an auto-tuned H-maser referenced to UTC-NIST puts an upper limit on long term drift of LITS-2 of $1.2(1.4)x10^{-16}$/day.

Introduction

Trapped ion frequency standards show great promise towards fulfilling several intermediate and long term frequency and timing needs. Ion trap based standards have the main advantage that the ion (atomic oscillator) is confined only by electromagnetic fields. Perturbations due to collisions are greatly reduced and ions can in principle be held indefinitely allowing for extremely long interrogation times. The $^{199}Hg^{+}$ ion is particularly well suited for frequency standards because the large mass and ≈40.5 GHz ground state hyperfine splitting reduce sensitivity to thermal and magnetic variations. Research standards LITS-1 and LITS-2 were developed to provide continuous high stability operation. These linear ion trap [1] standards (LITS) use a ^{202}Hg lamp to generate 194 nm light for optical state selection [2] and helium buffer gas to cool the ions to near room temperature [3].

The microwave $^2S_{1/2}(F=0,m_F=0)$ to $^2S_{1/2}(F=1,m_F=0)$ hyperfine transition of $^{199}Hg^{+}$ has a measured $Q > 2x10^{12}$ [5]. Good signal to noise is achieved with as many as $3x10^7$ ions in a linear ion trap. Several local oscillators (LO) have been used, including a good quartz crystal, a H-maser, or the Superconducting Cavity Maser Oscillator (SCMO) [4,5,6]. Short term performance of $7x10^{-14}/\tau^{1/2}$ [6] is obtained using a hydrogen maser as the local oscillator. With improvements to the optics configuration, we estimate the lamp based system is capable of $4x10^{-14}/\tau^{1/2}$.

Measured environmental sensitivity [6] indicate that an order of magnitude improvement compared to H-maser stability is possible with regulation levels still less stringent than for masers. Because a large number of mercury ions are confined at room temperature, the second order Doppler shift is the leading perturbation that will dictate the stability floor and the system accuracy. Current frequency accuracy is about 10^{-13}, though with an ion temperature measurement accurate to 1% [7] overall accuracies of 10^{-14} should be possible. A cryogenic, laser based $^{199}Hg+$ standard is currently under development at NIST [8]. This approach uses only a few ions which limits signal to noise, but has the potential of high absolute accuracy with long averaging times. Both approaches will benefit if current research to develop an ultra-violet diode laser capability is successful (see e.g. [9]), The JPL standards would achieve even better short term stability, and laser cooling would become much more practical.

In this paper, we report the first 9 day stability comparison between the JPL Hg+ trapped ion research standards LITS-2 and LITS-1. This comparison demonstrates stabilities well into the 10^{-16} range for averaging times longer than 100,000 seconds (Fig. 1). We also report a 40 day stability comparison between LITS-2 and two H-masers.

Long Term Stability and Environmental Sensitivity

The limiting long term stability depends on the frequency sensitivity of the hyperfine transition to confinement and environmental perturbations. Typical operating conditions, frequency offsets, and measured sensitivities have been previously reported [6]. The accuracy

[1]This work was performed by the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration.

0-7803-1945-1/94 $3.00 © 1994 IEEE

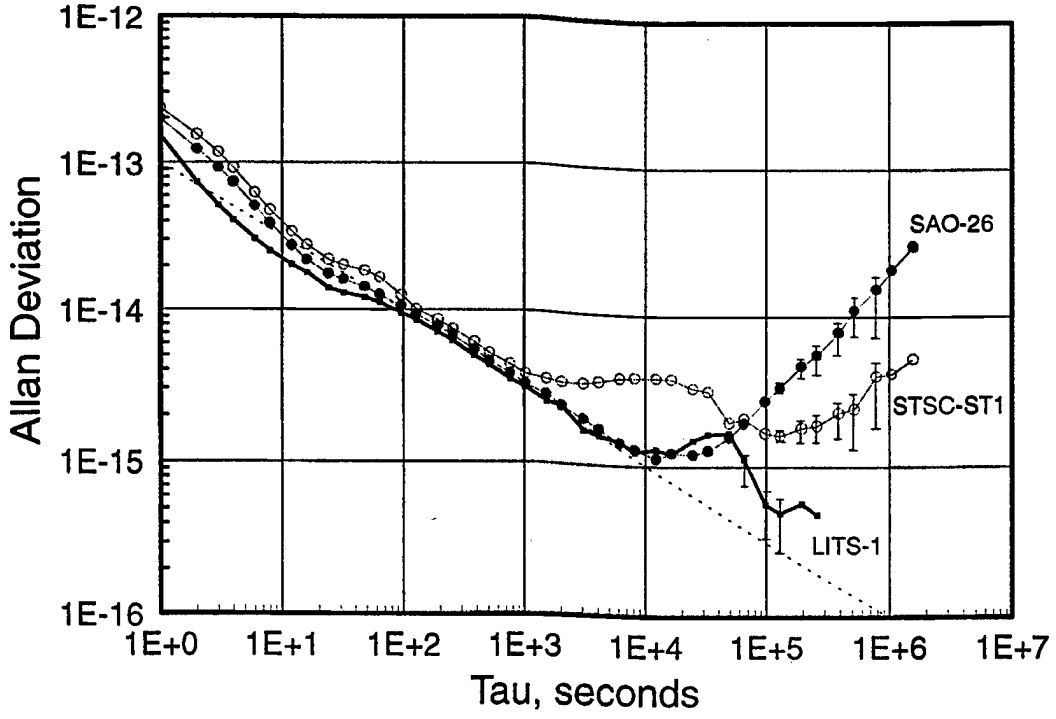


Figure 1: Forty day stability comparison of the Hg+ standard LITS-2 against (a) SAO-26 H-Maser, (b) STSC ST-1 Auto-tuned H-Maser, and (c) a nine day comparison against the Hg+ standard LITS-1.

and limiting stability of the trapped ion standards depends on how well these offsets are understood and held constant.

To measure the stability between the two $^{199}Hg^{+}$ standards each steers a separate VLG-11 [10] H-maser receiver (Fig. 2). These local oscillators consist of a good crystal oscillator phase locked to a common H-maser oscillator. Each LO is steered at approximately 20 second intervals based on the error signal determined by the microprocessor controlling the ion trap interrogation cycle. Both receivers provide a 100 MHz output and one is offset by 1 Hz. The 1 Hz beat is measured and the stability analyzed. As shown in Figure 2, the steered output of each LO is also compared against other available reference standards.

For this initial comparison both standards were operated with a 16 second microwave interrogation cycle and a performance of $1x10^{-13}/\tau^{1/2}$. Figure 1 shows the Allan deviation of LITS-2 compared to three separate references, LITS-1, the H-maser SA0-26 [10], and the auto-tuned H-maser STSC-ST1 [11]. The SA0 maser is useful for determining performance for averaging intervals less than 20,000. The STSC-ST1 maser has poorer short term stability, but is useful as a long term reference [12]. The STSC-ST1 maser is also independently compared to UTC-NIST via GPS to provide a reference to the international time scale.

The data shown in Figure 1 has no drift removed, though a $\sqrt{2}$ has been removed from the LITS-2 vs. LITS-1 comparison for averaging times greater than 20 seconds.

LITS-2 and LITS-1:

The stability between the two trapped ion standards reaches approximately $6x10^{-16}$ at 100,000 seconds. This point consists of 6 samples and the uncertainty is shown in figure 1. The peak at approximately 50,000 seconds resulted from a poor regulation circuit on LITS-1. This is made graphically clear in Figure 3 which shows the Allan deviation of each trapped ion standard compared against SAO-26. An oscillation is observed in the frequency residuals of both comparisons involving LITS-1. The oscillation is not present in LITS-2, which has better control electronics. In this preliminary 9 day measurement the differential drift between LITS-2 and LITS-1 is $3.2(2.7)x10^{-16}$/day. This small drift correlates well with a known sensitivity and measured drift of the RF trapping potential of LITS-1 during the comparison. The long term drift of the SAO maser is measured independently by both LITS-1 and LITS-2 during the same time interval of $4.4(0.3)x10^{-15}$/day and $3.7(0.6)x10^{-15}$/day respectively. The drift rate of the maser changes over time (see also [12]). For the 40 day comparison between SAO-26 and LITS-2 (Fig. 1) the measured drift is $2.4(0.3)x10^{-15}$/day.

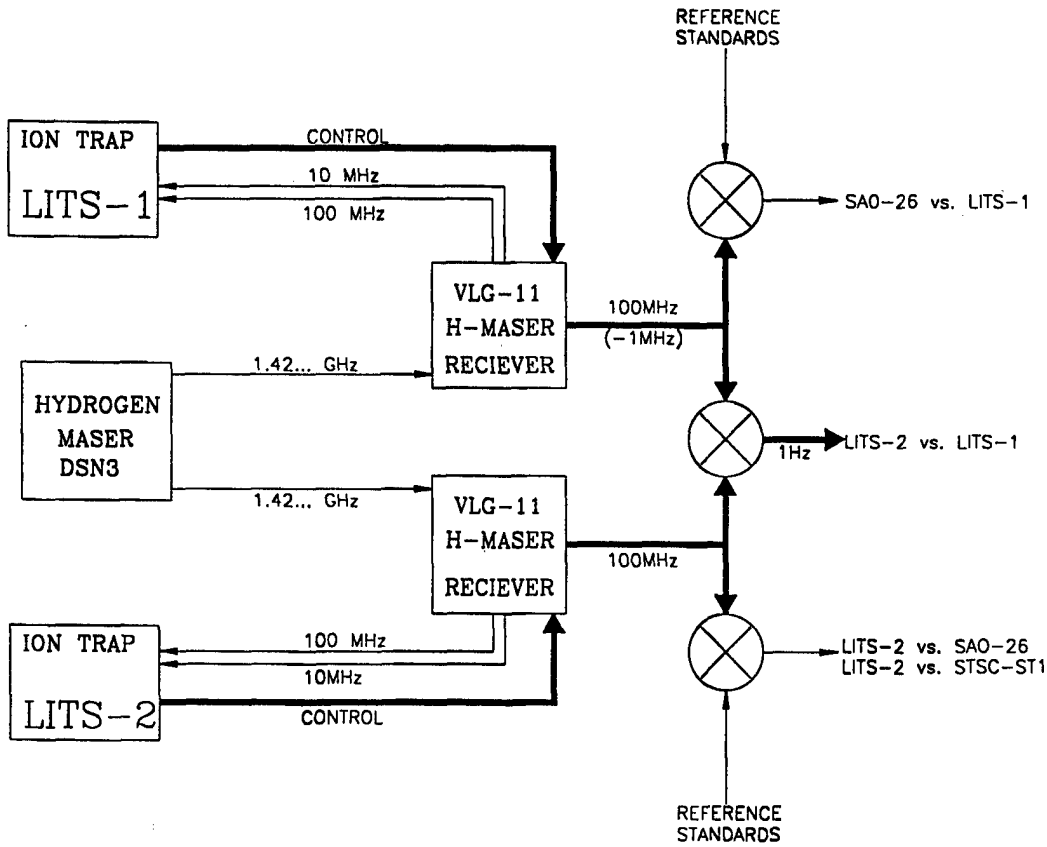


Figure 2: The measurement scheme used to compare the two trapped ion standards LITS-1 and LITS-2. Both Ion traps steer a separate VLG-11 H-maser receiver.

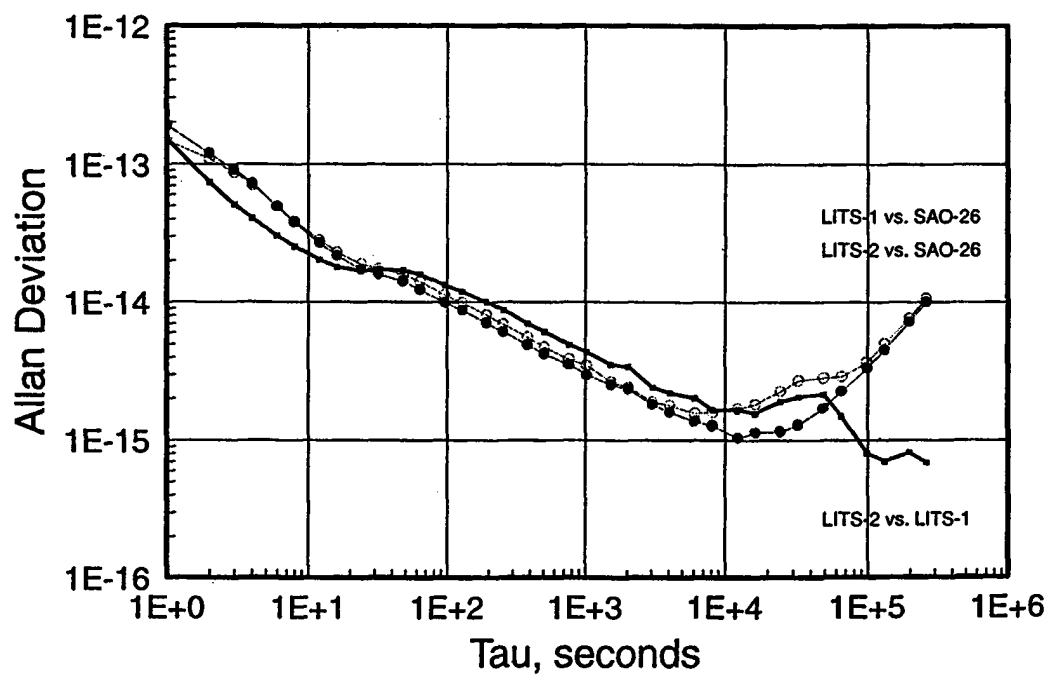


Figure 3: Nine day conparison between the Hg+ trapped ion standards LITS-2 and LITS-1. Both standards are also compared to the H-maser SAO-26 over the same time interval.

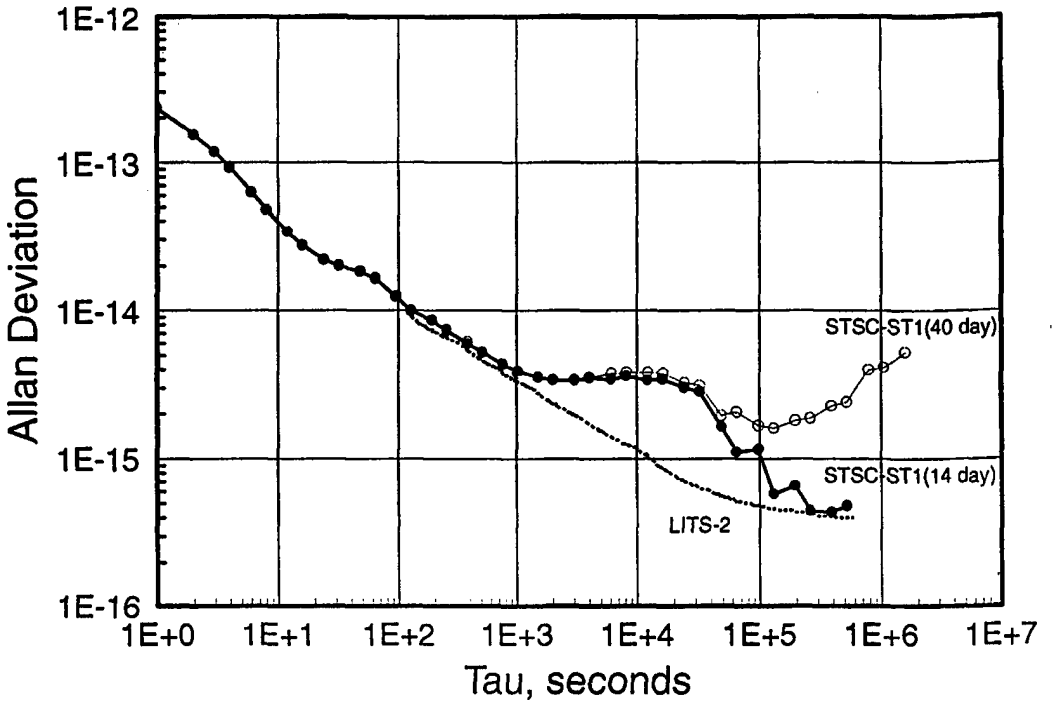


Figure 4: The Stability of LITS-2 compared against the H-maser STSC ST-1 for a selected 14 day period between frequency jumps in the H-maser. Also shown is the complete 40 day measurement.

LITS-2 and STSC-ST1:

In Figs. 1 and 4 the Allan deviation of a 40 day comparison between LITS-2 and the auto-tuned H-maser STSC-ST1 is shown. The differential drift between these two standards over this 40 day span is $4.7(1.6)\times10^{-16}$/day. A closer examination of the time residuals shows a large 10^{-14} frequency shift 10 days into the measurement. This shift can not be accounted for in LITS-2 and is apparently due to a frequency jump in the STSC-ST1 maser. This frequency change is confirmed in long term time residuals in GPS measurements comparing the STSC-ST1 H-maser to UTC-NIST. Similar frequency jumps a few times a year have also been reported elsewhere [12]. For the purpose of characterizing the stability of LITS-2, long term reference of the maser to UTC-NIST indicates stable reference windows between frequency jumps in STSC-ST1.

Figure 4 shows the Allan deviation for a two week interval of the 40 day comparison. During this time interval the frequency stability of the STSC-ST1 H-maser is particularly good as confirmed by the GPS comparison with NIST. In this 14 day comparison, the differential drift between LITS-2 and the STSC-ST1 maser is measured to be $1.2(1.4)\times10^{-16}$/day. For averaging times longer than 100,000 seconds this measurement is in agreement with the performance of LITS-2 as measured by LITS-1 (Fig. 1).

Electronic Improvements and Reducing Sensitivity With The Extended Linear Ion Trap (LITE)

Because of the low sensitivity to thermal and magnetic perturbations, averaging to 10^{-15} stability is accomplished with only minimal electronic control and isolation from environmental perturbations [6]. LITS-1 and LITS-2 are research laboratory standards and though portable, are not highly regulated. The data presented here was obtained with the standards thermally regulated to 0.05 C and a low field differential magnetic shielding factor of only 800. The trapping potentials are run "open loop" and the ion number is not actively servoed. Several improvements to the control electronics are currently under development which should allow the standards to average with characteristic $1/\tau^{1/2}$ behavior to near 1×10^{-16}.

In addition to relying on further electronic improvements for improved stability there are ways to reduce fundamental sensitivity and still maintain a practical, room temperature, lamp based system. An extended version of the linear ion trap (LITE) is currently under development [13] which takes advantage of the capability to easily move ions. By moving ions between two regions of a linear ion trap, the two often conflicting tasks of ion loading and optical state selection can be separated from the microwave interrogation region which requires an excellent magnetic

environment. Moving the ions into a long interrogation region reduces the linear ion density without sacrificing signal to noise. This not only reduces sensitivity to second order Doppler perturbations but may allow operation at lower magnetic fields.

Conclusion

A second ^{199}Hg+ trapped ion frequency standard LITS-2, now under continuous operation, provides a capability for measuring stability beyond all existing frequency standards for averaging times longer than 20,000 seconds. In a recent frequency stability comparison between ion trap standards LITS-1 and LITS-2, each standard steered a separate VLG-11 hydrogen maser receiver demonstrating stabilities of 6×10^{-16} for averaging times up to 9 days. The Allan deviation of each standard was $\sigma_y(\tau)= 1\times10^{-13}/\tau^{1/2}$ with the differential drift measured to be $3.2(2.7)\times10^{-15}$/day. This remaining small drift is predominantly in LITS-1 and correlates well with a measured drift in the trapping potential. A 40 day comparison of LITS-2 against an auto-tuned H-maser referenced to UTC-NIST provides an upper limit on the drift of LITS-2 of $1.2(1.4)\times10^{-16}$/day.

With both standards operating at the previously demonstrated short term performance of $\sigma_y(\tau)=7\times10^{-14}/\tau^{1/2}$ a stability of 1×10^{-16} should be possible in 5×10^5 seconds given sufficient magnetic shielding and stability in the control electronics. Measured environmental sensitivities indicate this can be accomplished with regulation still less stringent than for hydrogen masers. In addition, use of a new extended linear ion trap (LITE) configuration should further reduce remaining sensitivity to ion number and magnetic field fluctuations, allowing for even higher stabilities.

Acknowledgements

We thank A. Kirk who proposed the measurement scheme described in Fig. 2 and W. Diener for assistance with the stability measurements.

References

[1] 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.

[2] F. G. Major and G. Werth, "Magnetic Hyperfine Spectrum of Isolated ^{199}Hg$^+$ Ions", *Appl. Phys.* **15**, 201-208, 1981.

[3] 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.

[4] J.D. Prestage, R.L. Tjoelker, G.J. Dick, L.Maleki, "Ultra-Stable Hg$^+$ Trapped Ion Frequency Standard", *J. Mod. Opt.* **39**, No. 2, 221-232, 1992.

[5] J.D. Prestage, R.L. Tjoelker, G.J. Dick, L.Maleki, "Hg+ Trapped Ion Standard Performance With the Superconducting Cavity Maser Oscillator", IEEE Trans. Instr. and Meas. **42**, April 1993.

[6] 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, p 132, June 1993.

[7] J.D. Prestage, R.L. Tjoelker, G.J. Dick, and L. Maleki, "Doppler Sideband Spectra of Hg Ions in a Linear Ion Trap", Proc. 47th Ann. Symp. Freq. Control, p. 144, June 1993.

[8] M.G. Raizen, J.M. Gilligan, J.C. Bergquist, W.M. Itano, and D.J. Wineland, Phys. Rev. A, Vol. 45, p. 6493, 1992.
M.E. Poitzsch, J.C. Bergquist, W.M. Itano, & D.J. Wineland, "Progress On A Cryogenic Linear Trap For 199Hg+ Ions", Proc. 48th Ann. Symp. Freq. Control, these proceedings, June 1994.

[9] G.F. Neumark, R.M. Park, and J. M. DePuydt, "Blue-Green Diode Lasers", Phys. Today **47**, p.26 1994.
See also, S. Nakamura, T. Mukai, and M. Senoh, "Candela-class high-brightness INGaN/AlGaN double-heterostructure blue-light-emitting diodes", Appl. Phys. Lett. **64**, p. 1687, 1994.

[10] Smithsonian Astrophysical Observatory; Cambridge, Massachusetts 02138 USA.

[11] Sigma Tau Standards Corporation; Tuscaloosa, Alabama 35403 USA.

[12] 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.

[13] J.D. Prestage, R.L. Tjoelker, G.J. Dick, and L. Maleki, "Improved Linear Ion Trap Package", Proc. 47th Ann. Symp. Freq. Control, p. 148, June 1993.
J.D. Prestage, R.L. Tjoelker, G.J. Dick, and L. Maleki, "Progress on the Linear Ion Trap Extended (LITE)", Proc. 48th Ann. Symp. Freq. Control, these proceedings, June 1994.

PROGRESS ON A CRYOGENIC LINEAR TRAP FOR $^{199}Hg^+$ IONS*

M.E. Poitzsch,† J.C. Bergquist, W.M. Itano, and D.J. Wineland
Time and Frequency Division, NIST
Boulder, Colorado 80303, USA

Abstract

We have observed linear "crystals" of up to tens of laser-cooled $^{199}Hg^+$ ions in a new cryogenic linear rf ion trap that operates at liquid helium temperature and is designed for use as a prototype 40.5 GHz frequency standard with high accuracy and stability.

$^{199}Hg^+$ Atomic Clock System

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]. The work at NIST on $^{199}Hg^+$ [5] has been devoted to obtaining a system that will provide high accuracy as well as high stability. To achieve this goal, we have incorporated laser cooling to suppress the second-order Doppler shift. In this paper, we report preliminary results using a cryogenic ($\sim$ 4 K) trap system. This should yield high vacuum, thereby reducing ion loss and frequency shifts due to background gas collisions, and should provide the basis for superconducting magnetic shielding.

At Hewlett-Packard [3] an rf-trapped $^{199}Hg^+$ ion frequency standard (using buffer gas cooling) has been demonstrated to have high frequency stability. It contained $N \sim 2 \times 10^6$ ions and had a fractional second-order Doppler shift of $\sim -2 \times 10^{-12}$. At JPL [4], 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). 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 is -2×10^{-18} [5]. The fractional frequency shift of the 40.5 GHz clock transition with magnetic field is $0.24B^2$, where B is expressed in teslas. Thus, a $^{199}Hg^+$ ion confined in an ion trap at near-zero magnetic field and laser-cooled to the Doppler limit should constitute a highly accurate 40.5 GHz microwave frequency standard, assuming sufficient magnetic shielding. 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.

0-7803-1945-1/94 $3.00 © 1994 IEEE

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. In addition, the 4 K temperature allows us to operate a superconducting shield around the ion trap region to help in shielding out changes in the magnetic field.

We have constructed and are testing a prototype apparatus based on the above 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 by positively biasing rings at either end of the four-rod quadrupole, whose separation 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 superconducting 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 field-emitter point 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. The superconducting shield consists of a thin coating of lead, electroplated onto the inside of the copper vacuum vessel.

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. If left undisturbed, these crystals are very stable over periods of several hours. One rough measure of the background gas pressure is the rate at which trapped "impurity" ions of different species (which show up as non-fluorescing spots in the crystal) exchange places with their $^{199}Hg^+$ ion neighbors. This seems to occur very infrequently in our cryogenic vacuum. We hope to be able to demonstrate sub-hertz linewidths of the clock transition and assess the stability and accuracy of this frequency standard prototype.

Assuming a 10 s Ramsey interrogation time, the short-term fractional frequency stability of an ensemble of 20 ions could be $< 3 \times 10^{-13}\tau^{-1/2}$ if we succeed in detecting the ions independently and with nearly 100% detection efficiency. With sufficient magnetic shielding, a fractional inaccuracy of $< 1 \times 10^{-16}$ appears attainable.

In addition, this apparatus contains features (the superconducting coil pairs) that should allow us to investigate new effects based on motional Zeeman coherences. These include a novel cooling scheme (proposed by Harde [11]) 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 [12].

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] H. Harde, in Int. Conf. on Quantum Electronics Tech. Digest, Series 1990, Vol. 8. Washington, DC: OSA, p. 178, 1990.

[12] D.J. Wineland, J.J. Bollinger, W.M. Itano, F.L. Moore, and D.J. Heinzen, Phys. Rev. A, Vol. 46, p. R6797, 1992.

SPACE FLYABLE Hg^+ FREQUENCY STANDARDS*

John D. Prestage and Lute Maleki
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109

Abstract

We discuss a design for a space based atomic frequency standard (AFS) based on Hg^+ ions confined in a linear ion trap. This newly developed AFS should be well suited for space borne applications because it can supply the ultra-high stability of a H-maser but its total mass is comparable to that of a NAVSTAR/GPS Cesium clock, i.e., about 11 kg. This paper will compare the proposed Hg^+ AFS to the present day GPS Cesium standards to arrive at the 11 kg mass estimate. The proposed space borne Hg^+ standard is based upon the recently developed extended linear ion trap architecture which has reduced the size of existing trapped Hg^+ standards to a physics package which is comparable in size to a Cesium beam tube. The demonstrated frequency stability to below 10^{-15} of existing Hg^+ standards should be maintained or even improved upon in this new architecture. This clock would deliver far more frequency stability per kilogram than any current day space qualified standard.

Introduction

Isolation from environmental perturbations drives the technology and development of all atomic frequency standards. The choice of atom used as the frequency discriminating element in a passive atomic frequency standard determines the sensitivity to external perturbations and consequently the degree to which shielding from these environmental changes will be necessary to reach a given level of stability. Because shielding, usually magnetic and thermal, adds a great deal to the bulkiness and complexity, it is especially important for space borne clocks, where low mass and high reliability are paramount, that the standard be inherently immune to environmental changes so that only modest shielding is required. Mercury ions with their relatively large hyperfine clock resonance frequency and large atomic mass are far more immune to environmental changes than any of the other microwave standards, i.e., hydrogen masers, rubidium gas cells, and cesium beam standards. For example, the fractional frequency shift with magnetic field, $(1/\nu_0)d\nu/dH \propto H/\nu_0^2$ where H is the magnitude of the C-field and ν_0 is the hyperfine clock transition frequency. For the same C-field bias the 40.5 Hg^+ transition is less sensitive to ambient field changes than Cesium (by 19×), Rubidium (by 35×) and Hydrogen (by 837×). This is very important for space borne standards where the spacecraft magnetic environment is much more variable than an earth based frequency standards laboratory. The extra shielding required for the harsher environment thus contributes even more to the clock mass which comes at a premium.

Similarly, H-masers require 10^{-4} °C active temperature regulation to reach 10^{-15} frequency stability while GPS Cesium clocks require 0.1 °C temperature stability to reach 10^{-14}. By contrast, the Hg^+ frequency standard requires 0.1 °C temperature stability to reach 10^{-15} frequency stability, showing that Hg^+ is additionally the most immune to environmental temperature changes. These comparisons demonstrate that this standard is an attractive candidate for space borne applications.

Applications of Space-based Ultra-stable Clocks

There are several uses, scientific and operational for stable atomic frequency standards on board spacecraft in earth orbit, planetary orbit or flyby, and interplanetary cruise mode. In generating Doppler data the ground based antenna observation time is reduced greatly for s/c navigation at the outer planets when the two way uplink-s/c transpond-downlink round trip is replaced by the one-way downlink from the s/c to earth station. This reduces the burden on ground based tracking facilities as the number of spacecraft to be tracked grows larger. There are other advantages for the one way downlink. In the absence of an uplink, media propagation errors and noise sources are reduced by one half. Additionally, when the ground station is configured for listen only operation, the receiver noise temperature is lower[1].

The scientific uses for ultra-stable atomic standards on board spacecraft include tests of gravitational theories and detection of low frequency gravitational waves. For example, an ulta-stable clock aboard a Solar Probe approaching the sun's center to within 4 solar radii could measure the gravitational redshift with an accuracy of a few parts in 10^9, nearly 5 orders of magnitude improvement over present day tests[2].

0-7803-1945-1/94 $3.00 © 1994 IEEE

Another class of scientific applications that will be improved upon with an on-board clock of better stability is the remote sensing of planetary atmospheres during occultation of the s/c. The pressure vs. temperature of the planet's atmosphere can be derived from the phase variation of the radio signal during occultation [1,3].

Navigation on and near the earth's surface is being revolutionized through the spaced based atomic clocks which comprise the Global Positioning System. Rubidium and Cesium standards aboard these satellites generate time coded signals to broadcast to Earth based receivers. Small handheld receivers process signals simultaneously recieved from different spacecraft to determine the users position with an accuracy of meters.

Survey of Space-based Atomic Frequency Standards

GPS/GLONASS Cesium Clocks

Although still in its youth, the technology of space based clocks is being refined through the development of the NAVSTAR/GPS and GLONASS programs. The GPS program has launched 40 satellites into orbits of 12 hours at altitude 20,000 km each with three to four AFSs and backup VCXOs [4,5].

Similary, the Russian GLONASS program has launched several navigation satellites with multiple redundant clocks on board.

In this section we review the physical characteristics of the cesium clocks aboard the NAVSTAR and GLONASS spacecraft since these are probably the most advanced space based clocks yet developed. Although the GPS satellites have flown more Rb standards, the Cs standards are considered primary because of their much lower frequency drift and environmental sensitivities [5]. Additionally, as will be shown, the Hg^+ LITE standard can be made similar in physical size and layout to the Cs beam standard.

Table I summarizes some of the features of the GPS and GLONASS clocks. The masses of the GPS Cs AFSs ranges from 13 kg for Block II standards to 8 kg for the Block IIR replacement clocks[6]. These values are for the clock alone and do not include approximately 1.8 kg (4 pounds) of radiation shielding to protect the standard from natural and potential manmade radiation sources.

The GLONASS clocks are somewhat more massive and bulky [7]. The MALAKHIT is an improvement over the GEM spaceborne Cs AFS and although more massive, it is expected to be longer lived and more immune to environmental changes [7].

The temperature variation in orbit can lead to pronounced frequency changes in the Rb AFS [8] and to a lesser extent in the Cs AFS [9]. The Rb AFS are sensitive enough that +0.1 C active temperature control was implemented [8]. Temperature variations experienced by the s/c and on-board clocks are strongly correlated to the NAVSTAR orbit orientation relative to the earth-sun line. For example, twice a year, for about a 25 day interval, NAVSTAR 16 goes into the earth's shadow during a portion of each 12 hour orbit. This causes a 1-3 C drop in the average temperature of the s/c and frequency pulling of the CAFS ranging from $2\text{-}7\times10^{-13}$ [9].

H-masers

Although several H-maser frequency standards have been developed for spaceborne applications [10-14], only the NASA/SAO Gravity Probe A suborbital flight (June, 1976) has carried an ultra-stable frequency standard into space [2,15]. In a 2 hour flight a 45 kg H-maser was lifted to an altitude of 10,000 km above the earth's surface where the gravitational redshift speeds the clock rate by 4×10^{-10} relative to an earth based clock. The measured H-maser stability in flight was $\sigma_y(\tau=1000 \text{ secs})=6\times10^{-15}$ and verified the gravitational redshift as predicted by Einstein's General Relativity to about 1 part in 10^4.

Table II summarizes some physical characteristics of two, more recently developed H-masers for space based operations [11-13]. These H-masers are higher performance but somewhat more massive than the 1976 GP-A maser. The mass of the maser is made large by the large size of the high Q TE_{011} cylindrical cavity together with the inherent magnetic sensitivity of the hydrogen atom. Typically 4 to 5 layers of magnetic shielding are required to prevent ambient field changes from pulling the atomic frequency and degrading stability. The shields must be large enough to enclose the approximately 30 cm(12") diameter cavity and consequently contribute 38 kg to the total $\approx$70 kg mass of the maser [12]. A dielectrically loaded cavity of reduced size for space based H-maser applications with somewhat reduced performance is described in ref [14].

Cs Clock parameters	GPS	GLONASS
Mass	13 kg (Block II) 8 kg (Block IIR) (29-17 lbs)	40 kg (GEM) 52 kg (MALAKHIT) (88-114 lbs)
Power (Watts)	26 W (II) 23 W (IIR)	80 W (GEM) 90 W (MALAKHIT)
Package Size (mm) (inch)	 150x150x430 6x6x17	 414x421x655 16x16x26
Stability,$\sigma_y(\tau)$ 1 day 10 days Frequency Drift	 1.5×10^{-13} $4\text{-}5\times10^{-14}$ $\pm$few$\times10^{-15}$/day	 $1\text{-}5\times10^{-13}$ $7\text{-}25\times10^{-14}$
Temperature Sensitivity	$1\text{-}2\times10^{-13}$/C no active control	$2\text{-}5\times10^{-13}$/C

Table I: A summary of present day space based Cesium atomic frequency standards in use in the GPS and GLONASS programs.

H-maser parameters	NASA/SAO	ESA
Mass	67 kg	70 kg
Power	55 Watts	70 Watts
Size	17"OD×34"h (44cm×86cm)	14"OD×27"h (35cm×70cm)
Stability,$\sigma_y(\tau)$ 10^3 to 10^4 sec 1 day	 $\leq10^{-15}$ few$\times10^{-15}$	 $\leq10^{-15}$ few$\times10^{-15}$
Temperature Sensitivity	requires 10^{-4} C temperature control	requires 10^{-4} C temperature control
Magnetic Shielding Ref [11-13]	5 layers & active field compensation 2×10^6 axial shielding	4-5 layers

Table II: A summary of two recently developed hydrogen maser frequency standards for use in space based applications.

Review of the Linear Ion Trap Extended (LITE)

The proposed space-borne frequency standard discussed in this paper is based upon the Hg^+ extended linear ion trap [17] currently under development at the JPL Frequency Standards Lab. This new linear ion trap architecture separates the resonance region (where the multiplied output of the local oscillator is compared to the stable atomic frequency) from the state-selection/state-interrogation region where the ions are optically pumped into the lower atomic hyperfine state. The charged ions can easily be moved from one region to another along the axis or node line of a single extended linear ion trap. One end of the extended linear trap serves as the state selection region with an optical system, etc. while the other end serves as the resonance region with a 40.5 GHz microwave source, magnetic shields, etc. Only the microwave resonance region need be magnetically shielded and since there are no optics in the resonance region (unlike the previous design [18]) the shields can be made quite small.

The LITE frequency standard now being developed at JPL is shown in Fig. 2. This first laboratory version is 18" in length with a 5 1/4" od triple layer set magnetic shield around the resonance region. This particular version can be reduced in size by about one-half by use of cylindrical electrodes as described in Ref [19]. In the resulting smaller trap, the cylindrical electrodes would also be the vacuum wall additionally removing the need for electrical feedthroughs. In this manner a linear ion trap of diameter ≤ 1" with several eV well depth for Hg^+ ions could be constructed. This is much smaller than the diameter of a Cesium beam tube.

Table III shows a breakdown of the contributors to the mass of an ion trap of the architecture discussed above. The trap is assumed to be about 16" long with an diameter of 1". The state selection\ion loading region is about 4" with the resonance region making up 12" of the length. The triple layer magnetic shield is assumed 14" long with 1/2" spacing between layers. These design estimates are based on and extrapolated from Hg^+ trap systems which have demonstrated performance well below 10^{-15} [20].

The support electronics are to a large extent similar to those of a generic Cesium frequency standard. For example, since they are both passive frequency standards, a frequency multiplier chain from the local oscillator to the atomic resonance frequency is necessary, to 40.5 GHz for Hg^+ and to 9.2 GHz for Cs. Table IV shows the support electronics that are common to both units and some additional electronics required by the Hg^+ standard. A very good crystal VCXO [21] would enable the Hg^+ standard to reach few parts in 10^{16} stability at $\tau = 10^5$ seconds, i.e., $2\times10^{-13}/\sqrt{\tau}$. This is more than is required for Cesium standard operation but such LOs exist and have been space qualified [21]. The state of the art for space qualified Cs standards are the GPS Block IIR [6], having undergone a few generations of refinement in the GPS clock development program. The super VCXO is about 1760 gms [21] while the GPS Block IIR Cesium VCXO is about 315 grams as shown in table IV.

Mass Estimate for the Hg^+ LITE Standard

Taking the Block IIR Cesium as current state of the art for compact space qualified Cs standards, we estimate the mass of a space qualified Hg^+ clock. The Cesium standard of IIR design has total mass 7.7 kg (17 lbs) of which the Cesium tube contributes 4.1 kg (9 lbs) [6]. Our Hg^+ tube is estimated in Table III to be somewhat heavier at 5.4 kg. Table IV shows that the electronics package for an Hg^+ clock with ultra-stable performance is 1760-315+500 = 1945 gms more in mass (primarily because of the better LO). We thus arrive at a mass estimate for an Hg^+ standard of $1.945 + 1.3 \approx 3.2$ kg more than the Block IIR Cesium mass of 7.7 kg, that is, a space qualified Hg^+ standard of mass 10.9 kg. This estimate is uncertain by about 1 kg since efficient packaging can be difficult to estimate before full miniaturization is implemented. There is little doubt, however that a high performance Hg^+ clock could be built that is $\leq 20\%$ of the mass of the proposed space qualified H-masers.

Hg^+ LITE Consumables

The Hg^+ standard described here needs a supply of Helium (to be used as a buffer gas) and, of course, Hg for the generation of ions. In ground based laboratory ion trap standards, helium operating pressure is about 10^{-5} Torr and is pumped away into a vacuum pump at a speed of about 1 liter/second. This throughput of 10^{-5} Torr liters/second would deplete a 1 liter helium bottle filled to 5 atmospheres in just over 12 years.

Similarly, the mercury vapor in an operating Hg^+ standard is at a pressure of 10^{-9} Torr or less and is pumped at even less speed than the helium. The mercury is thus consumed at a rate of less than 1 milligram per 3 years, again showing no problem in depleting the supply of mercury.

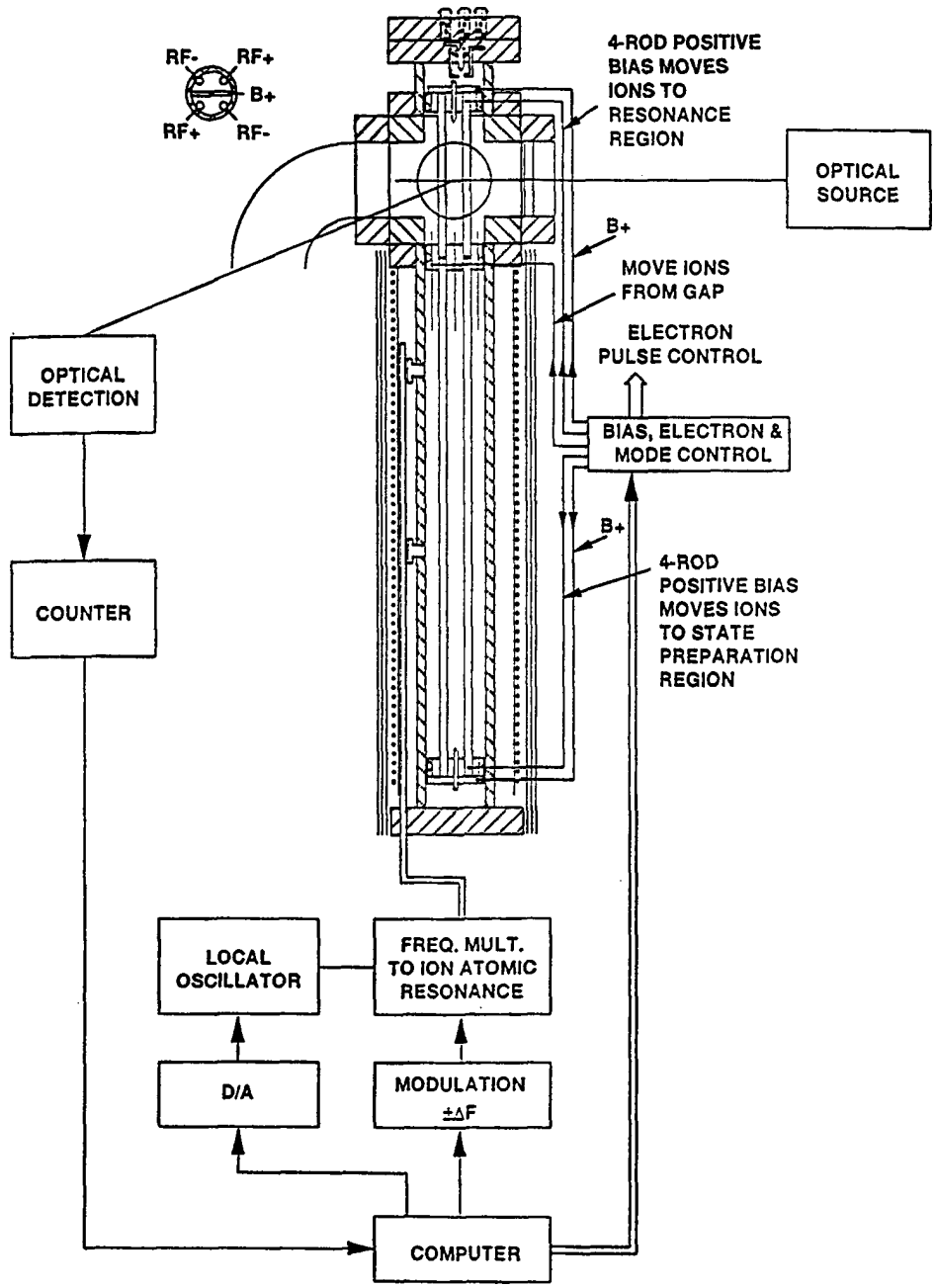


Figure 1: Schematic view of the linear ion trap extended architecture. The state preparation region is at the upper end of a ≈40 cm long linear trap. A dc break separates the resonance region (lower part of the long trap) from the upper part. By applying dc biases, ions are moved from the state selection region into the microwave resonance region and then back. Only the microwave resonance region need be magnetically shielded, etc. The resulting physics package is comparable in size to a cesium beam tube.

Vacuum can/Trap electrodes	790 gm
Triple layer Magnetic Shield	227 + 409 + 590 = 1226 gm
HgO Oven	150 gm
Fused Silica Windows & Mount	150 gm
Lamp & Mount	150 gm
Collection Mirror, Lens & Mount	400 + 35 + 500 gm
C-field Coils	250 gm
Electron gun & Mount	10 gm
Getter pump & Ion pump	1000 gm
Helium leak & Storage Bottle	750 gm
Total	5.4 kg

Table III: An itemized list of the contributors to the mass of the Hg^+ LITE atomic frequency standard. This list summarizes components in the physics package of the Hg^+ standard and is analogous to the Cesium beam tube of the Cesium frequency standard.

Electronics	Hg^+ LITE	Cesium
LO to Atom Frequency Multiplier	to 40.5 GHz	to 9.2 GHz
Servo LO to Atom	Microprocessor, Frequency modulation, etc.	modulate, tune VCXO etc.
Bias/power supplies	HgO oven, C-field, ion pump, e-gun, PMT, ion gauge	Cs oven, C-field, ion pump, e-mult, hot wire, ion accel
VCXO	Super VCXO, $\sigma_y=10^{-13}$ 1760 gms	VCXO 315 gm
Additional supplies	Helium leak heater, Lamp driver, Trap rf drive. (Estimate 500 gm)	Not Applicable

Table IV: A comparison of the support electronics for the Hg^+ LITE standard and the GPS Cesium standard. Much of the elctronics is common to both clocks. The Hg^+ clock needs a super VCXO to reach stabilies of a few parts in 10^{16} at 10^5 seconds. Some additional heaters and rf supplies are needed for the Hg^+ standard.

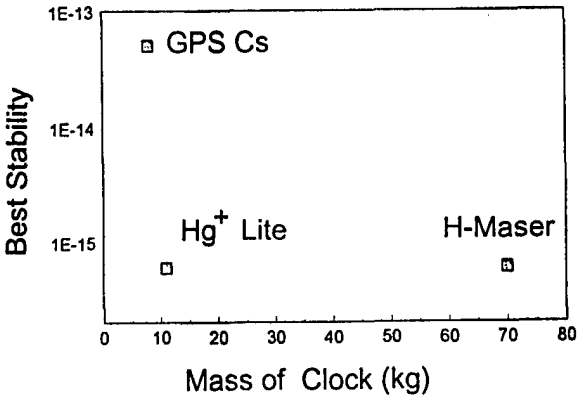


Figure 2. A summary of clock frequency stability vs. clock mass for three space based frequency standards discussed in this paper. The proposed Hg^+ LITE can supply stability of a maser in a total package with mass comparable to the current day GPS Cs standards.

Summary and Conclusions

The ultra-stable Hg^+ standard described here can deliver far more stability per kilogram than any of the present day space based standards. This is graphically illustrated in Figure 2 where H-masers, GPS Cesiums and the proposed Hg^+ standard are plotted on a stability vs. mass diagram. The Hg^+ offers the best of both of the other space based standards, that is, the stability of an H-maser for the mass of a Cesium clock.

*This work was performed by the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration.

References

[1] J. S. Border and E. R. Kursinski, "Deep Space Tracking and Frequency Standards," in Proceedings of the 45th Annual Symposium on Frequency Control, 1991, pp.594-607.

[2] R. F. C. Vessot, "Past and Future Tests of Relativistic Gravitation with Atomic Clocks," in Proceedings of the 1st William Fairbank Meeting on Relativistic Gravitational Experiments in Space, Rome, Sept. 1990.

[3] E. R. Kursinski, "Application of High Stability Oscillators to Radio Science Experiments using Deep Space Probes," in Proceedings of the 22th Annual PTTI Applications and Planning Meeting, 1990, pp. 253-268.

[4] GPS breakdown: GPS 1-3 each carried 3 RFSs, GPS 4-11 each carried 3 RFSs and 1 CFS, GPS 13-40 each carried 2 RFSs and 2 CFSs. GPS 1-12 is known collectively as Block I, GPS 13-21 is called Block II and GPS 22-40 is called Block IIA. The Block IIR replacement clocks represent a reduction in mass and power over the Block II and IIA Cesium clocks. See ref [5].

[5] M. J. Van Melle, "Cesium and Rubidium Frequency Standards Status and Performance on the GPS Program," in Proceedings of the 25th Annual PTTI Applications and Planning Meeting, 1993, pp. 329-344.

[6] J. A. Wisnia, "CAFS- A Cesium Atomic Frequency Standard for GPS Block IIR," in Proceedings of the 24th Annual PTTI Applications and Planning Meeting, 1992, pp. 209-220.

[7] Y. G. Gouzhva, A. G. Gevorkyan, A. B. Bassevich, P. P. Bogdanov, and A. Y. Tyulyakov, "Comparative Analysis of Parameters of GLONASS Spaceborne Frequency Standards when used onboard and on Service Life Tests," 1993 IEEE International Frequency Control Symposium, 1993, pp. 65-70.

[8] H. Bethke, D. Ringer, and M. Van Melle, "Rubidium and Cesium Frequency Standards Status and Performance on the GPS Program," in Proceedings of the 16th Annual PTTI Applications and Planning Meeting, 1984, pp. 127-141.

[9] T. B. McCaskill, W. G. Reid, J. A. Buisson, and H. E. Warren, "Analysis of the Frequency Stability of on Orbit GPS NAVSTAR Clocks," 1993 IEEE International Frequency Control Symposium, 1993, pp. 23-32.

[10] H. T. M. Wang and R. R. Hayes, "Design Considerations and Performance of a Spaceborne Hydrogen Maser Frequency Standard," in Proceedings of the 20th Annual PTTI Applications and Planning Meeting, 1988, pp. 111-122.

[11] R. F. C. Vessot, E. M. Mattison, G. U. Nystrom, L. M. Coyle, R. Decher, S. J. Feltham, G. Busca, S. Starker, and S. Leschiutta, "Hydrogen Masers for Space," in Proceedings of the 6th European Frequency and Time Forum, 1992, pp. 19-25.

[12] L. G. Bernier, "Basic Design of a Spaceborne Hydrogen Maser," in Proceedings of the 6th European Frequency and Time Forum, 1992, pp. 283-286.

[13] G. Busca, L. G. Bernier, S. Starker, and S. Feltham, "In Orbit Demonstration of a H-Maser Clock System," in Proceedings of the 20th Annual PTTI Applications and Planning Meeting, 1991, pp. 385-399.

[14] G. Busca, L. Bernier, P. Silverstrin, S. Feltham, B. A. Gaygerov, and V. M. Tatarenkov, "Hydrogen Maser Clocks in Space for Solid-Earth Research and Time-Transfer Applications: Experiment Overview and Evaluation of Russian Miniature Sapphire Loaded Cavity," in Proceedings of the 25th Annual PTTI Applications and Planning Meeting, 1993, pp. 467-480.

[15] R. F. C. Vessot and M. W. Levine, "A Test of the Equivalence Principle Using a Space-Borne Clock," General Relativity and Gravitation, vol. 10, pp. 181-204, 1979. See also R. F. C. Vessot et. al., "Tests of Relativistic Gravitation with a Space-borne Hydrogen Maser," Phys. Rev. Lett., vol. 45, pp. 2081-2084, 1980.

[16] R. F. C. Vessot, E. M. Mattison, G. U. Nystrom, L. M. Coyle, R. L. Walsworth, R. Nicoll, D. Graveline, and R. Decher, "Design of a Hydrogen Maser for Space and Progress on Cryogenic H-maser Research," in Proceedings of the 7th European Frequency and Time Forum, 1993, pp. 707-711.

[17] J. D. Prestage, R. L. Tjoelker, G. J. Dick, and L. Maleki, "Improved Linear Ion Trap Physics Package," in Proceedings of the 1993 International Frequency Control Symposium, 1993, pp. 144-147.

[18] J. D. Prestage, G. J. Dick, and L. Maleki, "Linear Ion Trap Based Atomic Frequency Standard," IEEE Trans. Instr. Meas., Vol. 40, pp. 132-136, April 1991.

[19] R. K. Melbourne, J. D. Prestage, and L. Maleki, "Analytical Potential in a Linear RF Quadrupole Trap with Cylindrical Electrodes," J. Appl. Phys., Vol. 69, pp. 2768-2775, March 1991.

[20] R. L. Tjoelker, J. D. Prestage, G. J. Dick, and L. Maleki, "Recent Stability Comparisons with the JPL Linear Trapped Ion Frequency Standards," in Proceedings of the 1994 IEEE International Frequency Control Symposium, 1994.

[21] J. R. Norton, "BVA-Type Quartz Oscillator for Spacecraft," Proceedings of the 45th Annual Symposium on Frequency Control, 1991, pp. 426-430.

1994 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 an improvement in the architecture of the physics package used in the Linear Ion Trap (LIT) based frequency standard recently developed at JPL. This new design is based on the observation that ions can be moved along the axis of an LIT by applied dc voltages. The state selection/interrogation region can be separated from the more critical microwave resonance region where the multiplied local oscillator (LO) signal is compared to the stable atomic transition. This separation relaxes many of the design constraints of the present units. Improvements include increased frequency stability, and a substantial reduction in size, mass and cost of the final frequency standard.

Introduction

The primary reason that ions confined to electromagnetic traps are the basis for stable frequency standards is that in the containerless environment of a trap, processes that equalize atomic state populations and destroy coherence within the state prepared atomic ensemble are very weak. In past work we have measured coherence times in an ensemble of trapped Hg^+ ions of over 30 seconds on the 40.5 GHz transition. Such weak relaxation has permitted an atomic line-Q on the 40.5 GHz transition of 2×10^{12} [1,2].
One disadvantage of a trapped ion based frequency standard is the relatively low density of the confined ions. This leads to low signal to noise ratio (SNR) in the detected atomic resonance and consequently limits clock stability. This situation was greatly improved by the use of the linear ion trap[2] where the number of trapped ions was increased by more than 10 over conventional hyperbolic traps. The stability reached with the improved SNR of the LIT together with the high line-Q has led to a 10 fold improvement in clock stability over conventional ion based clocks. Indeed, the present LIT based clock shows stability competitive with the best H-masers for averaging times less than 10,000 seconds and exceeds H-maser stability beyond 10,000 seconds [3] making it the most stable of all clocks for long term stability. The LIT clock is, however, relatively recent in its development and its configuration continues to evolve. The purpose of this article is to propose an improvement in the architecture of the current physics package which will lead to improved long term stability, and a substantial reduction in size, mass and cost of the final frequency standard.

Current LIT Operation

The present form of the LIT physics unit for frequency standard operation is shown in Figure 1. Ions are created inside the linear trap by an electron pulse which ionizes a weak vapor of parent neutral atoms introduced into the vacuum system from a heated HgO isotopic source. They are held in the four bar linear trap shown inside the region enclosed by mag-

[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.

0-7803-1945-1/94 $3.00 © 1994 IEEE

netic shields. Before the stable atomic frequency of the trapped ions can be compared to the multiplied output of an LO or flywheel oscillator a population difference between the hyperfine levels of the ground state must be created. This is done via optical pumping with UV light from a laser or a discharge lamp. As shown in Fig. 1, optical pumping is carried out by a ^{202}Hg lamp whose output light is collected and focussed onto the central portion of the ion cloud.

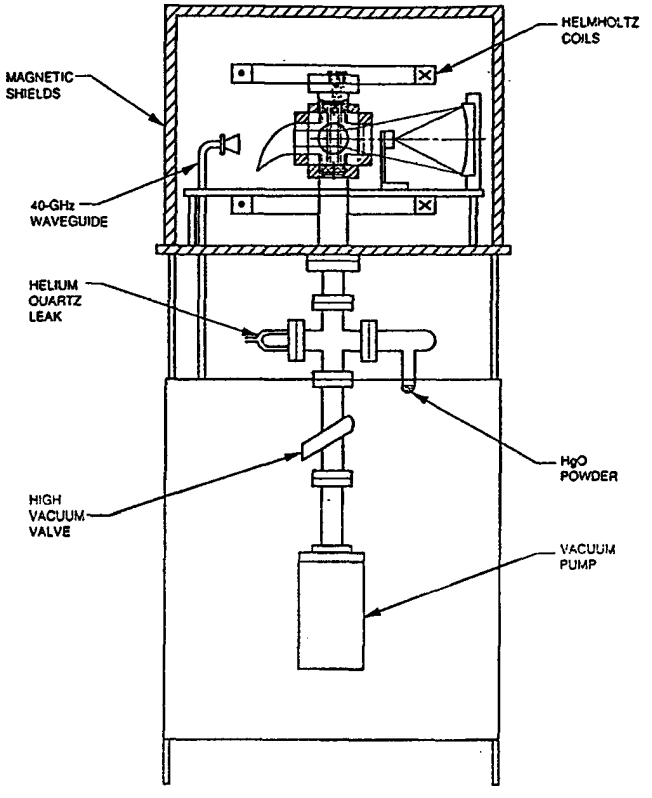


Figure 1: Present physics unit for the LIT based frequency standard. ^{199}Hg ions are state prepared and interrogated with the multiplied output of a local oscillator in the ion trap housed inside the magnetic shielded region. The apparatus is about 2 m high by 1 m on a side.

After state preparation, microwave radiation derived from the LO is passed onto the ions to make a frequency comparison with the stable atomic resonance. During this period it is crucial that the atomic resonance frequency not be perturbed by any changes in the trap environment. Such fluctuations will be transferred to the servoed LO thereby degrading clock stability. Because the state selecting light will shift the atomic clock resonance it is switched off during the microwave interrogation time. Following the microwave pulse, the lamp is turned on again to determine whether the microwave radiation has changed the population of the hyperfine levels of the atomic ion. Any frequency detuning of the LO from the reference atomic frequency will change the fluorescent light intensity measured when the lamp is turned on. These fluorescence changes are converted to a voltage and fed back to the frequency control port of the LO to keep it on frequency with the atomic oscillator.

It is apparent that the trap region is operated in two modes in the process of steering an LO. First, to prepare the atomic ions for microwave frequency comparison to the LO. In the second mode the atomic frequency is compared to the multiplied output of the LO. During the state preparation/interrogation mode there are no stringent requirements on environmental isolation or regulation whereas during the resonance mode great care must be exercised in the regulation of the atomic environment to insure stable operation.

Improved LIT Design

Figure 2 shows a configuration of an ion trap based clock where these modes of operation are carried out in two separate regions. The trap shown is a lengthened version of the the LIT of Fig. 1. In the new architecture of Fig. 2 the region labelled ion loading and fluorescence is somewhat smaller in size to the trap shown in Fig.1. A second linear trap has been added for a microwave resonance region.

A dc break forms a junction between these two regions to allow ions 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 negative 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 upper 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 upper region is dc biased positive with

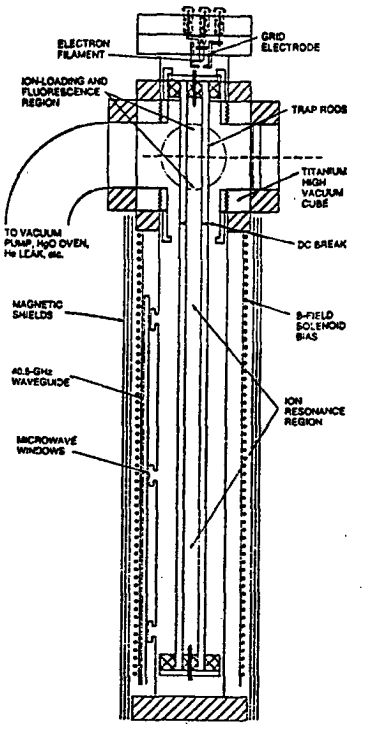


Figure 2: Improved physics unit for the trapped ion frequency standard.[199]Hg ions are created and state prepared in the upper region of the trap, then transported into the lower region, where the Hg^+ atomic resonance frequency is compared with the local oscillator frequency. The ions are then moved 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.

respect to the resonance region all ions will be transported from the fluorescence region to the ion resonance region. The helium buffer gas will provide the required energy damping with a 1/3 second time constant[4]. The dc potential along the length of the trap which controls which trap the ions occupy is shown in Fig. 3.

The separation of the clock resonance region from the optical/ion loading region relaxes many of the design constraints of the present units. The microwave resonance region can now be designed with no consideration of optical issues. A simple set of cylindrical shields and solenoid supply a very uniform and stable magnetic environment. In Figure 2 it is seen that the volume of the resonance region is now reduced by about 100 as the magnetic shields are 10 times smaller in diameter. Since the resonance region requires the most stringent temperature regulation of the entire physics package the size reduction enables simplification of thermal control of the standard.

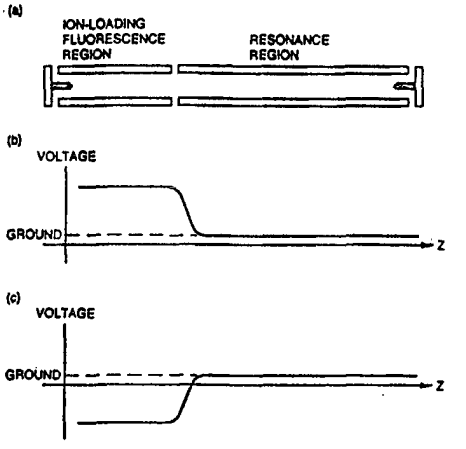


Figure 3: The improved LIT with separated state selection and microwave resonance regions: (a) the complete ion trap showing the dc junction; (b) a positive dc bias is applied to the loading and interrogation trap rods forcing the ions into the resonance region; and (c) ions are drawn back into the fluorescence region by applying a negative bias.

Similarly, the state selection optical design is now done with no worry of perturbing the atomic resonance. In practice, this means that the optical components are no longer required to be non-magnetic as in the previous design. The custom made Copper-Titanium flanged UV transmitting windows can be replaced with S-steel flanged windows which are each about 2 k$ less expensive. The 4 k$ non-magnetic UV light collecting horn is replaced by an S-steel elbow connecting the trap region to the vacuum pumps. The light source can be moved much closer to the ion cloud raising the possibility of a collimated UV source with no collection mirror thus further reducing the finalpackage size. Similar modifications could be made in the collection arms to reduce size. These design changes would not allow stable clock operation if the ions underwent microwave clock resonances inside the optical state selection region as in the previous design, because of the large magnetic shifts of the atomic levels induced by the close proximity of the light source and detectors. Another simplification gained by relaxing the non-magnetic requirement is in the electron gun assembly and its heater current supply. The electron source for ion creation is now remote from the atomic resonance region allowing use of the standard (magnetic) filament base and a floatable dc current drive. In the design of Fig. 1 ions are created inside the resonance region and require a floating audio frequency

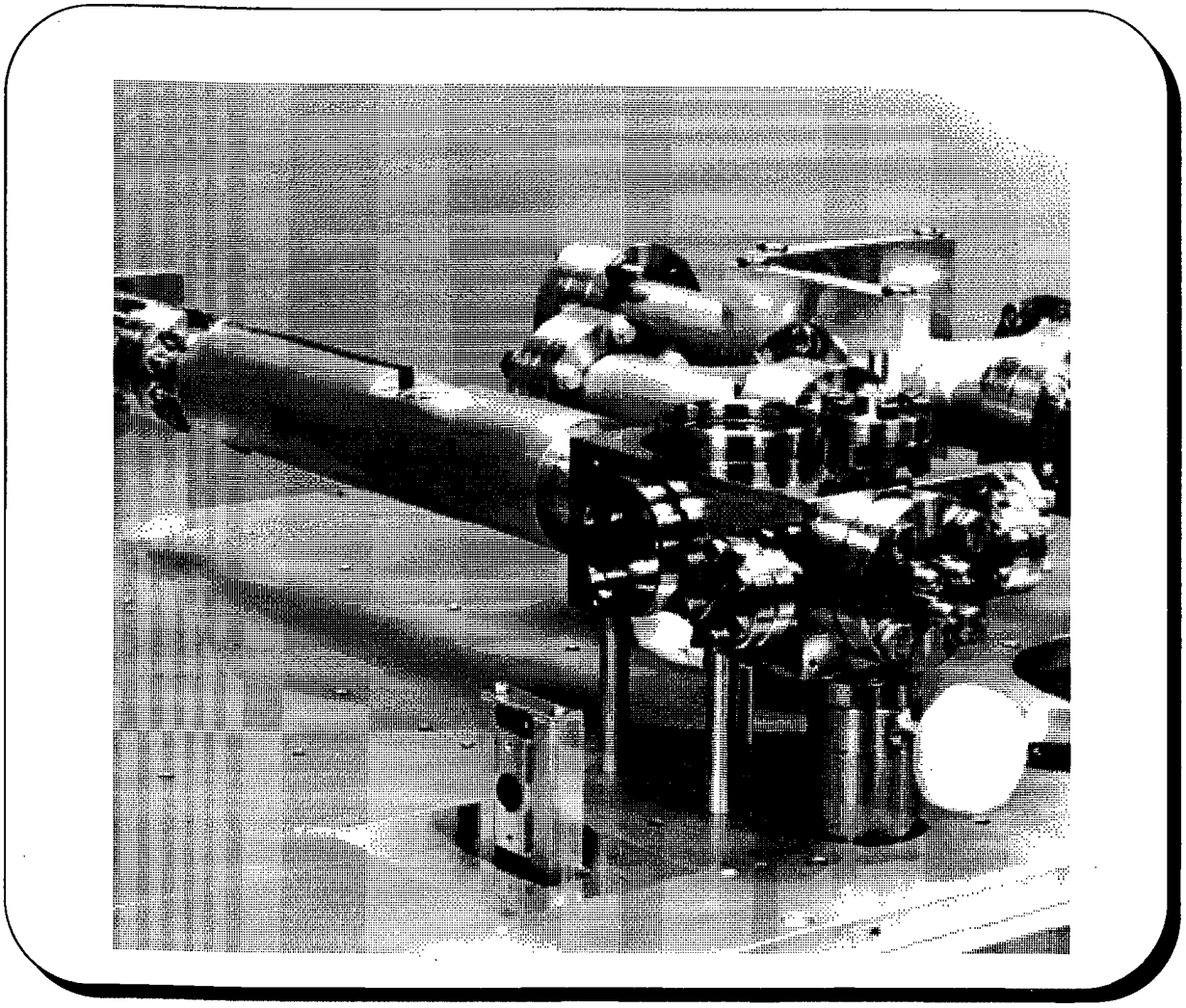


Figure 4: The new linear ion trap architecture with separated resonance and state selection regions. The resonance region is the 1 7/8" diameter by 12" long OFHC copper tube in the left side of the picture. The optical state selection region is in the right side foreground. In the background we see some of the support vacuum plumbing. The length of the Hg^+ trap and vacuum housing is 18" and is analagous to a Cesium beam tube in a Cesium standard. The magnetic shields and optical components are not installed in this photo.

filament heater current to prevent the residual dc magnetic field from the electron gun from shifting the atomic ion frequency.

Additionally, the largest frequency offset and potential instability can be reduced greatly by selecting a somewhat longer trapping length in the atomic resonance region. This frequency shift stems from the finite (i.e., non-zero) ion cloud diameter where ions spend time in regions of large rf trapping fields where their motion results in frequency pulling via the second order doppler or relativistic time dilation effect. The magnitude of this offset (for large clouds) depends only on the linear ion density, N/L, and is given by [5,6]

$$(\Delta f/f)_{\text{2nd Doppler}} = -(q^2/8\pi\epsilon_0 mc^2)N/L.$$

An increase in the resonance trapping length to about 200 mm would reduce the clock sensitivity to ion number variations by about a factor of 4 from that in the present system of Fig. 1. Since ion number stabilization to the 0.1% level has been demonstrated for 10^{-15} frequency standard operation in that system, an ion number induced instability noise floor below $2.5\text{x}10^{-16}$ should be achievable in this modified configuration.

Design Details

The system currently under development (Fig. 4) is housed in a vacuum chamber about 18" long. The resonance section of the vacuum can is cylindrical, 1 7/8"×12" long. It is fabricated from OFHC copper with a 1" bore along the axis to allow the 4 rod quadrupole trap to sit inside. The state selection region is a Stainless-Steel block with three UV transmitting fused silica windows welded into the window openings for optical pumping light entry and fluorescence collection. The UV light is passed into a 45° elbow vacuum component (which connects the trap region to the vacuum pump) to prevent backscattering of the UV optical pumping light beam into the fluorescence collection system. In the previous design this was accomplished with an expensive pyrex horn attached to a non-magnetic vacuum flange.

As shown in figure 2 the transition from the state selection to resonance region forces the magnetic shields to be open at one end. To obtain good shielding near the transition we need a small diameter for the vacuum can in the resonance region to enable small diameter cylindrical shields. The initial set of shields will be a triple layer set of 2.25" inside diameter with 3/4" spacing between shield layers.

The trap is inside the 1" diameter bore of the vacuum can, the four trap rod centers being located equally spaced on a 5/8" circle. The close proximity of the grounded vacuum container walls causes some shielding of the rf trapping fields from the interior of the trap. We have calculated the shielding effect of the nearby wall by the method of image charges [6]. If the vacuum wall is radius R_c and the trap rods are equally spaced on a circle of radius R_t, an image line charge of opposite sign located at $R_c(R_c/R_t)$ will maintain the vacuum wall at radius R_c at an equipotential. Figure 5 shows the equipotential surfaces for our configuration, $R_c/R_t = 1.6$. We find that the lowest order correction to the secular frequency due to the shielding effect of the grounded vacuum walls is a weakening of the secular frequency by the factor $[1 - (R_t/R_c)^4]$.

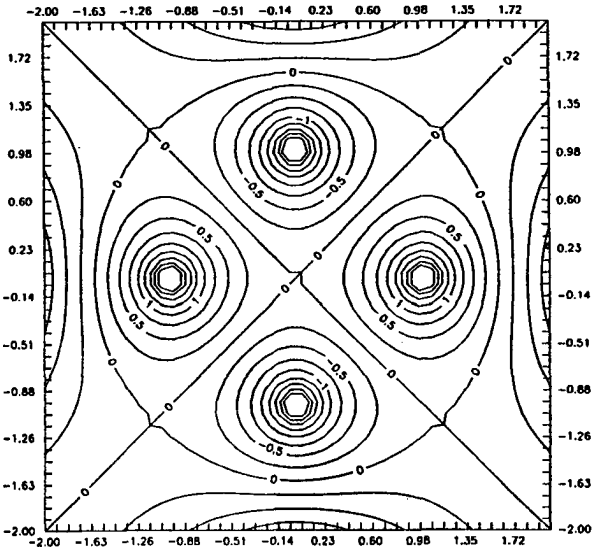


Figure 5: Equipotetial surfaces for a four rod linear ion trap, of radius R_t, inside a grounded cylinder at radius $R_c/R_t = 1.6$.

Conclusions

An improved architecture for a linear ion trap based frequency standard is described. Its design is a natural extension of the original LIT and separates the state selection and ion creation region from the atomic resonance region. Many of the design constraints of the present configuration are eliminated and a smaller, cheaper more stable frequency standard will result.

References

[1] J. D. Prestage, R. L. Tjoelker, G. J. Dick and L. Maleki, "Ultra-stable Hg^+ trapped ion frequency standard," Journal of Modern Optics, vol. 39, pp. 221-232, 1992.

[2] J. D. Prestage, R. L. Tjoelker, R. T. Wang, G. J. Dick, and L. Maleki, "Hg^+ Trapped Ion Standard Performance with the Superconducting Cavity Maser Oscillator as L.O.," Proc. 1992 IEEE Frequency Control Symposium, pp. 58-63, 1992.

[3] 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, June 1994.

[4] 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.

[5] 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.

[6] J. D. Prestage, R. L. Tjoelker, G. J. Dick, and L. Maleki, "Doppler Sideband Spectra for Ions in a Linear Trap," in Proceedings of the 1993 IEEE International Frequency Control Symposium, 1993, pp. 148-154.

[7] J. R. Reitz and F. J. Milford, Foundations of Electromagnetic Theory. Reading: Addison-Welsey, 1967, 2nd Edition, ch. 3, pp. 61-63.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PROGRESS TOWARD THE DEVELOPMENT OF A YTTERBIUM ION STANDARD*

D. J. Seidel† and L. Maleki

Jet Propulsion Laboratory, California Institute of Technology,
4800 Oak Grove Drive, Pasadena, California 91109

Abstract

A ytterbium ion standard is currently under development at JPL for both ground and space applications. We have chosen the ytterbium ion for its large hyperfine splitting of 12.6 GHz, and accessibility of its first excited electronic energy levels with light from frequency-doubled semiconductor and solid state lasers. Performing microwave-optical double resonance spectroscopy we have achieved 25 mHz Ramsey fringes with a signal-to-noise ratio better than 350:1 in the shot noise limit, which corresponds to an inferred stability on the order of 1.8 x $10^{-14}/\sqrt{\tau}$.

Introduction

A ytterbium ion standard has been under development at JPL for ground and space applications. The trapped mercury ion standard has achieved fractional frequency stability of 7 x $10^{-14}/\sqrt{\tau}$ for $20s \leq \tau \leq 10,000s$ and has made possible stabilities of 7 x 10^{-16} at 100,000 seconds of averaging time, a performance unsurpassed by any other frequency standard [1,2]. However ytterbium is also an attractive ion for frequency standards applications since preparation and observation of its clock transition at 12.5 GHz may be made utilizing compact semiconductor or solid state laser systems [3,4]. This feature enables designs methodologies that reduce the size and the mass of the standard without necessarily compromising its performance. It also allows laser cooling [5–8] of the ions to improve stability performance of the standard. Thus a small ion standard based on ytterbium could be ideal for spacecraft applications.

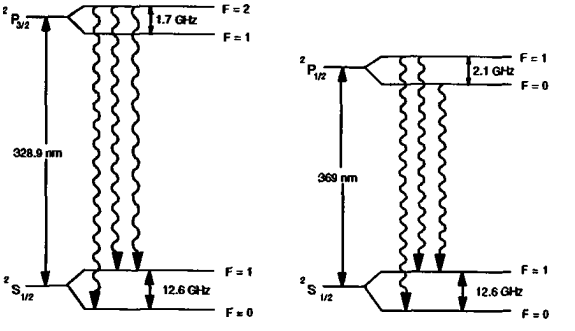


Figure 1: Hyperfine structure of the ground and first excited states in $^{171}Yb^+$.

In a previous paper [9] we had described results on the laser spectroscopy of ytterbium ions trapped in a linear rf/dc hybrid trap. In the present work we report on observation of Ramsey fringes obtained by excitation of the clock transition. Here again a laser was used to prepare and interrogate the ions. The resulting signal and the observed line Q of the ions allow us to infer stability of a standard based on this scheme. Since the 12.5 GHz ground state hyperfine transition of ytterbium is about four times more susceptible to magnetic field perturbations than the corresponding transition in the mercury ion, we also have performed an extensive study to characterize the magnetic sensitivity of ytterbium in our particular trap configuration. Despite its apparent advantages, ytterbium has difficulties with population trapping which effectively removes the ion from the cycling process [9–14]. We have also studied this problem and determined a means for effectively reducing population trapping influences. Finally, we will describe future plans to perform actual stability

*This work was carried out at the Jet Propulsion Laboratory, California Institute of technology, under a contract with the National Aeronautic and Space Administration.

†Also at Department of Physics, University of California - Riverside, Riverside, CA

0-7803-1945-1/94 $3.00 © 1994 IEEE

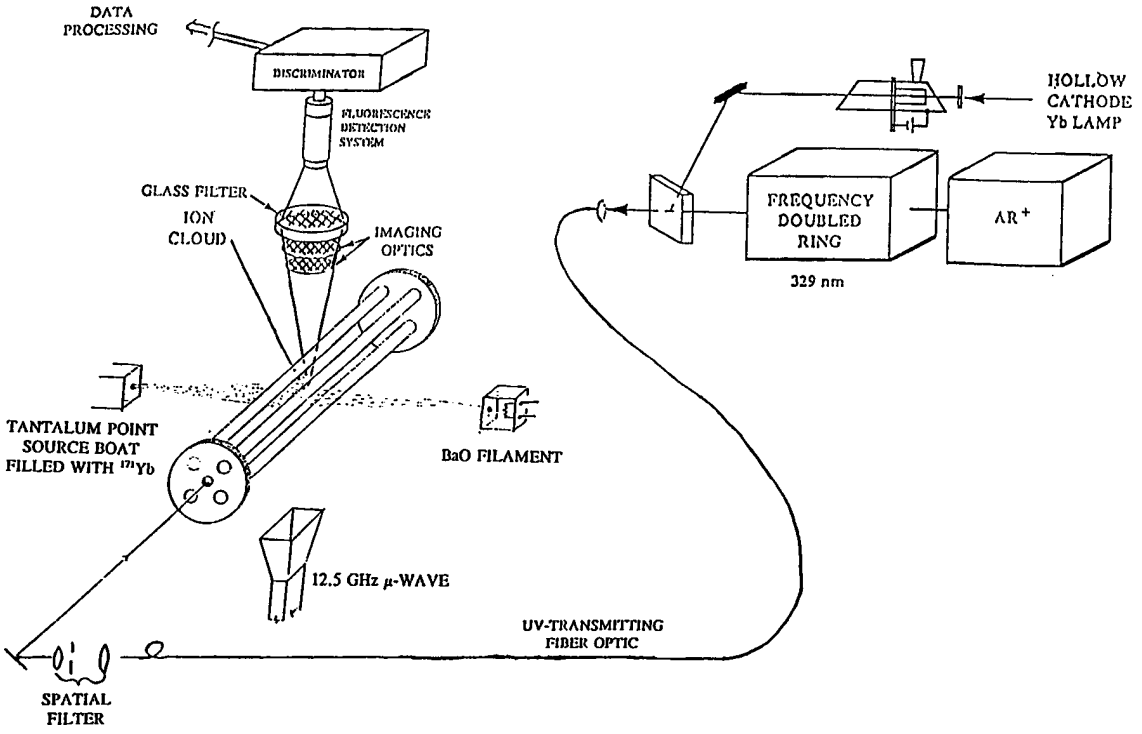


Figure 2: Simplified experimental setup used to carry out spectroscopy on trapped $^{171}Yb^{+}$ ions.

measurements with the ytterbium ion standard.

Experimental setup

The ytterbium isotope 171 ion has a hyperfine structure similar to the mercury isotope 199 ion, with an energy splitting corresponding to 12.64 GHz, as compared to mercury's 40.5 GHz. Its $^2P_{1/2}$ or $^2P_{3/2}$ excited states can be reached from the ground state with 369nm or 329nm radiation respectively (see fig 1). The use of lasers for optical pumping ytterbium ions readily leads to a ten fold improvement in the signal-to-noise ratio (SNR) over mercury's lamp based system, while only giving up a factor of 40.5/12.6 in line Q for equivalent interrogation times. Since the short term stability is given by

$$\sigma(\tau) \propto \frac{1}{Q_\ell}\frac{1}{SNR}\sqrt{\frac{t_i}{\tau}} \qquad (1)$$

(where Q_ℓ is the line Q $= f/\delta f$, and t_i is the total cycle time from one measurement to the next) there is the potential for a two to three fold improvement in the short term stability. In our experiment we were able to cycle the ions, a process that scatters many optical photons for each microwave photon absorbed. This was accomplished through excitation of the ($^2S_{1/2}$,F=1) $\Leftrightarrow$ ($^2P_{3/2}$,F=2) transition while only pumping in the doppler wings of the ($^2S_{1/2}$,F=1) $\Leftrightarrow$ ($^2P_{3/2}$,F=1) clearing transition. The cycling of the ($^2S_{1/2}$,F=1) $\Leftrightarrow$ ($^2P_{3/2}$,F=2) transition is possible due to the narrow linewidth ($\sim$ 1 MHz) of the laser and large 1.7 GHz hyperfine splitting of the $^2P_{3/2}$ state.

Figure (2) shows the experimental setup used to study the clock transition in trapped ytterbium ions. The hybrid linear rf/dc trap, which is modeled after the one first used with mercury ions at JPL [15], is composed of four rf electrodes in the form of rods with dimensions of 3.18 mm in diameter and 88.9 mm long. The rods are configured equally spaced with their center on a 17.48 mm diameter circle. Hollow cylindrical dc endcaps with inner radii of 1.19 mm and outer radii of 3.18 mm are placed on the axis of the quadrupole trap. The endcaps protrude into the trap 15.24 mm from the end, leaving a total trap length of 58.42 mm from endcap to endcap. The effective trap length, or length of the ion cloud, in the trap depends on the dc voltage applied to the endcaps. For our experiment an rf drive amplitude of

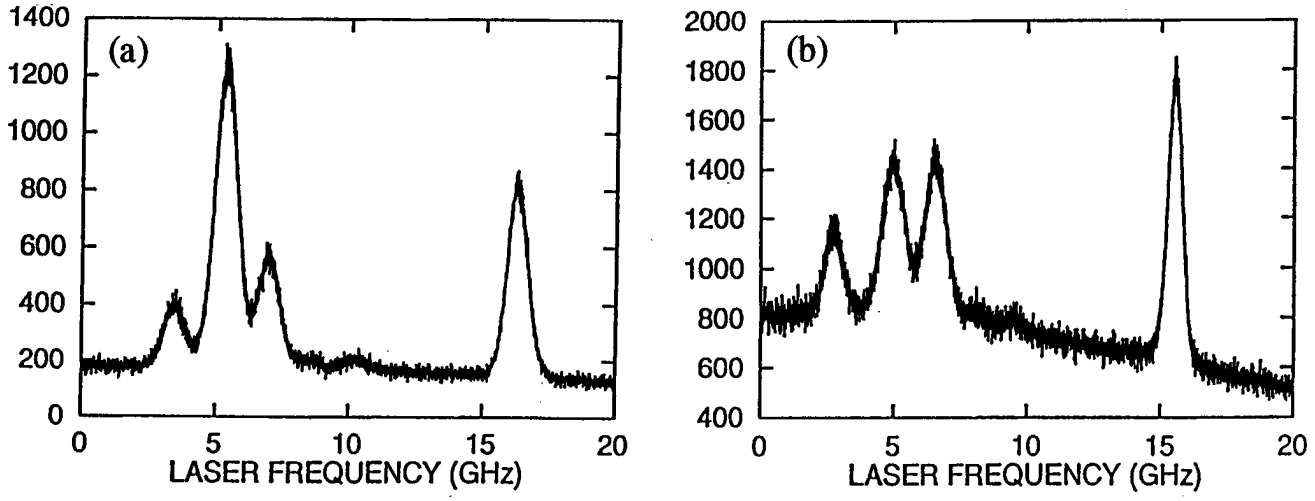


Figure 3: Resonance lineshape spectra vs. uv-laser power. (a) Laser power ~ 50 nW. (b) Laser power ~ 250 nW.

110 V_{rms} at 483 kHz was applied on the four rods, with +15 dc volts applied on the endcaps. To reduce the effect of rf heating on the ions a helium buffer gas was introduced through a quartz leak to collisionally cool the ions. In addition, nitrogen gas was introduced through a variable leak to collisionally quench the lower lying metastable states, which otherwise contain the ions in a "dark" state. The helium background in the trap was maintained at a nominal pressure of 1 x 10^{-5} torr while the nitrogen pressure was approximately 1 x 10^{-6} torr. We estimate that these parameters allow the trap to hold a 2 mm diameter cloud of 10^7 ions. The mixture of nitrogen and helium was quite effective in cooling the ion cloud and quenching the dark state.

The $4f^{14}(^1s)6s^2S_{1/2} \Leftrightarrow 4f^{14}(^1S)6p^2P_{3/2}$ transition was excited using 328.9 nm radiation generated by a coherent 699-21 ring dye laser running with DCM special laser dye and an intracavity lithium iodate doubling crystal. Up to 2 milliwatts of uv power was generated in this manner, and a homemade hollow cathode lamp was used to maintain the laser was on resonance. In order to move the trap away from the large magnetic fields associated with the argon ion laser and stainless steel optical table the uv-radiation was then injected into a 10 meter length of fiber optic cable and introduced axially into the trap.

The magnetic field inhomogeneities associated with the argon ion laser and steel optical table, as well as the ambient earth's field, are quit large, leaving a very poor environment for carrying out precision spectroscopy on the hyperfine structure of an atom. For this reason it was necessary to place the ion trap inside a three axis set of five foot diameter Helmholtz coils. With this set up we were able to cancel the ambient magnetic field in the room and set up a magnetic field environment stable enough to allow greater than 20 second coherence times.

Optical Resonance Lineshapes

Resonance fluorescence lineshapes at two different laser powers are shown in figure 4. Lower optical power improves the signal to noise by reducing the background scattered light and improving the photon counting efficiency by ensuring the counting system is running below saturation. Axial temperature measurements are also improved by reducing the artificial narrowing of the the resonance line by rapid pumping of the ions into an alternate level while the laser is scanning through the transition.

Microwave-Optical Double Resonance Spectroscopy

Ramsey fringes were obtained by first preparing the ions in the ($^2S_{1/2}$,F=0) state through optical pumping by tuning the laser to overlap the ($^2S_{1/2}$,F=1) $\Leftrightarrow$ ($^2P_{3/2}$,F=1) transition. With the laser off the ions were irradiated with a pair of 400 ms $\pi/2$ microwave pulses separated by a time t_i. If the microwave frequency matched the energy splitting of the ground state hyperfine levels the ions would be effectively transferred from the ($^2S_{1/2}$,F=0) back up to the ($^2S_{1/2}$,F=1) level. The laser which was tuned

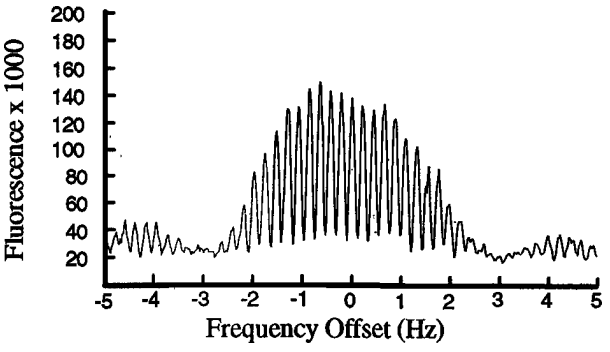


Figure 4: 100 mHz ramsey fringes corresponding to a 5 second separation between $\pi/2$ pulses.

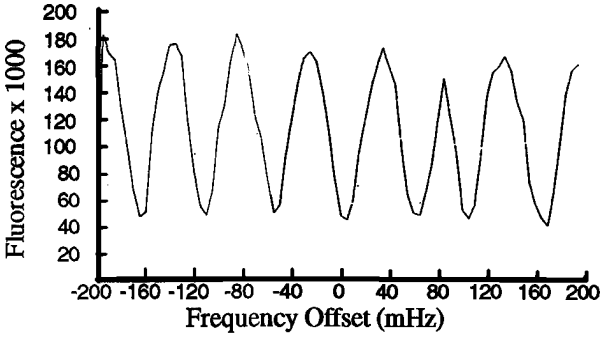


Figure 5: 25 mHz ramsey fringes corresponding to a 20 second separation between $\pi/2$ pulses.

to overlap the $(^2S_{1/2},F=1) \Leftrightarrow (^2P_{3/2},F=2)$ transition was then turned on and the ion fluorescence was monitored for two seconds. The level of fluorescence corresponds to the number of ions in the upper hyperfine state, which in turn depends on how close the microwave frequency matched the energy difference between the ground state hyperfine levels. Once the fluorescence was recorded the laser was then scanned over the $(^2S_{1/2},F=1) \Leftrightarrow (^2P_{3/2},F=1)$ transition to ensure all the ions were back in the $(^2S_{1/2},F=0)$ level and ready for the next pair of microwave pulses. Figures 4 and 5 show the effect of slowly stepping the microwave frequency after each pair of microwave pulses.

As shown in figure 5 we achieved 25 mHz Ramsey fringes with single shot photon counts of 130,000 above a 50,000 count background; this corresponds to a signal-to-noise ratio exceeding 350:1 in the shot noise limit, and inferred stability of $\sigma_y(\tau) \sim 1.8\ x\ 10^{-14}/\sqrt{(\tau)}$. It is clear from figure 5 that we are not yet realizing the full 350:1 SNR, probably because of magnetic field fluctuations. At the 100 mgauss magnetic field that we set up to separate the $\Delta M_F=0$ from the $\Delta M_F=\pm 1$ transitions the transition frequency sensitivity to magnetic fluctuations is 62mHz/mgauss. The relatively high sensitivity of the 12.64 GHz resonance transition to magnetic fluctuations coupled with the lack of magnetic shields in our present experimental system prohibits actual measurements of long term stability.

Zeeman and Motional Sideband Spectroscopy

The frequency of the $(^2S_{1/2},F=0,M_F=0) \Leftrightarrow (^2S_{1/2},F=1,M_F=0)$ transition in the presence of a magnetic field is given by

$$\nu = \nu_o + 311\text{Hz}\left(\frac{B}{1\text{ gauss}}\right)^2. \qquad (2)$$

Since we are interested in fractional frequency stability and this function has a quadratic nature it is desirable to work at values of applied magnetic field as low as possible since the sensitivity to magnetic fluctuations is given by

$$\left.\frac{\partial\nu}{\partial B}\right|_{B=B_o} = 622\left(\frac{B_o}{1\text{ gauss}}\right)\frac{\text{Hz}}{\text{gauss}}. \qquad (3)$$

With this in mind we decided to investigate the Zeeman and motional sidebands at various orientations of the applied static magnetic field (C-field) with respect to the microwave polarization and trap axis orientation. Similar work was carried out by Enders et. al. on ytterbium ions confined in a hyperbolic trap [16].

Figures 6 and 7 each show 26 possible magnetic field orientations that almost completely map out all possible orientations. They were obtained by tuning the laser to overlap the $(^2S_{1/2},F=1) \Leftrightarrow (^2P_{3/2},F=1)$ transition while scanning the microwaves 500 kHz in a phase continuous manner over a period of 10 seconds and summing together 50 scans. The total static magnetic field was kept constant at approximately 86 mgauss and the following static magnetic field orientations were setup $\mathbf{B}(x,y,z) =$ $(0,0,\pm1)$; $(0,\pm1,0)$; $(\pm1,0,0)$; $(0,\pm1,\pm1)$; $(\pm1,0,\pm1)$; $(\pm1,\pm1,0)$; $(\pm1,\pm1,\pm1)$. The trap axis is oriented along the x direction. Figure 6 shows the 26 static magnetic field orientations when the microwave polarization was oriented perpendicular to the trap

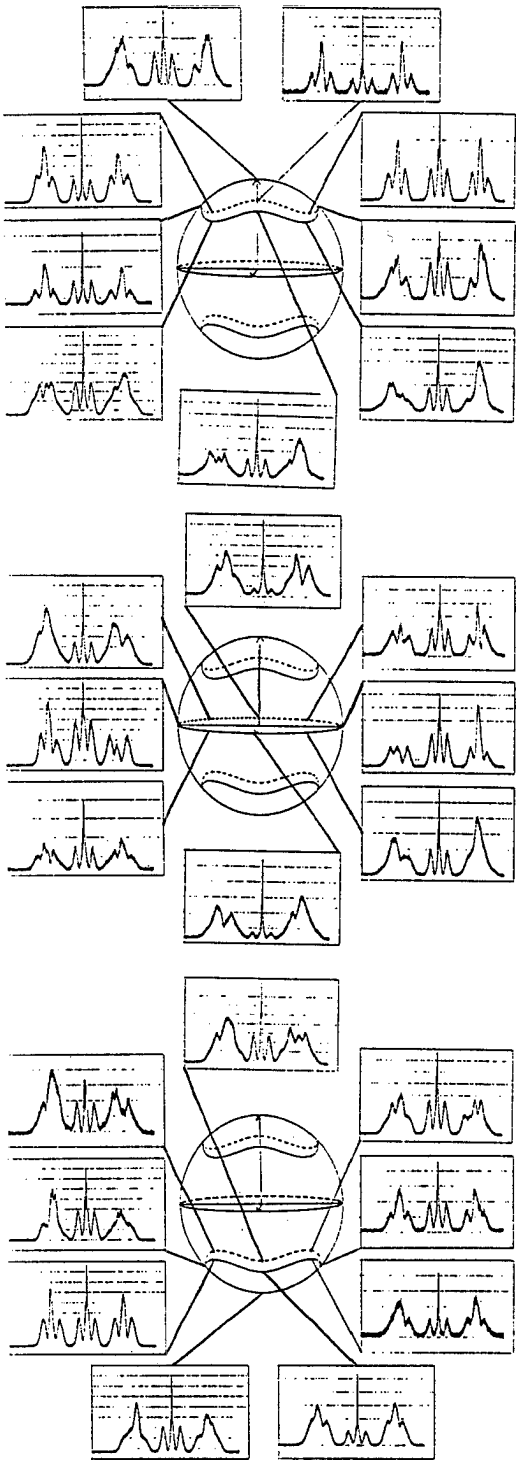


Figure 6: Zeeman and motional sideband spectra at 86 mgauss total magnetic field and microwave polarization parallel to the trap axis. The positions indicated on the spheres correspond to the 26 static C-field orientations $\mathbf{B}(x,y,z) = (0,0,\pm1); (0,\pm1,0); (\pm1,0,0); (0,\pm1,\pm1); (\pm1,0,\pm1); (\pm1,\pm1,0); (\pm1,\pm1,\pm1)$.

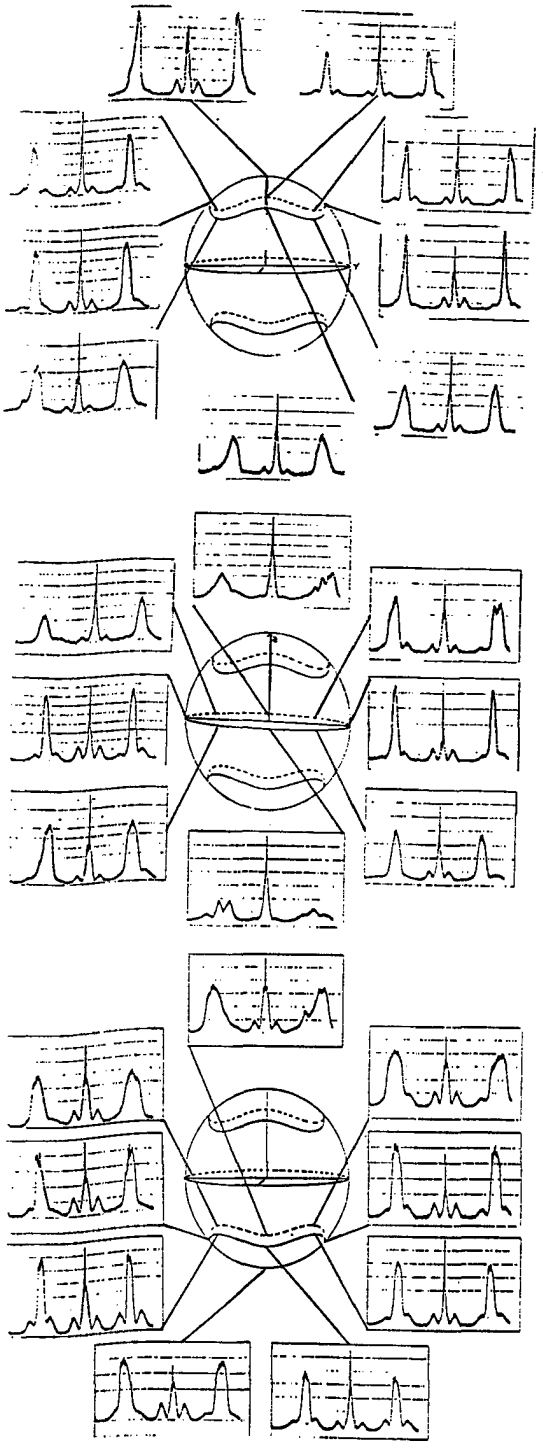


Figure 7: Zeeman and motional sideband spectra at 86 mgauss total magnetic field and microwave polarization perpendicular to the trap axis. The positions indicated on the spheres correspond to the 26 static C-field orientations [$\mathbf{B}(x,y,z) = (0,0,\pm1); (0,\pm1,0); (\pm1,0,0); (0,\pm1,\pm1); (\pm1,0,\pm1); (\pm1,\pm1,0); (\pm1,\pm1,\pm1)$].

axis, and figure 7 shows the same 26 magnetic field orientations with the microwave polarization oriented parallel to the trap axis. In both cases the the microwave propagation direction was perpendicular to the trap and oriented in the z-direction.

Conclusion

We have carried out extensive microwave-optical double resonance spectroscopy of trapped $^{171}Yb^{+}$ in order to determine the feasibility of ytterbium as the active atom in a trapped ion frequency standard. An inferred stability on the order of 1.8×10^{-14} has been obtained with 25 mHz Ramsey fringes. These results show great promise for ytterbium as a candidate ion in a frequency standard, we intend to perform actual stability measurements in the near future after the addition of magnetic shields, which are needed to reduce the effect of environmental magnetic fluctuations on the splitting of the 12.64 GHz microwave resonance transition.

A new ion trap and vacuum system is currently under design with emphasis on small size, ease of magnetic shielding, and high photon collection efficiency. The collection efficiency of the current system is only $\sim 0.1\%$ which can easily be improved ten to fifty fold. In addition, the current system is not suited for the addition of magnetic shields which are necessary to perform stability measurements.

While the design and fabrication of the next generation ytterbium system is taking place we will further study the influence of different buffer gases on the lifetime of ytterbium's meta-stable states. The construction of a stabilized compact 369 nm laser source will also be addressed so that ytterbium will be able to advance to a stand alone field frequency standard.

References

[1] R.L. Tjoelker, J.D. Prestage, G.J. Dick, and L. Maleki, "Long Term Stability of Hg^{+} Trapped Ion Frequency Standards," in *Proc. 47th Annual International Frequency Control Symposium*, IEEE Publication 93CH3244-1, 133, 1993.

[2] R.L. Tjoelker, J.D. Prestage, G.J. Dick, and L. Maleki, "Recent stability Comparisons with the JPL Linear Trapped Ion Frequency Standards," to be published in *Proc. 48th Annual IEEE International Frequency Control Symposium*, 1994.

[3] A. Williams, D.J. Seidel, and L. Maleki, "Generation of 369.4 nm radiation by efficient doubling of a diode Laser," in *OSA Proceedings on Advanced solid-State Lasers*, vol. 15, 250-252, 1993.

[4] Chr. Tamm, "A Tunable Light Source in the 370 nm Range Based on an Optically Stabilized Frequency-Doubled Semiconductor laser," *App. Phys. B*, vol. B56, no. 5, 295, May 1993.

[5] H. A. Klein, A. S. Bell, G. P. Barwood, and P. Gill, "Laser Cooling of Trappped Yb^{+}," *Appl. Phys. B*, vol. 50, pp. 13-17, 1990.

[6] H.A. Klein, A.S. Bell, G.P. Barwood, P. Gill, and W.R.C. Rowley, "Studies of Laser-Cooled Trapped Yb^{+}," *IEE Transaction on Instrumentation and Measurement*, vol. 40, no. 2, pp. 129-131, April 1991.

[7] A.S. Bell, P. Gill, H.A. Klein, A.P. Levick, Chr. Tamm, and D. Schnier, "Laser Cooling of Trapped Ytterbium Ions Using a Four-Level Optical-Excitation Scheme," *Phys. Rev. A*, vol. 44, no. 1, pp. R20-R23, 1991.

[8] 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.

[9] D.J. Seidel, A. Williams, R.W. Berends, and L. maleki, "The development of a Ytterbium Ion Frequency Standard," in *Proc. 46th Annual IEEE Frequency Control Symposium*, IEEE Publication 92CH3083-3, 70, 1992.

[10] H. Lehmitz, J. Hattendorf-Ledwoch, R. Blatt, and H. Harde, "Population Trapping in Excited Yb Ions," *Phys. Rev. Lett.*, vol. 62, pp. 2108-2111, May 1, 1989

[11] A. Bauch, D. Schnier, and Chr. Tamm, "Collisional Population Trapping and Optical Deexcitation of Ytterbium Ions in a Radiofrequency Trap," *J. Mod. Opt.*, vol. 39, no. 2, pp. 389-401, Feb. 1992.

[12] Chr. Tamm and D. Schnier, " A Tunable Three-Level Neodymium-Doped Fiber Laser and its application for Depletion of the $4f^{14}5d^{2}D_{3/2}$ Level in Optically-Excited, Trapped Ytterbium Ions," *Opt. Comm.*, vol. 87, no. 5/6, pp. 240-244, 1992.

[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 a linear Paul trap," in *Proc. 47th Annual IEEE International Frequency Control Symposium*, IEEE Publication 93CH3244-1, 139, 1993.

[14] J. Yoda and K. Sugiyama, "Disappearance of Trapped Yb^{+} Ions by Irradiation of the Resonance Radiation," *J. Mod. Opt.*, vol. 39, no. 2, pp. 403-409, Feb. 1992.

[15] J.D. Prestage, G.J. Dick, and L. Maleki, "New Ion Trap for frequency Standard Applications," *J. Appl. Phys.*, vol. 66, no. 3, pp. 1013-1017, August 1, 1989.

[16] V. Enders, Ph. Courteille, W. Neuhauser and R. Blatt, "Motional sidebands in the microwave spectra of ions in an rf trap," *J. Mod. Opt.*, vol. 39, no. 2, pp. 325-334, February 1992.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SECOND HARMONIC LEVEL MONITORS IN CESIUM ATOMIC FREQUENCY STANDARDS

B. Jaduszliwer and J. P. Hurrell
Electronics Technology Center, The Aerospace Corporation
P.O. Box 92957, Los Angeles, CA 90009

The Cesium Frequency Standards (CFS) to be flown on board GPS Block IIR satellites will carry a Second Harmonic Level Monitor (SHLM), the output of which will be the only telemetry datum available on the CFS health. The SHLM replaces the Beam Current Monitor (BCM) currently operating in GPS CFS. We have carried out an analysis of the validity and reliability of the SHLM as a CFS health indicator. The results of a rigorous calculation of the Cesium Beam Tube (CBT) output signal under the square-wave microwave frequency modulation scheme to be used in Block IIR CFS show that almost any process causing the output of the BCM to decrease will lead to a proportionate decrease in the SHLM output, but the SHLM output will be more sensitive to changes in some CBT operating parameters than the BCM output. In particular, that is the case for changes in the ionizer temperature and microwave power. We also show that the signal-to-noise ratio at the SHLM will be less than that at the BCM.

Introduction

The Cesium Frequency Standards (CFS) to be flown on board GPS Block IIR satellites will carry a Second Harmonic Level Monitor (SHLM). The output of this monitor will be the only telemetry datum available on the status and health of the CFS physics package, in contrast with the Block IIa satellites, in which a number of operating CFS physics package parameters can be monitored from the ground. A SHLM is incorporated in some commercial compact CFS as one among many monitor functions, but there is a general lack of data in regard to the validity and reliability of the SHLM as a CFS health and status indicator.

The physics package of a CFS, usually referred to as the cesium beam tube (CBT), is a complex device, whose purpose is to allow the interrogation of individual cesium atoms in an atomic beam by externally generated microwaves; the corresponding intelligence is conveyed by the cesium beam signal. In order to lock the local oscillator frequency to the cesium atomic hyperfine transition ("clock transition") used as frequency standard, the external microwaves are phase- (or frequency-) modulated. Frequency-lock of the CFS is achieved by a feedback loop incorporating phase-sensitive detection of the beam signal; under lock conditions, the Fourier component of the beam signal at the modulation frequency is nulled. The mean value (DC level) of the cesium beam signal, averaged over many microwave modulation cycles, is usually known as the "beam current". The output of the Beam Current Monitor (BCM) is an important health and status indicator in GPS Block I and IIa CFS, as well as in commercial units. We have recently analyzed the different processes which may lead to secular variations in a CFS beam current [1], in order to enhance the use of the BCM as a diagnostics tool.

The SHLM monitors the level of the Fourier component of the cesium beam signal at twice the microwave modulation frequency. In order to enable the use of the SHLM as a health and status indicator, it is necessary to determine the sensitivity of its output to changes in CBT operating parameters. We have derived the expected amplitude of the second harmonic component of the cesium beam signal under the microwave modulation technique used in Block IIR CFS, and analyzed the dependency of the SHLM output on some CBT variables.

Beam signal second harmonic amplitude

The cesium atoms in the beam are interrogated by the microwaves in a cavity of the type first described by Ramsey [2] and shown schematically in Fig. 1. Two

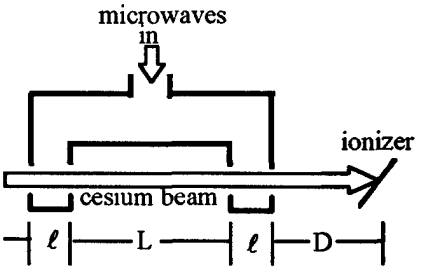


Figure 1: Schematical representation of the Ramsey-type microwave cavity

interaction regions of length ℓ are separated by a field-free region of length L. We assume that the atomic beam is well collimated, i.e., the atomic velocities are given by $\mathbf{V} \cong v\mathbf{i}$, where i is a unit vector along the axis of the CBT. Atoms spend a time $\tau = \ell/v$ in each of the interaction regions, and $T = L/v$ in the central field-free region. The microwave field mean angular frequency is ϖ, and in the GPS Block IIR CFS the microwave field is square-wave frequency-modulated, with modulation amplitude ω_m and modulation period T_m or, equivalently, triangular-wave

0-7803-1945-1/94 $3.00 © 1994 IEEE

phase-modulated with amplitude $\phi_0 = \omega_m T_m/4$ and the same period, as shown in Fig. 2 (a).

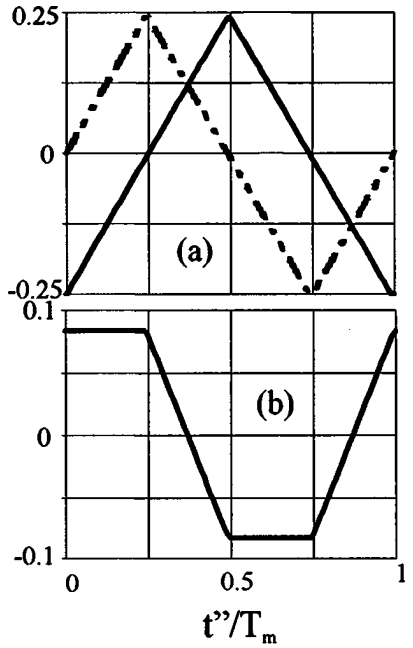


Figure 2: (a) Microwave field phase seen by an atom in each one of the arms of the Ramsey cavity under square-wave frequency modulation conditions. The atom exits the cavity at time t". Full line: ϕ_1 (phase in the first arm); dashed line: ϕ_2 (phase in the second arm). (b) Phase difference $\Delta\Phi = \phi_2 - \phi_1$ between both cavity arms, as seen by a traveling atom. For both (a) and (b) phases have been normalized by $\omega_m T_m$.

In the Appendix to this memorandum we derive from first principles the probability $P(\tau)$ that an atom that spends a time τ in each interaction region will undergo a hyperfine state transition while traveling through the Ramsey cavity, and use that result to calculate the time-dependent cesium beam signal. To show conceptually what to expect, we will for now assume that $\tau << T$, following Vanier and Audoin [3], and neglect the probability that the frequency may change while the atom is traveling within one of the interaction regions; we will consider only frequency changes which take place while the atom is traveling through the field-free region. If b is the Rabi frequency of the atoms in the cavity (determined by the amplitude of the magnetic component of the microwave field), then for small detunings from the clock transition angular frequency ω_0, $\Omega = \omega - \omega_0 << b$, the probability of transition $P(\tau)$ is given by

$$P(\tau) \cong \tfrac{1}{2} \sin^2 b\tau\left[1+\cos(\Omega T+\Delta\Phi)\right] , \qquad (1)$$

where $\Delta\Phi$ is the phase difference of the microwave field in both cavities, as seen by the traveling atom. When the microwave field is modulated, $\Delta\Phi$ will depend explicitly on time, and thus $P(\tau)$ will become time-dependent. If the cesium atoms in the beam have a velocity distribution g(v) (normalized to unit area), then the normalized distribution of travel times in the microwave interaction regions will be given by $f(\tau) = (l/\tau^2)\, g(\ell/\tau)$. The time-dependent cesium atomic beam intensity can be obtained by folding $P(\tau)$ with $f(\tau)$:

$$I(t') \cong \frac{I_0}{2}\int_0^\infty f(\tau)\sin^2 b\tau\left\{1+\cos\left[\Omega T+\Delta\Phi(t)\right]\right\}d\tau, \qquad (2)$$

where $t' = t + (L+D)\,\tau/\ell$ and D is the distance between the second interaction region and the detector. An atom of speed v needs to enter the first interaction region at time t in order to arrive at the detector at time t'. I_0 is the atomic beam intensity that would be measured if the beam were monovelocity; it is adjusted so that

$$I_p = I_0 \int_0^\infty f(\tau)\sin^2 b\tau\, d\tau . \qquad (3)$$

In the absence of microwave modulation; I_p is the cesium atomic beam intensity obtained under those conditions for $\varpi = \omega_0$.

The time dependence of the cesium beam signal is determined to a large extent by $\Delta\Phi(t)$, shown in Fig. 2(b) for triangular-wave phase modulation. Using these results, I(t) can be derived graphically for a monovelocity cesium atomic beam, as shown by Vanier and Audoin [3]. The results are shown in Fig. 3. Signal peaks appear with half the periodicity of the modulation, indicating that the

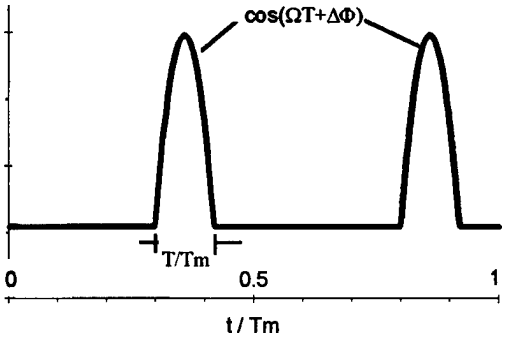


Figure 3: Result of graphical derivation of the beam signal I(t) for a monovelocity cesium atomic beam under triangular-phase wave microwave modulation. The beam signal is proportional to $\cos(\Omega T+\Delta\Phi)$.

first significant Fourier component of the beam signal (under frequency-lock conditions) is the second harmonic of the modulation frequency. In the Appendix we show that Eq. (2) is valid even if we relax the restriction $\tau << T$, provided that the CFS is frequency locked (i.e., $\varpi = \omega_0$).

<u>Delayed ionization effects on the cesium beam signal</u>

After traveling the length of the CBT, the cesium atoms impinge on a hot metal wire, where they are surface-ionized. The resulting ions are accelerated towards the first dynode of an electron multiplier (EM), which amplifies the collected ion current by a factor of about 10^4. The output of the EM is the cesium beam signal. Since characteristic times for surface ionization can be comparable to the

modulation time-scale, delayed ionization effects may play a significant role in determining the time dependency of the beam signal. This has been analyzed in the past [4], and it can be shown that, if I(t') is the cesium atom flux incident on the ionizer at time t', the emitted cesium ion current $I_+(t)$ is given by

$$I_+(t) = \gamma\, e^{-\gamma t} \int_{-\infty}^{t} e^{\gamma t'} I(t')\,dt', \qquad (4)$$

where γ is the cesium ion desorption rate off the ionizer wire surface; it depends on the ionizer material, surface condition and temperature. Hughes and Levinstein [5] have shown that, in general, the temperature dependence of γ can be expressed as

$$\gamma = \gamma_\infty\, e^{-Q/k\Theta}, \qquad (5)$$

where Q is the ionic heat of desorption, k is Boltzmann's constant and Θ the ionizer absolute temperature. Both Q and γ_∞ are determined by the ionizer material and surface condition. Since ion collection and electron multiplication are very fast processes compared to the microwave modulation period, they do not contribute significantly to the time dependence of the beam signal, which is then determined by Eq. (4), using I(t') as given by Eq. (2).

The time-dependent beam signal

Sequential use of Eqs. (2) and (4) allows the calculation of the cesium beam signal as a function of time when square-wave frequency modulation is used in a frequency-locked CFS. We developed computer code to calculate $I_+(t)$. We assume values for L, ℓ and D which are representative of those found in compact CBTs. The modulation frequency is $1/T_m$ = 100 Hz, and the frequency modulation amplitude is $\omega_m/2\pi$ = 200 Hz. The mean ion dwell time on the ionizer surface, $\tau_+ = 1/\gamma$, is a program input, as are the parameters required to characterize the atomic speed distribution. The program computes the detected atomic flux, I(t), and the ion current, $I_+(t)$; the time-dependent beam signal is proportional to $I_+(t)$. The program also calculates the output of the BCM, proportional to $\langle I_+ \rangle$,

$$\langle I_+ \rangle = \frac{1}{T_m} \int_0^{T_m} I_+(t)\,dt, \qquad (6)$$

and the output $\langle |i_+| \rangle$ of the SHLM, modeled as a phase-sensitive detector operating at twice the microwave modulation frequency:

$$\langle |i_+| \rangle = \frac{1}{T_m} \int_0^{T_m} i_+(t)\, S(t - t_0)\,dt, \qquad (7)$$

where $i_+(t)$ is the AC component of the beam signal, $i_+(t) = I_+(t) - \langle I_+ \rangle$. S(t') is the square wave reference at the second harmonic of the microwave modulation frequency:

$$S(t') = 1 \text{ for } 0 < t' \le T_m/4 \text{ and } T_m/2 < t' \le 3T_m/4 \qquad (8a)$$

$$S(t') = -1 \text{ for } T_m/4 < t' \le T_m/2 \text{ and } 3T_m/4 < t' \le T_m \qquad (8b)$$

The phase parameter t_0 is adjusted to maximize $\langle |i_+| \rangle$. Effects of finite preamplifier bandwidth have been neglected.

Figure 4 shows I(t) and $I_+(t)$, calculated using a 100 m/s monovelocity cesium atomic beam. The atomic flux at the ionizer, I(t), shows the same time dependence shown in Fig. 5, while the ion current $I_+(t)$ displays the effects of a 0.5 ms ionizer dwell time.

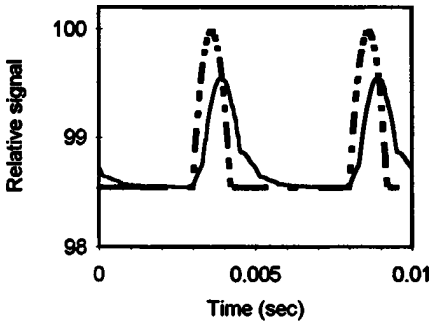


Figure 4: Calculation of CBT waveforms for a monovelocity beam. Dashed line: atomic flux at the ionizer, fully consistent with the one obtained graphically in Fig. 3. Full line: cesium beam signal (CBT output).

Evaluation of Eq. (2) in a more general case requires knowledge of the distribution of atomic speeds in the cesium beam. Measurements of speed distributions performed at Aerospace and elsewhere have shown substantial variability from one CBT to another, even when produced by the same manufacturer. This variability can be traced to normal fabrication tolerances when aligning the components of the CBT, as well as state-selecting magnets strength. A simple, 3-parameter empirical model of the speed distribution represents the full range of variation:

$$g(v) = A e^{-(v-v_0)^2/2\sigma^2} \quad v \le v_0$$
$$g(v) = B\left[e^{-(v-v_0)^2/2\sigma^2} + \xi (v/\alpha) e^{-(v/\alpha)^2} \right] \quad v \ge v_0 \qquad (9)$$

where the normalization constants A, B are related by

$$A = B\left[1 + \xi (v_0/\alpha) e^{-(v_0/\alpha)^2} \right]. \qquad (10)$$

B is determined by numerical integration in order to normalize g(v) to unit area and α is the most probable velocity in a maxwellian atomic beam at the cesium oven source temperature. The free parameters are v_0 (distribution peak), σ (width parameter) and ξ (high-speed distribution tail parameter). An example for g(v) is shown in Fig. 5 with v_0 = 110 m/s, σ = 15 m/s and ξ = 0.2; Fig. 6 shows the results obtained for I(t) and $I_+(t)$ using this speed distribution.

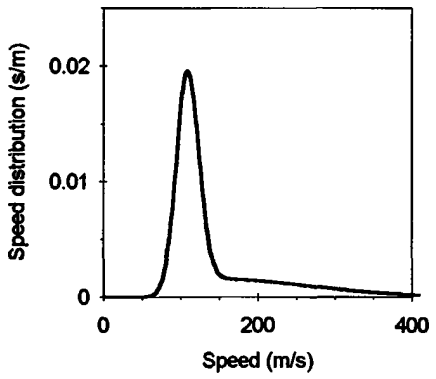


Figure 5: Three-parameter atomic beam speed distribution. $v_0 = 110$ m/s, $\sigma = 15$ m/s, $\xi = 0.2$

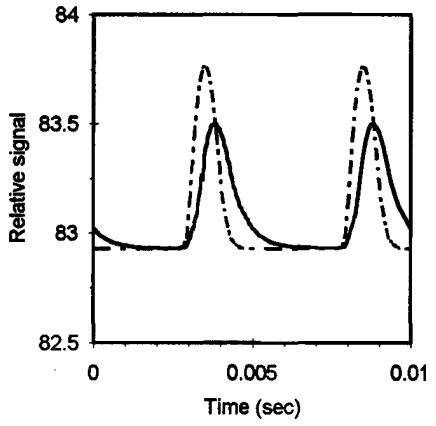


Figure 6: Calculation of CBT waveforms for a cesium atomic beam having the velocity distribution shown in Fig. 5. Dashed line: atomic flux at the ionizer. Full line: cesium beam signal (CBT output).

Second harmonic level monitor sensitivities

The SHLM output will be sensitive to I_0, as indicated by Eqs. (2), (4) and (7). Furthermore, changes in detection efficiency or EM gain will affect $\langle I_+ \rangle$ and $\langle |i_+| \rangle$ equally. Thus, almost any of the processes described in [1] leading to a reduction of the BCM output will also lead to a proportionate reduction of the SHLM output. On the other hand, as we will show, there are processes for which the SHLM sensitivity may be much higher than that of the BCM. We have examined two such processes in more detail: changes in the ionizer temperature and in the injected microwave power.

Sensitivity of the SHLM to ionizer temperature changes

If the atomic speed distribution is centered at v_0 and has a width $2\Delta v$, then the typical difference ΔT in the ionizer arrival times of a "slow" and a "fast" atom which entered the microwave cavity simultaneously will be

$$\Delta T \cong \frac{L+D}{v_0 - \Delta v} - \frac{L+D}{v_0 + \Delta v} \cong \frac{2(L+D)\Delta v}{v_0^2}, \qquad (11)$$

where we assume $\Delta v << v_0$. Both time-of-flight differences and delayed ionization effects will smear the temporal structure of $I_+(t)$ and thus reduce the SHLM output, but for mean ionizer dwell times $1/\gamma = \tau_+ << \Delta T$, the contribution from delayed ionization will be negligible. For $\tau_+ \geq \Delta T$, the SHLM output will become quite sensitive to changes in τ_+. In our computer program, we take L+D = 0.29 m; for $v_0 = 110$ m/s and $\Delta v = 15$m/s, we obtain $\Delta T \cong 7.2 \times 10^{-4}$s. Thus, we expect that for $\tau_+ \geq 700$ μs, the second harmonic signal level $\langle |i_+| \rangle$ will drop sharply with increasing τ_+.

Figure 7 shows the results obtained for the second harmonic signal level for increasing values of ionizer dwell time τ_+. The speed distribution parameters were $v_0 = 110$ m/s, $\sigma = 15$ m/s and $\xi = 0.2$; these parameters are consistent with the mean speed and characteristic width discussed above, and $\langle |i_+| \rangle$ indeed drops sharply as τ_+ increases beyond 700 μs. The ionizer dwell time $\tau_+ = 1/\gamma$

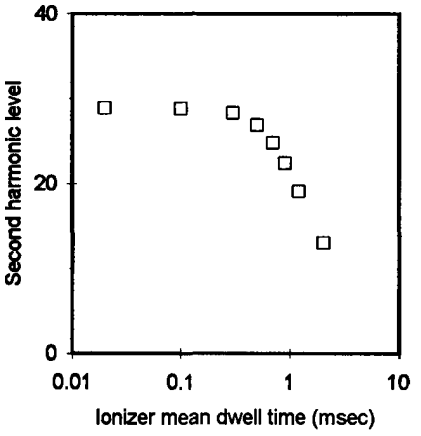


Figure 7: SHLM output vs. ionizer dwell time. Data for this figure was calculated using the speed distribution shown in Fig. 5

is determined by the ionizer temperature, as shown by Eq. (5). In actual CBT operation, the ionizer temperature, in turn, is determined by the ionizer current. Equation (5) can be used to model the dependence of the second harmonic signal level on ionizer temperature Θ:

$$\Theta = \frac{Q}{k \ln(\tau_+ / \tau_\infty)}, \qquad (12)$$

where $\tau_\infty = 1/\gamma_\infty$. Both Q and τ_∞ depend on ionizer material and surface conditions. For instance, for cesium atoms incident on a tungsten ionizer [6], $Q \cong 1.84$ eV and $\tau_\infty \cong 7.94 \times 10^{-12}$ s. An additional contribution to the ionizer temperature dependence of the SHLM output is introduced by the temperature dependence of the surface ionization efficiency ε, given by the Saha-Langmuir equation [7]:

$$\varepsilon(\Theta) = \left\{1 + (g_0/g_+)\, e^{-(W-I)/k\Theta}\right\}^{-1}. \qquad (13)$$

$\varepsilon(\Theta)$ can then be evaluated for cesium atoms (ionization energy I ≅ 3.87 eV, statistical weight factor $g_0/g_+ = 2$) incident on a tungsten wire ionizer (work function W ≅ 4.54 eV).

As $L_+(t)$ will be proportional to $\varepsilon(\Theta)$, both $\langle I_+ \rangle$ and $\langle |i_+| \rangle$ will be affected equally by the temperature dependence of the ionization efficiency, but only $\langle |i_+| \rangle$ will express the temperature dependence of τ_+. Figure 8 shows how the BCM and SHLM output signals depend on ionizer temperature. The speed distribution parameters are the same ones used for Fig. 5, and a tungsten ionizer is assumed. Above 900° C, neither the BCM nor the SHLM output change significantly with ionizer temperature.

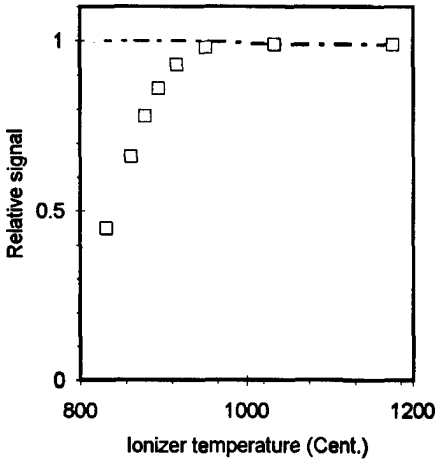


Figure 8: Dependency of the outputs of the SHLM (squares) and BCM (dashed line) on ionizer temperature. These data were calculated assuming a tungsten ionizer, and the atomic beam speed distribution shown in Fig. 5

Below 900° C, the BCM output remains essentially constant, (at least, as long as the ionizer is hot enough to enable the cesium ions to escape) while the SHLM output becomes very sensitive to temperature changes. If we define the critical ionizer temperature Θ_0 as that for which the mean ion dwell time τ_+ equals the spread ΔT in atomic travel times, then below Θ_0 we expect the output of the SHLM to change sharply with temperature and the output of the BCM to be approximately constant; above Θ_0 both monitor outputs will be insensitive to temperature changes. Changing the parameters of the atomic beam speed distribution or the ionizer material will change Θ_0, but not the overall pattern of temperature dependence of the BCM and SHLM outputs.

Sensitivity of the SHLM to microwave power variations

The probability $P(\tau)$ that an atom of speed $v = \ell/\tau$ will undergo a hyperfine state transition while traveling through the Ramsey cavity is given by Eq. (1); the maximum value it may take is $\sin^2 b\tau$. If $b\tau = \pi/2$, $P(\tau) = 1$, i.e., the atom will certainly undergo a hyperfine transition. The Rabi frequency b is determined by the microwave magnetic field amplitude B_0 within the Ramsey cavity, which, in turn, is determined by the injected microwave power. In a monovelocity beam, it is possible to set the microwave power so that $P(\tau) = 1$ for all atoms in the beam. In an actual beam having a speed distribution with a finite spread, it is impossible to satisfy $b\tau = \pi/2$ for all atoms, and so the microwave power is optimized by maximizing some desirable property of the signal (e.g., the slope of the feedback loop error signal at zero detuning). Microwave power variations will affect $P(\tau)$ for every atom in the beam, and both $\langle I_+ \rangle$ and $\langle |i_+| \rangle$ can be expected to change in response to those variations. Figure 9 shows how the SHLM and BCM outputs depend on microwave power for a particular choice of speed distribution (as before, $v_0 = 110$ m/s, $\sigma = 15$ m/s and $\xi = 0.2$), and $\tau_+ = 500$ μs.

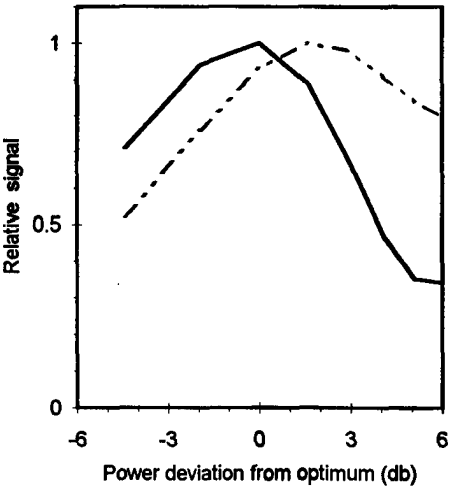


Figure 9: Dependency of the SHLM (full line) and the BCM (dashed line) on microwave power. Power is "optimized" for the peak of the speed distribution. The speed distribution shown in Fig. 5 was used in the calculations.

In these calculations, the power was arbitrarily optimized for the peak of the speed distribution (i.e., $b\ell/v_0 = \pi/2$). It is evident from the figure that the outputs from both monitors are not proportional to each other; the SHLM output is at a maximum for the power optimized as described above, but the largest BCM output is obtained at a slightly higher power level. Additionally, the SHLM peak is somewhat narrower than the BCM peak, indicating a globally higher sensitivity to microwave power.

Discussion

We have used a realistic, fairly detailed model of the interaction of the cesium atomic beam and the microwave field within the Ramsey cavity to develop a powerful analysis and diagnostics tool which allows us to evaluate and understand better the capabilities of the SHLM as a CFS status and health indicator. Since we calculate both the SHLM and the BCM outputs, we can compare the responses of both monitors to different processes affecting the operation of a CBT. The model underlying our calculations assumes that the CFS feedback loop is locked, i.e., it can not describe the dynamics of the search for frequency lock or the behavior of the CFS under failed lock conditions.

A first conclusion is that almost any of the processes which may lead to reductions in the CBT beam current will also lead to proportionate reductions in the SHLM output. The main exception is (besides the trivial one of changes in the gain of non-common signal-processing electronics) the difference in sensitivities of both monitors to microwave power variations, analyzed in detail in the present memorandum. In general, a CFS which would display a gradual decay of the BCM output will display a gradual decay of the SHLM output.

Since the output of the SHLM depends directly on the time-dependent part of the cesium beam signal, any process which modifies the time dependency of the beam signal, even if it does not change its mean value, will also lead to changes in the SHLM output. Under these conditions, differences in the behavior of both monitor outputs may be expected. Processes within the CBT which could change the cesium atomic beam speed distribution (e.g., changes in the magnetization of the state-selecting magnets) fall within this category, although they are not very probable. The most likely situation in which these differences could be observed is one in which the ionizer current varies, leading to changes in the ionizer temperature. In this work we show that below a critical temperature determined by the ionizer material and the width of the atomic beam speed distribution, the output of the SHLM becomes much more sensitive to ionizer temperature than the output of the BCM.

An additional issue, which we have not yet discussed, is the signal-to-noise ratio (SNR) at the SHLM output. If we assume that the noise in the beam signal is determined by the Poisson statistics of atomic arrivals at the detector (the implicit assumption for white-noise limited performance of the CFS), then the rms statistical fluctuation associated with the detection of N atoms is $\sqrt{N}$. Under these conditions, the ratio of SHLM SNR to BCM SNR will be

$$\rho = \frac{SNR_{SHLM}}{SNR_{BCM}} = \sqrt{\frac{\langle |i_+| \rangle}{\langle I_+ \rangle}}. \qquad (14)$$

Since all the atoms in the beam are counted in $\langle I_+ \rangle$, but only a fraction of them participates in $\langle |i_+| \rangle$, $\rho < 1$. For our choice of computational parameters, we typically obtained $\rho \cong 0.23$, dropping as far down as 0.14. Our analysis indicates that having both a SHLM and a BCM available would result in an enhanced differential diagnostics capability for the CFS. If that option is not available and a single monitor must be chosen, the SHLM may offer some advantages over the BCM, inasmuch as it is more sensitive to the time-dependency of the beam signal. (The precision with which the CFS can be locked to the atomic transition is determined, to a large extent, by the slope of the error signal in the feedback loop, which in turn is determined by the sensitivity of the loop to the presence of a beam signal component at the first harmonic of the modulation frequency, i.e., a time-dependent signal.) For instance, one may question the ability to lock precisely a CFS if the ionizer temperature is substantially lower than the critical temperature Θ_0, since the first-harmonic component in the beam signal under out-of-lock conditions would be greatly attenuated by the long ionizer response time. The output of the BCM would not reflect this situation at all, but the output of the SHLM would be greatly reduced, thus signaling the presence of a problem. More analysis is required to investigate the possibility of false negatives (i.e., situations in which the output of the SHLM would display large variations for CBT processes which do not have a significant impact on clock performance).

Appendix: derivation of the atomic transition probability

The quantum-mechanical problem of a two-level atom interacting with an harmonic electromagnetic field can be solved exactly. Let the two levels have energies E_p and E_q, with $E_q - E_p = \hbar\omega_0$. For magnetic dipole coupling between a ground-state alkali atom and the field (neglecting nuclear magnetic dipole contributions), the time-dependent part of the hamiltonian is $H' \cong \mu_B B_0 \cos\omega t$, where ω is the angular frequency of the field, B_0 its magnetic amplitude and μ_B the Bohr magneton. The Rabi frequency of the atom in the field will be $b = \mu_B B_0/\hbar$, and the detuning $\Omega = \omega_0 - \omega$. If the probability amplitudes for states p and q at time t are, respectively, $C_p(t)$ and $C_q(t)$, then at a time $t+\delta t$ they will be given by [8]:

$$C_p(t+\delta t) = \{[i\cos\theta \sin(\alpha\,\delta t/2) + \cos(\alpha\,\delta t/2)]\, C_p(t) + [i\sin\theta\sin(\alpha\delta t/2)\, e^{i\omega t}]C_q(t)\}e^{i(\omega-\omega 0)\delta t/2} \qquad (A1a)$$

$$C_q(t+\delta t) = \{[i\sin\theta \sin(\alpha\,\delta t/2)e^{-i\omega t}]C_p(t) -[i\cos\theta\sin(\alpha\delta t/2)-\cos(\alpha\delta t/2)]C_q(t)\}e^{-i(\omega+\omega 0)\delta t/2} \quad \text{(A1b)}$$

In Eqs. (A1a) and (A1b), $\alpha = [\Omega^2 + b^2]^{1/2}$; $\cos\theta = \Omega/\alpha$, and $\sin\theta = -b/\alpha$. We also assume near-resonance conditions, $\Omega << b$, so that $\alpha \cong b$, $\cos\theta \cong 0$ and $\sin\theta \cong -1$. Equations (A1a,b) can be then rewritten as (A2a) and (A2b), respectively:

$$C_p(t+\delta t) \cong \left\{\cos\frac{b\delta t}{2}C_p(t) - i\sin\frac{b\delta t}{2}e^{i\phi}C_q(t)\right\}e^{i(\omega-\omega_0)\delta t/2}$$

$$C_q(t+\delta t) \cong \left\{-i\sin\frac{b\delta t}{2}e^{-i\phi}C_p(t) + \cos\frac{b\delta t}{2}C_q(t)\right\}e^{-i(\omega+\omega_0)\delta t/2}$$

where ϕ is the phase of the field at time t. In a field-free region, these equations reduce to

$$C_p(t+\delta t) = e^{-i\omega_p\delta t}C_p(t) \quad \text{(A3a)}$$

$$C_q(t+\delta t) = e^{-i\omega_q\delta t}C_q(t), \quad \text{(A3b)}$$

where $\omega_p = E_p/\hbar$ and $\omega_q = E_q/\hbar$.

The atom enters a Ramsey-type microwave cavity at $t_1=0$. The usual initial conditions obtain: $C_p(0) = 1$ and $C_q(0) = 0$ (i.e., the atom is certainly in state p). The atom exits the first interaction region at time $t_2 = \tau$, enters the second interaction region at time $t_3 = T+\tau$, and exits the cavity at time $t_4 = T+2\tau$. We use Eqs. (A2a,b) and (A3a,b) to calculate iteratively $C_p(t_{i+1})$ and $C_q(t_{i+1})$, using $C_p(t_i)$ and $C_q(t_i)$ as initial conditions. The probability $P(\tau)$ that the atom will undergo a hyperfine transition in transit through the cavity is $P(\tau) = |C_q(t_4)|^2$.

We now assume that the microwaves are square-wave frequency modulated, with mean frequency $\varpi = \omega_0$ (i.e., the CFS is locked), modulation amplitude ω_m and modulation period T_m. Under these conditions, $\Omega = \omega_0 - \omega = \mp\omega_m$. Atomic transit times through the Ramsey-type cavity are of the order of 10^{-3} s, while the microwave modulation period is typically of the order of 10^{-2} s; thus, a large fraction of the atoms will travel through the cavity at constant microwave frequency. For those atoms, under near-resonance conditions ($\omega_m << b$) the transition probability is given by Ramsey [9]

$$P(\tau) \cong \tfrac{1}{2}\sin^2 b\tau\{1+\cos\omega_m T\}, \quad \text{(A4)}$$

The sign of Ω is immaterial, since the cosine is an even function of its argument. To calculate the transition probability for those atoms which are in transit as the microwave frequency changes by $\pm 2\omega_m$, we need to insert an extra step in the iterative calculation of the probability amplitudes $C_p(t)$ and $C_q(t)$, at the time t' at which the microwave frequency is switched. The calculation must be then performed separately for $0 \le t' < \tau$, $\tau \le t' < \tau+T$ and $\tau+T \le t' < T+2\tau$.

When the frequency is switched while the atom is in the first interaction region ($0 \le t' < \tau$) or the second one ($\tau+T \le t' < T+2\tau$), the transition probabilities $P(\tau)$ are again given by Eq. (A4) in both cases, although the corresponding probability amplitudes $C_q(T+2\tau)$ differ by a phase factor. If the frequency is switched while the atom is traveling between both interaction regions, a time T' after entering the field-free region within the cavity ($T' < T$), the iterative calculation yields a transition probability

$$P(\tau) \cong \tfrac{1}{2}\sin^2 b\tau\{1+\cos[\omega_m(2T'-T)]\}. \quad \text{(A5)}$$

Equation (A5) reduces to Eq. (A4) when $T' = 0$ or $T' = T$, as expected; the results derived in this Appendix are fully consistent with those obtained graphically in Section II.

References

[1] B. Jaduszliwer, "Signal Decay in Cesium Clocks: A Review of its Causes and Effects", TOR-92(2470)-1, The Aerospace Corp., 1 March 1992.

[2] N. D. Ramsey, "A New Molecular Beam Resonance Method", Phys. Rev. **76**, 996 (1949).

[3] J. Vanier and C. Audoin, "The Quantum Physics of Atomic Frequency Standards" (Adam Hilger, Bristol, 1989), Vol. 2, pp 729-733.

[4] B. Jaduszliwer, "Atomic Transit and Delayed Ionization Effects on Cesium-Beam Frequency Standards", IEEE Trans. UFFC **34**, 690 (1987).

[5] F. L. Hughes and H. Levinson, "Mean Absorption Lifetime of Rb on Etched Tungsten Single Crystals: Ions", Phys. Rev. **113**, 1029 (1958).

[6] E. G. Nazarov, "Observation of atomic desorption dynamics with a voltage-modulation method", Sov.Phys. Tech Phys. 24, 707 (1979).

[7] J. Vanier and C. Audoin, "The Quantum Physics of Atomic Frequency Standards" (Adam Hilger, Bristol, 1989), Vol. 2, p. 937.

[8] N. F. Ramsey, "Molecular Beams" (Clarendon Press, Oxford, 1956), p.127.

[9] *ibid.*, p. 128.

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

GLONASS ON-BOARD TIME/FREQUENCY STANDARD - ARCHITECTURE AND OPERATION

Dr. Pyotr P. Bogdanov, Prof. Yuri G. Gouzhva, Prof. Arvid G. Gevorkyan,
Dr. Arkady B. Bassevich, Dipl.-Eng. Arkady E. Tyulyakov

Russian Institute of Radionavigation and Time (RIRT),
2, Rastrelli square, St.Petersburg, 193124, Russia

Abstract

While being one of the main units of the Global Navigation Satellite System (GLONASS), the spaceborne time/frequency standard (STFS) determines, to a great extent, a whole number of system technical and operation performances. Among these is, first and foremost, an accuracy of satellite time synchronization against system time, as well as probability of target task solving, satellite active life and time for recovery of system integrity. This paper presents principles of STFS construction and functioning along with its main parameters.

Introduction

According to its purpose, GLONASS onboard STFS provides following functions: generation and keeping of onboard time with a prescribed accuracy against the system time; generation of synchrofrequencies grid from 5-MHz reference signal of frequency standard; transfer of various signals to other satellite hardware; switching of backup units; generation and transfer of telemetric data on clock functioning. Architecture enables to use any combination of backup units. When a functioning STFS fails, a backup standard is switched on which, after proper warm-up time, reproduces prescribed metrological parameters. Then satellite time is generated and updated in order to synchronize it against system time with nanosecond accuracy.

Proceeding from system accuracy parameters, a cesium-beam frequency standard was used as STFS reference generator. Its daily frequency instability is no more than 5×10^{-13} providing a mutual synchronization accuracy of satellite times within the system at the level of no more than 20 ns for 12 hours and 40 ns for 24 hours. The clock structure enables to use any type of frequency standard as a reference generator.

Principles of constructing STFS

A proper STFS version was choosed, while using as a basis a solution of optimal designing task with respect to system reliability parameters, its life expectancy and time for integrity recovery, with consideration for limitations on STFS weight, dimensions, consumed power and available components. The STFS version was taken having twofold "cold" redundancy of reference oscillator (RO) and unit "cold" redundancy of equipment for synchronization, time and control (ESTC) and power supply units.

Block-diagram of GLONASS STFS is presented in Fig.1. View of frequency standard and equipment for synchronization, time and control are shown in Fig.2 and Fig.3.

STFS architecture enables to use any combination of units, for example, the first reference oscillator together with the second ESTC and the first power supply unit. The switching of standby units is performed according to instructions from a ground station. The cesium-beam frequency standard with daily instability of no more than 5×10^{-13} is used as a reference oscillator at present, but the STFS structure enables to use any type of reference oscillator. Control over STFS functioning is realized by means of amplitude/signal sensors.

ESTC comprises generator of synchrofrequencies set, generator for time digitizing, synchrosignals output amplifiers, equipment for time phasing and correction and checking/control unit. The generator of synchrofrequencies set provides a generation of synchrosignals and its transfer to the other satellite

0-7803-1945-1/94 $3.00 © 1994 IEEE

units. The digitizer provides generation and transfer to the other satellite units a sequential binary-decimal time code in a structure of second-minute-hour-day number within 4 years.

An equipment for time phasing and correction provides a realization of time phasing in two modes:

- with single pulse putting into a proper state a generator of synchrofrequencies set and a second counter of digitizer. This pulse is generated on the ground station being synchronous with a proper system time minute mark and transferred to the satellite via a channel of command/measurement system. Therefore an accuracy of satellite times phasing for this mode is rather low;

- with phasing code, by performing an offset of time to be generated for the prescribed value. A phasing accuracy for this mode is 0.2 μs.

Time correction can be realized also in two modes:

- with packet of pulses, which number is in accordance with the true value of current time;

- with correction code setting the direct true value of current time digitizing.

Apart from this, there is in STFS a mode of time correction for +1 s or -1 s according to occasional instructions from the time program of satellite operation. These modes are introduced for providing a synchronism against system time which is corrected in GLONASS when an appropriate UTC (SU) correction takes place [1].

In order to check STFS functioning, the telemetric data sensors of four types are used:

- amplitude sensors that store data on parameter to be checked in signal amplitude;

- signal sensors based on relay functioning;

- event sensors which signals cause an interrogation of a certain group of telemetric data sensors;

- temperature sensors for checking temperature in the place of spaceborne cesium-beam frequency standard location.

The amplitude sensors check a functioning of a cesium-beam tube in the first place: microwave power of atomic transition excitation, atomic beam current etc., as well as such a generalized parameter as "dual frequency" characterizing a routine functioning of cesium-beam frequency standard.

Signal sensors, by closing relay contacts, fix a status of switched-on equipment sets, passing of control instructions etc. Event sensors register, for example, such states as a loss of output signal of atomic-beam frequency standard being the factor of the most importance for STFS functioning.

STFS functioning in real environment

STFS switching-on is performed on the start site; in such a case, a power is given to the quartz oscillator (QO) only of the first STFS (mode of lower power supply). After placing a satellite to the proper orbit, an instruction is given to switch on a high-stability mode. According to this instruction, search and automatic tuning of QO frequency to the atomic transition frequency are performed. When a tuning is completed, STFS is ready to routine functioning. Then a checking of working ability of the second and the third sets of cesium-beam frequency standard is performed, along with a checking of its ESTC. If the results of checking for all the standby sets are positive, then such operations are performed as switching of the first sets of equipment, compensation for STFS relativity frequency shift and satellite time setting.

Compensation for STFS relativity frequency shift is performed for fixed value 4.36×10^{-10} according to the height of the mean estimated orbit. Satellite time setting is realized by phasing and correcting time to be generated while using data from a ground station.

Checking of STFS functioning in the course of further operation comprises a self-testing of main units' functioning, telemetric data checking, STFS accuracy parameters checking and checking of time generation/synchronization accuracy. Results of self-testing of STFS main units' functioning are transferred to satellite computer where its are considered when a generalized parameter of satellite functioning is generated. Telemetric data are transferred to a ground station for analysis. STFS accuracy parameters are estimated every day, by processing results of comparing satellite time against system time. An operative checking of time generation accuracy is performed for each session of measuring satellite times'

offset, and checking of time synchronization accuracy (for system routine operation) is realized once in every 10 days.

Control over STFS operation in the course of satellite flight is performed using results of checking of its functioning and analysis of these results; such a control involves a switching to standby units in case of violating of operation of main ones, as well as phasing of satellite time when its offset against system time is more than 1 ms or when a generator of synchrofrequencies' set fails, correction of satellite time when a digitizer fails and correction of satellite time for +/- 1 s when an appropriate UTC (SU) takes place.

STFS main parameters

STFS main accuracy parameters as registered on GLONASS satellites being operated are given in [1,2].

In the nearest time, the flight tests of satellites are planned that will have STFS with improved accuracy/operational parameters. This is concerned, in the first place, with STFS daily instability that is diminished up to 1×10^{-13} and duration of continuous operation (up to 5 years). This enables to provide an accuracy of mutual synchronization of phases of navigation satellites' signals being no worse than 15 ns for 24 hours.

Conclusion

Principles of constructing and functioning of GLONASS STFS provide a realization of prescribed accuracy/operational parameters, including probability of failure-free functioning, life time, time for recovery of system integrity.

While using a cesium-beam frequency standard with daily instability of no more than 5×10^{-13} as a STFS reference oscillator, an accuracy of mutual synchronization of phases of satellite navigation signals is no worse than 20 ns for 12 hours and 40 ns for 24 hours.

References

[1] Y.G.Gouzhva, A.G.Gevorkyan, A.B.Bassevich, P.P.Bogdanov. High-Precision Time and Frequency Dissemination with GLONASS. GPS World, July 1992, pp.40-49.

[2] Y.G.Gouzhva, A.G.Gevorkyan, A.B.Bassevich, P.P.Bogdanov, A.E.Tyulyakov. Comparative Analysis of Parameters of GLONASS Spaceborne Frequency Standards when Used Onboard and on Service Life Tests. 1993 IEEE International Frequency Control Symposium, June 1993, pp.65-70.

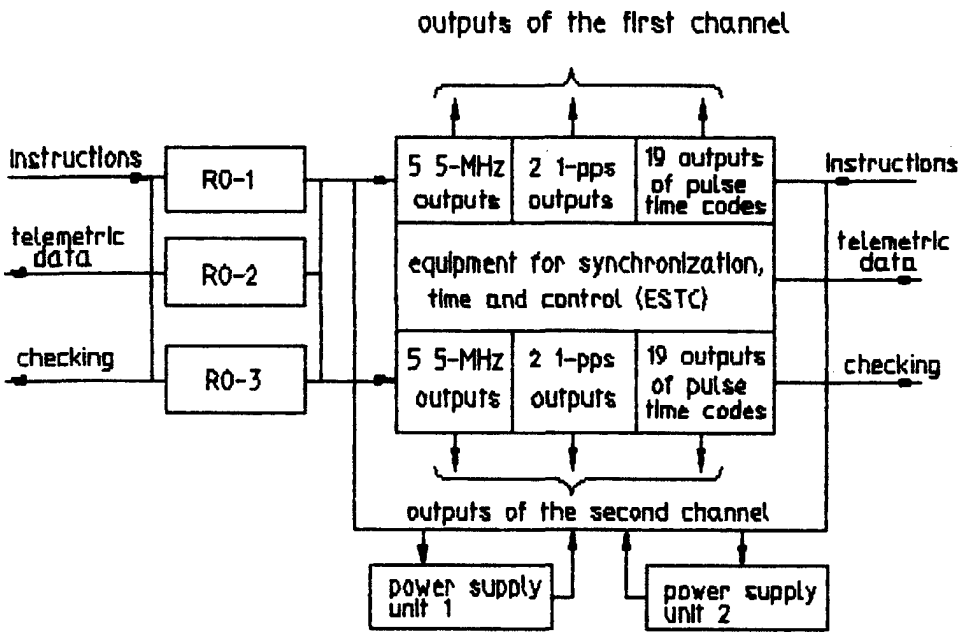


Fig.1 Block diagram of GLONASS STFS

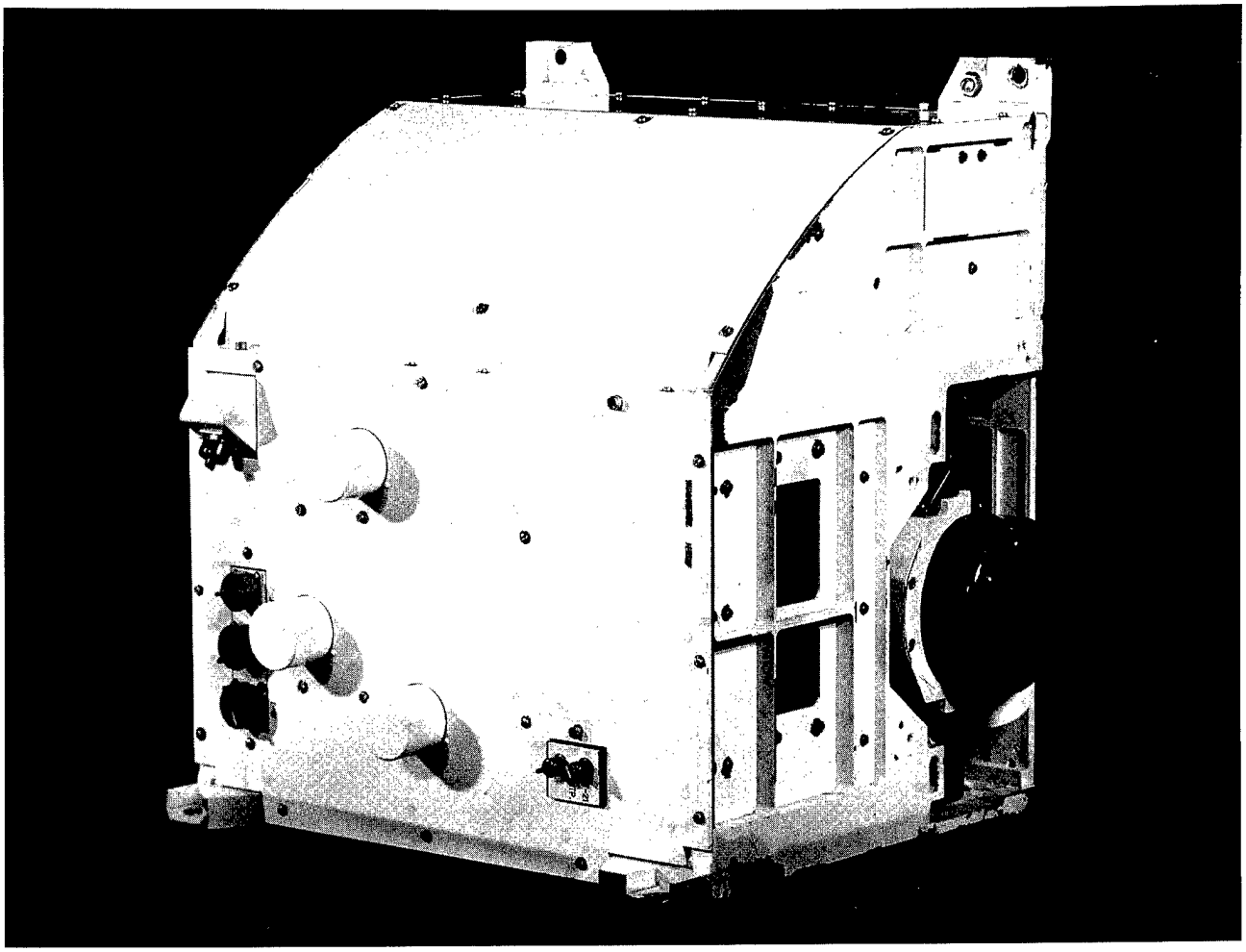


Fig.2 View of frequency standard

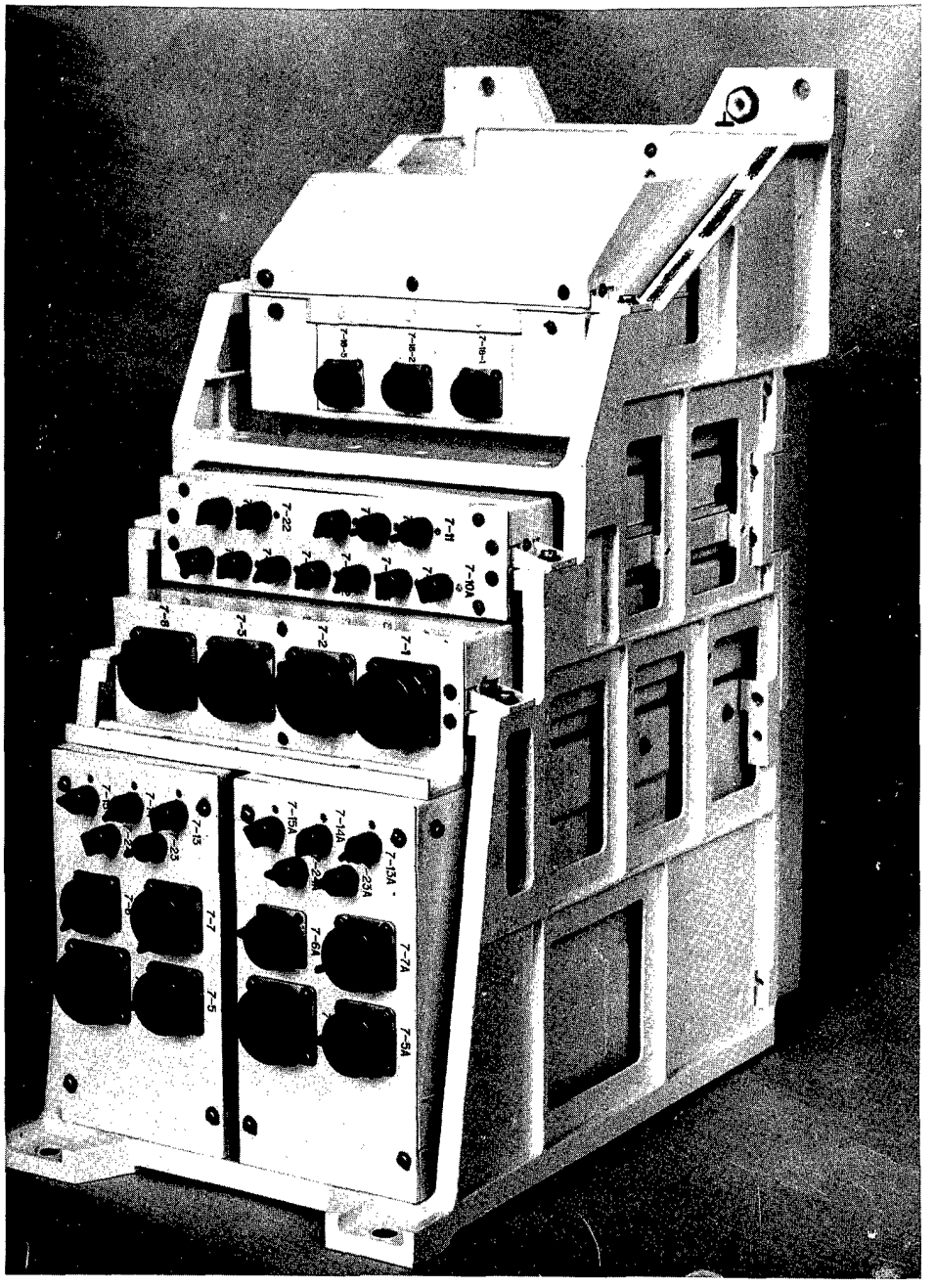


Fig.3 View of equipment for time synchronization, time and control

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

EFFECT OF SPACE VEHICLE TEMPERATURE ON THE FREQUENCY OF ON-ORBIT NAVSTAR CLOCKS

Wilson G. Reid
Thomas B. McCaskill
Orville J. Oaks

U.S. Naval Research Laboratory
Washington, D.C., 20375

James A. Buisson
Hugh E. Warren
Sachs Freeman Associates, Inc.

ABSTRACT

Analysis of the frequency of on-orbit Navstar clocks is performed by the Naval Research Laboratory† using both broadcast and post-processed precise ephemerides. The phase offset between the Navstar clock and the reference clock is computed from pseudo-range measurements obtained by the five GPS monitor sites and by the U.S. Naval Observatory, Washington, D.C., precise-time site using precise-positioning-service dual-frequency, receivers which correct for selective availability. The frequency offset of the early Block I rubidium clocks was found to be highly correlated with the temperature of the space vehicle and consequently with the earth eclipse seasons. By providing subsequent space vehicles with rubidium clocks having additional temperature control this correlation was suppressed. The Block I cesium clocks, on the other hand, showed no sensitivity to the space vehicle temperature. Recently, however, two Block II cesium clocks and a Block I rubidium clock with the additional temperature control have evidenced temperature sensitivity and have shown strong correlation with the onset of the eclipse seasons. Frequency-offset histories show this correlation for the cesium clocks for the first time. Scatter diagrams used to measure the degree of correlation of the frequency with temperature yield temperature coefficients larger than those measured in the laboratory prior to launch.

BACKGROUND

The Global Positioning System is a Department of Defense (DoD) space-based navigation system that provides precise position, time, and frequency to users located anywhere in the world. Twenty-four space vehicles are deployed in six planes at an inclination of fifty-five degrees. Each Navstar space vehicle broadcasts time-coded signals and a navigation message. Information in the navigation message is used to determine the position of the space vehicle at the time of observation and to relate the time of each of the space vehicle clocks to a common system time. Each user passively receives the time coded signals from four of the space vehicles and uses the information to compute position (three dimensions), velocity, and time. The system time is steered so that a precise-time user can determine the offset of his clock from Coordinated Universal Time (UTC) by using an additional correction provided in the navigation message.

Operational control of the system is exercised from the Master Control Station located at the Consolidated Space Operations Center (CSOC) at Falcon Air Force Base (FAFB) in Colorado Springs, Colorado. The control segment consists of the Master Control Station plus five monitor stations located in Colorado Springs, on the island of Hawaii, and on Kwajalein, Diego Garcia and Ascension Islands. *Figure 1* depicts the flow of information from a single space vehicle to each of the monitor stations. The monitor stations passively track all Navstar space vehicles in view, collecting pseudorange and pseudorange-rate data from each space vehicle. This tracking information is sent to the Master Control Station where the ephemeris and clock parameters for each space vehicle are estimated and predicted. The Master Control Station periodically uploads to each space vehicle the ephemeris and clock parameters which are included in the navigation message. Additional measurements referenced to the DoD Master Clock are col-

† *This work was sponsored by the GPS Joint Program Office.*

0-7803-1945-1/94 $3.00 © 1994 IEEE

lected at the U.S. Naval Observatory (USNO), Washington, D.C., for use in determining the difference between the GPS system time and UTC (USNO).

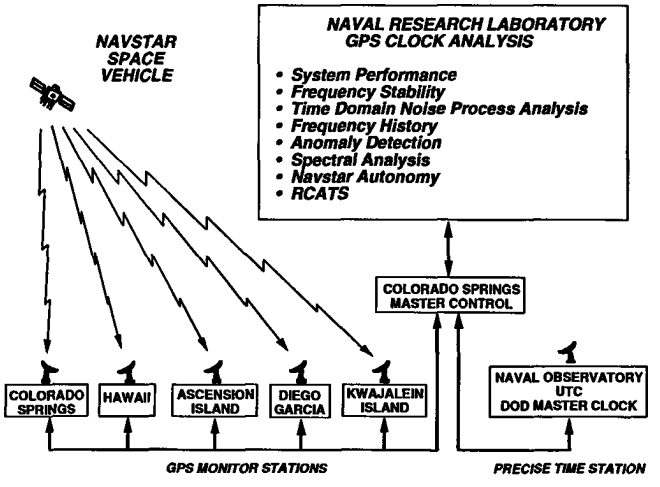


Figure 1

OVERVIEW

The frequency offset of early space vehicle rubidium clocks was found to be highly correlated with the temperature of the space vehicle [1] and consequently with the earth eclipse seasons [2]. For example, *Figure 2*, which is a partial history of the frequency offset of the first Navstar 3 rubidium clock to be activated (Serial Number 19), shows broad swings in the frequency as it passed through eclipse seasons (shaded regions) every six months during the six years from June 1985 to June 1991 when it was deactivated.

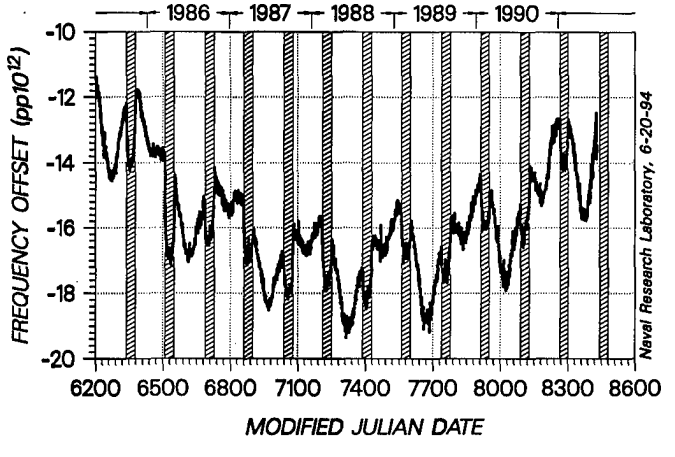


Figure 2

The rubidium clocks on subsequent space vehicles had additional temperature control which suppressed the correlation. For example, the first Navstar 8 clock to be activated was a rubidium clock (Serial Number 33) which operated for almost three years from July 1983 to May 1986 with no evidence of sensitivity to the eclipse seasons. *Figure 3*, which is a plot of the residuals of a linear fit to the frequency-offset history, shows the only temperature effect to be a step in the frequency of $-4.6pp10^{12}$ caused by a $-3\ ^oC$ step adjustment in the temperature of the oven.

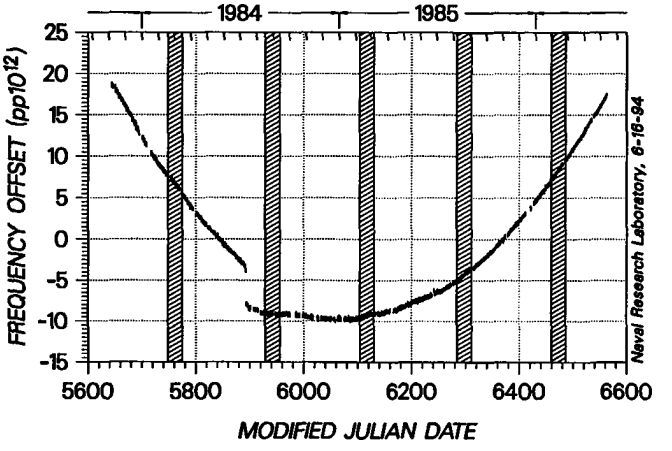


Figure 3

The Block I cesium clocks, on the other hand, appeared to show no sensitivity to the space vehicle temperature. For example, the Navstar 9 cesium clock (Serial No. 4), which operated continuously for over nine years from June 1984 to October 1993, had a frequency history that evidenced very little structure other than white frequency noise and a drift that slowly increased with the age of the clock.

Recently, however, two Block II cesium clocks and a Block I rubidium clock with the additional temperature control have evidenced temperature sensitivity and have shown strong correlation with the onset of the eclipse seasons. The first Navstar 14 cesium clock (Serial Number 8) to be activated operated continuously for three and a half years from February 1989 to August 1992. The frequency of this clock showed very strong correlation with the eclipse seasons—an effect that appeared to intensify as the clock aged. Similarly, the first Navstar 16 cesium clock (Serial Number 11) has operated continuously for three years since January 1991, and its frequency has shown very strong correlation with the eclipse seasons. Scatter diagrams used to measure the degree of correlation of the frequency with temperature yielded temperature coefficients larger than those measured by Rockwell International during prelaunch testing of these two space vehicles.

The Navstar 10 rubidium clock (Serial Number 28) showed the most interesting behavior of all. While the frequency of a typical clock that is sensitive to temper-

ature shows a characteristic pattern correlated with the temperature not only during the eclipse seasons but also during the period between eclipse seasons, the frequency of the Navstar 10 clock showed no such pattern. Instead, the frequency appeared insensitive to temperature variations up to some point during the eclipse season after which control appeared to be lost until the temperature variations returned to their pre-eclipse season values. Moreover, the profile of the frequency during the eclipse season showed wide variation from eclipse season to eclipse season which is not typical behavior for the Navstar clocks.

FREQUENCY DETERMINATION

The phase offset between the reference clock at the tracking station and the remote clock in the space vehicle is estimated using pseudorange measurements collected as the space vehicle passes over the tracking station. The one-day measurements of the phase offset were obtained from several 13-minute phase-offset measurements collected at USNO. By taking the measurement nearest the time of closest approach (the one having the highest elevation angle) the effect of the ionosphere, the troposphere, and the along-track orbit error is minimized. These measurements were obtained in turn from a linear least squares fit to 13 minutes of six-second phase-offset measurements evaluated at the beginning of the observation interval. The six-second phase-offset measurement was obtained from the difference between the pseudorange measurement and the predicted range obtained from the orbital elements [3] broadcast in the navigation message. The frequency offset is estimated by averaging the change in the phase offset over a sidereal day.

The Block I data to be presented was collected at the U.S. Naval Observatory using a single-frequency, time-transfer receiver with ionospheric corrections obtained from a model of the ionosphere contained in the navigation message. The broadcast signal for the Block I space vehicles is unaffected by selective availability (SA). The Block II data to be presented was collected at the Naval Observatory using precise-positioning service (PPS) receivers which measure the ionospheric delay and automatically correct for the effects of selective availability. In both cases the broadcast ephemeris was used to compute the position of the space vehicle in making the pseudorange measurement. The time-transfer receiver is driven by the DoD Master Clock which exhibits a long-term frequency stability that is significantly better than any of the space vehicle clocks.

TEMPERATURE MEASUREMENTS

Figure 4 presents the raw temperature measurements for the Navstar 3 Louver No. 2 Base Plate, which constitutes the heat sink for the rubidium clock, frequency standard No. 1, during a period of approximately two years. Immediately apparent is the quantization level of 0.33 degree Celsius. Also visible are the seasonal temperature variations which appear to be at a maximum during the eclipse seasons which are shown as shaded regions on the plot.

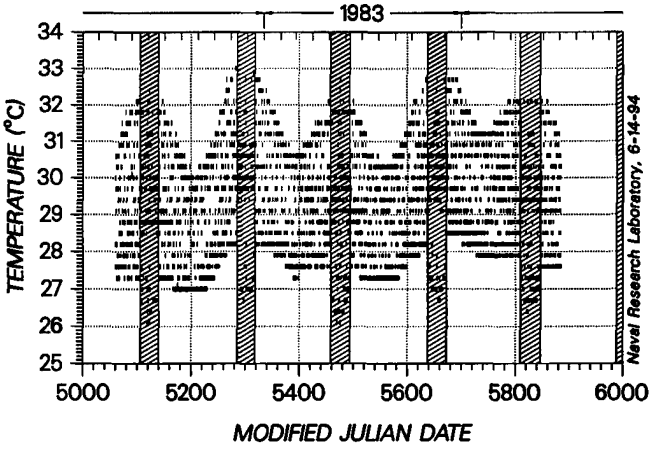


Figure 4

Figure 5 is a plot of the raw temperature measurements for a single day—20 May 1983—during the mid-1983 eclipse season. The measurements appear very sparse over the two orbits, i.e., about four measurements per orbit.

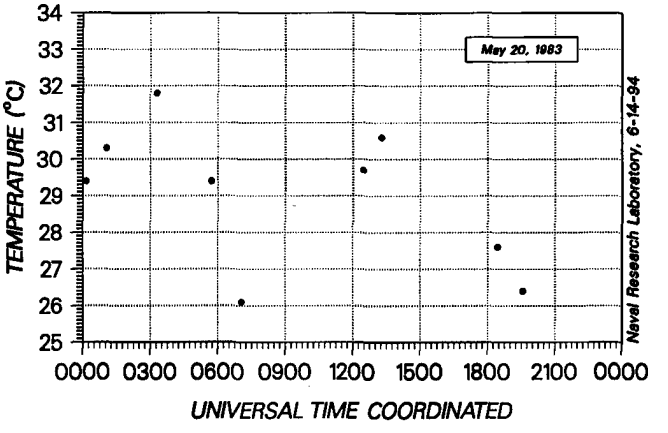


Figure 5

Given that adjacent orbits are essentially identical, one may superimpose the measurements from a number of successive orbits [4] as was done in *Figure 6*. Because of differences in the scheduled times of the measurements for different orbits, when the measurements are superimposed on a single orbit the profile of the orbital temperature variation can be more fully populated.

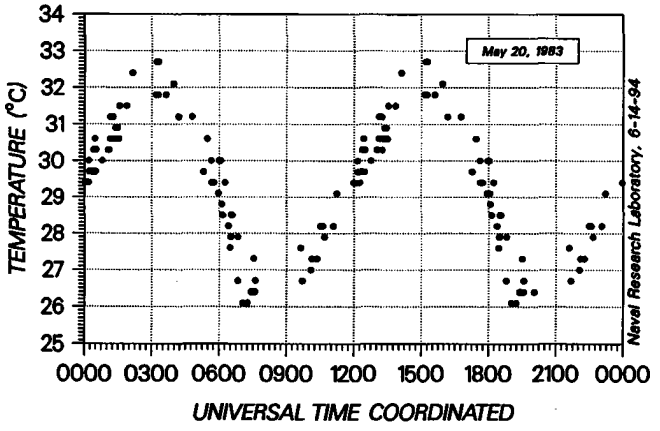


Figure 6

Figure 7 shows the effect on the temperature measurements of a moving average filter. The top trace in the figure presents the raw measurements. The second trace shows the results of applying a two-point moving average filter to the raw measurements. Each successive trace presents the results of applying a moving average filter with a length that is a factor of two larger than in the previous trace.

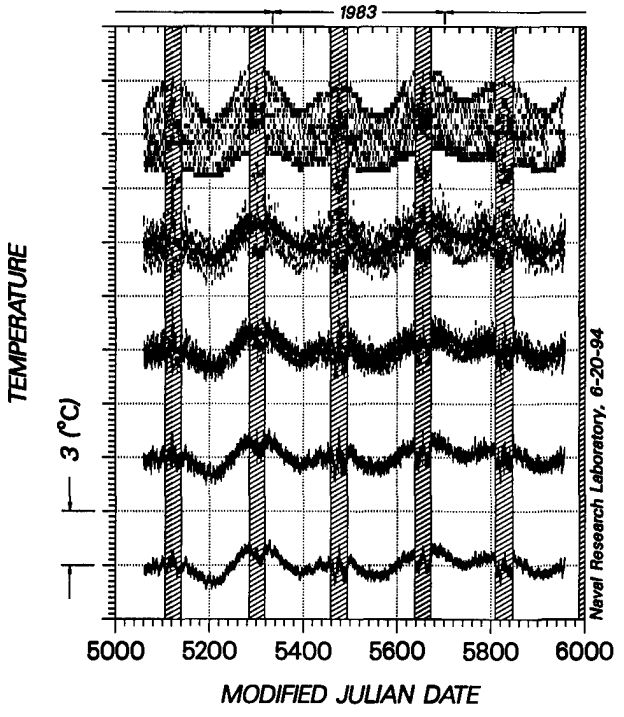


Figure 7

Figure 8 is a superposition on the raw measurements of the output of a 64-point moving average filter which, with an average of four measurements per orbit, spans approximately 16 orbits. This figure shows the effect of a moving average filter having a window equal to a multiple of the orbital period. The window must include only those successive orbits having similar temperature profiles to obtain the daily average temperature.

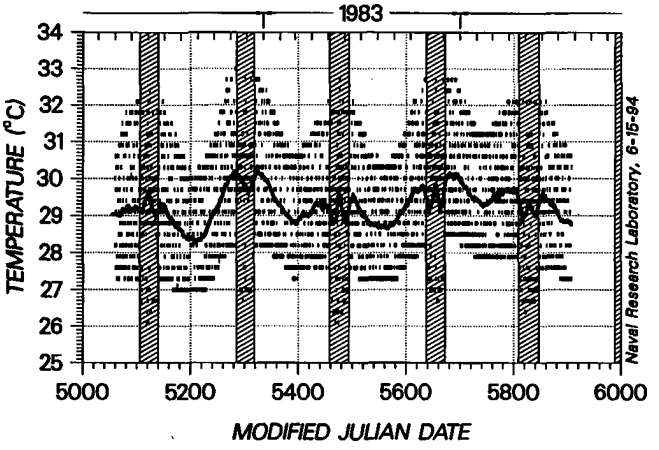


Figure 8

Figure 9 presents both the daily average temperature (upper trace) and the daily average frequency (lower trace) for the Navstar 3 rubidium clock for a period of about three years beginning in early 1982. The raw temperature measurements were smoothed with a 16-point moving average filter to obtain the daily average temperature. To examine the degree of correlation between the temperature and the frequency, those components of frequency variation that should be independent of temperature, i.e., random walk noise and drift, were modeled by a high order polynomial and were removed by subtracting the model from the frequency offset.

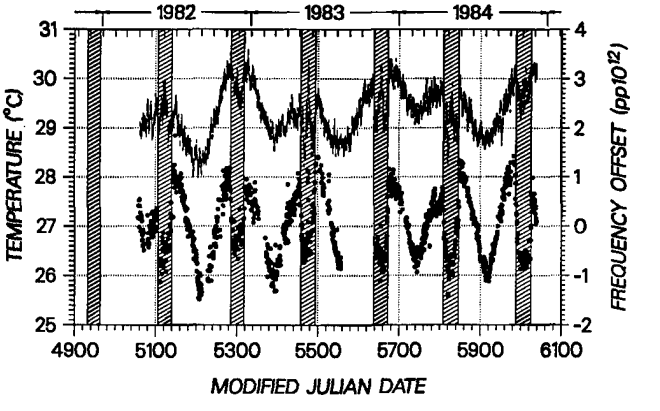


Figure 9

NAVSTAR 10

Figure 10 shows the frequency offset of the Navstar 10 rubidium clock for 1992 and 1993. During the eclipse seasons of June and December 1992 there appear to be disturbances in the frequency. Superimposed on the frequency is a high order polynomial which models the large components of random walk and linear drift present in this rubidium clock.

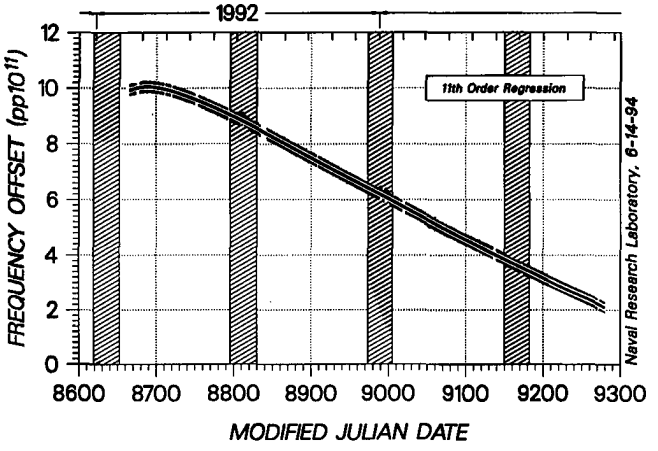


Figure 10

The scale has been expanded considerably in *Figure 11* by subtracting from the frequency history this model of the random walk and linear drift. Immediately obvious is the perturbation in the frequency of some 1.5 $pp10^{12}$ during the two eclipse seasons. Such a disturbance is notably absent from the eclipse season of June 1993.

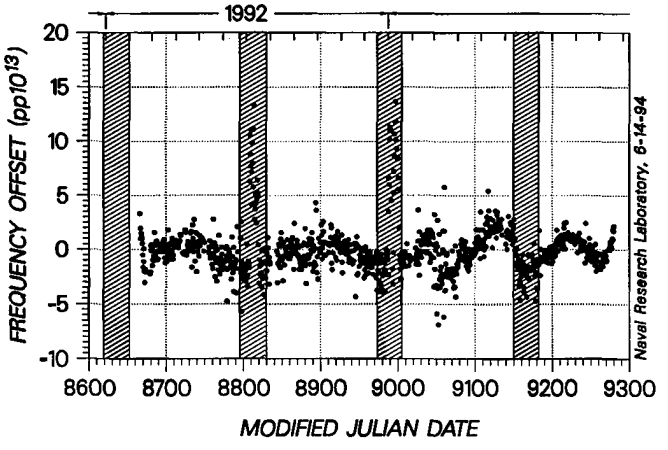


Figure 11

Figure 12, which is a plot of the raw temperature measurements for the Navstar 10 clock shows the temperature to be controlled much more tightly than for the rubidium clock on Navstar 3 which suffered orbital temperature variations of as much as seven degrees. In the case of Navstar 10 the temperature varies principally between 36.85 and 37.12 $°C$—a total variation of about 0.27 $°C$. This is approximately 26 times smaller than the temperature excursions on Navstar 3.

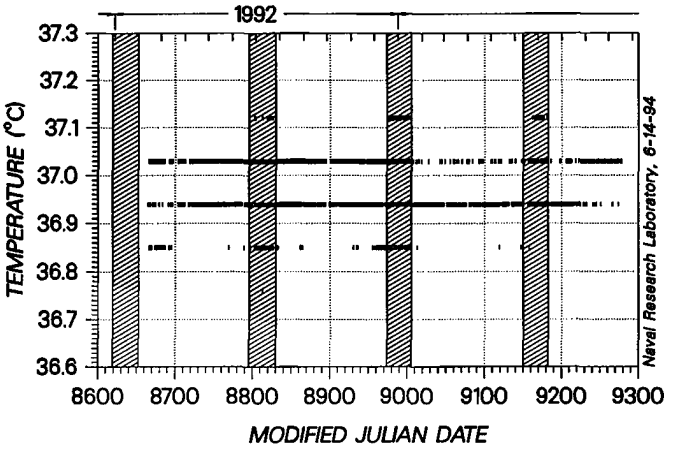


Figure 12

In *Figure 13* the raw temperature measurements from 21 successive orbits have been superimposed on each orbit and the resulting data smoothed with a 21-point moving average filter.

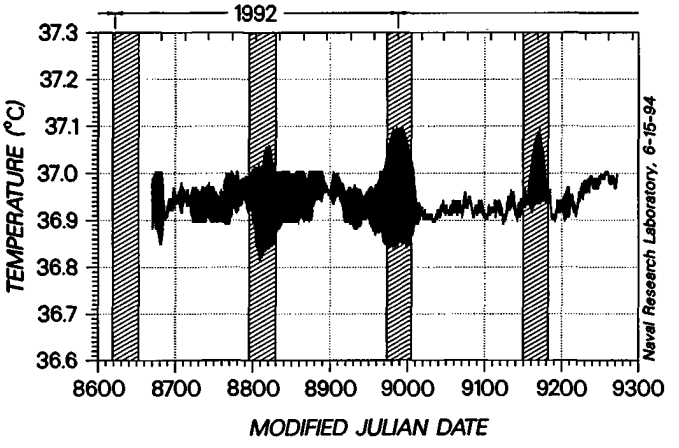


Figure 13

In *Figure 14* is a graph of both the superimposed and smoothed temperature measurements (upper trace) and the high order residuals of the frequency (lower trace) to assess the degree of correlation. Interestingly, as the temperature drops much below its average value during the first two eclipse seasons, the frequency is perturbed upwards. During the third eclipse season, the temperature appears not to go below its average value, and the perturbation characteristic of the first two

eclipse seasons is absent. Experience with the Block I rubidium clocks that did not have the additional temperature control, e.g., the Navstar 3 rubidium clock in *Figure 9*, suggests that the behavior during alternate eclipse seasons should be similar, but neither the temperature nor the frequency during the third eclipse season behave as they do during the first eclipse season.

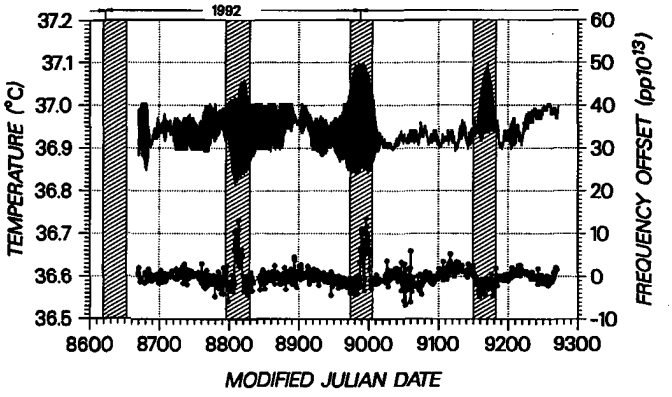


Figure 14

Looking for a reason for the difference in the behavior of the temperature during the third eclipse season it was noted that the temperature measurements were very sparse during that time. For example, during a four-day period in the middle of the third eclipse season there were only 16 measurements which equates to an average of four measurements per day or two measurements per orbit. For the corresponding four-day period in the first eclipse season, there were 35 measurements, or about four measurements per orbit.

Uneven spacing of the two measurements per orbit could result in sampling the temperature twice near its peak thereby giving a skewed result. But the measurements appeared to be evenly spaced. The absence of any perturbation in the frequency during the third eclipse season—in light of the correlation already noted between the temperature and the frequency during the first two eclipse seasons—lends additional support to the view that the sampling of the temperature was correct and that the temperature profile as plotted is correct. But that leaves unanswered the question as to why there should be such assymetry in the temperature profile.

NAVSTAR 14

Figure 15 presents the raw temperature measurements for the Navstar 14 cesium clock for a period of about three years beginning in 1990. For the period of 973 days there were 3572 measurements for an average of about four measurements per day or two measurements per orbit. Hence, the 32-point moving average of the raw temperature measurements presented in *Figure 16* would average the measurements from about 17 orbits—a span of time during which the neighboring orbits would have been similar. A significant feature of the temperature history in *Figure 16* is the presence of a positive linear drift.

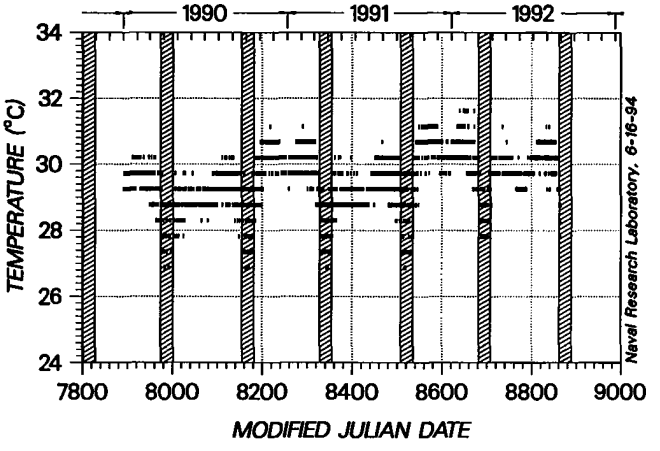


Figure 15

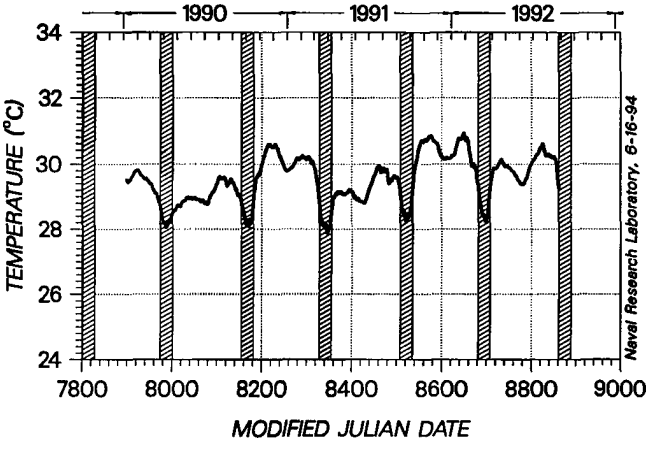


Figure 16

Figure 17 is a plot of the one-day average frequency offset for the Navstar 14 cesium clock. The increased density of measurements following the second eclipse season of 1990 resulted from taking measurements on two passes per day. A degree of correlation with the eclipse seasons appears in evidence beginning with the second eclipse season of 1990 when the frequency appears to dip during each eclipse season.

The correlation of the frequency with the eclipse seasons is more obvious in *Figure 18* where the frequency has been smoothed with a 32-point moving average filter to remove some of the noise. The smoothed data has been superimposed on the raw data for comparison. Two features of the smoothed data are worthy of note.

First, the degree of correlation appears to increase with age of the clock. For example, there is no apparent correlation between the frequency and the first eclipse season whereas during subsequent eclipse seasons the dip in the frequency increases. Second, the frequency of the clock appears to fall off after the first eclipse season of 1991 as though the aging of the space vehicle clock had changed.

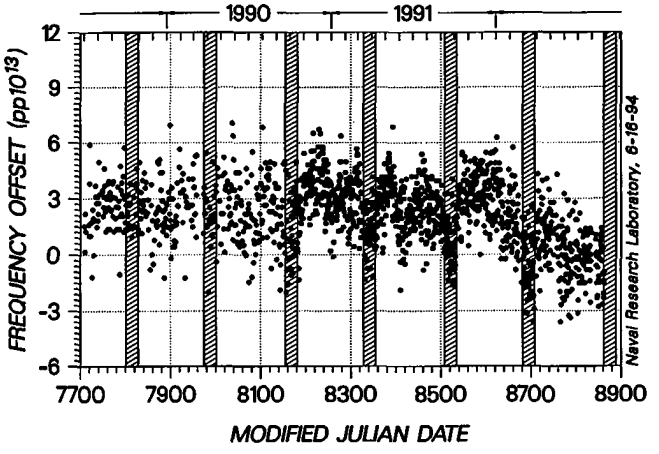


Figure 17

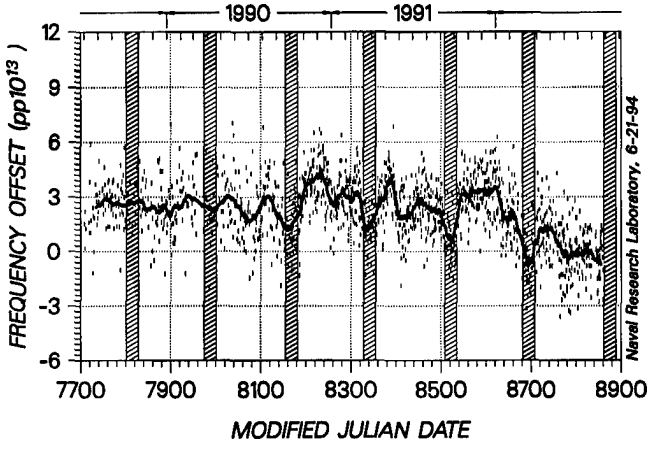


Figure 18

Figure 19 is a comparison of the temperature and frequency of the Block II cesium clock on Navstar 14. The lower trace is the 32-point moving average of the frequency history with a high order polynomial model removed. This corrects for the change in aging noted above. The upper trace is the 32-point moving average of the temperature history with the linear drift removed to match the frequency history which also has the linear drift removed.

As noted earlier, applying the moving average filter to the raw temperature measurements yields the daily average temperature. The purpose of applying the filter to the frequency offset is to suppress the noise that might otherwise obscure the correlation when the temperature coefficient is small—as it typically is with cesium clocks—causing the temperature-induced variations in the frequency to be buried in the noise. Characteristic of the moving average filter is the fact that it discriminates against periodic components with a period equal to the length of the filter window. Since the eclipse seasons have a period of approximately 178 days, the moving average filter used should not interfere with the correlation.

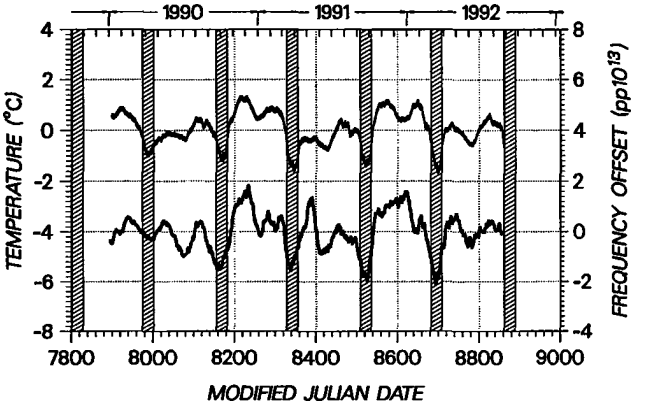


Figure 19

Figure 20 is a scatter plot showing the results of plotting the frequency as a function of the temperature. The slope of the linear regression is a measure of the temperature coefficient which is estimated to be 0.96 $pp10^{13}$ per degree Celsius. This value is twice as large as the Rockwell provided prelaunch measured value of 0.48 $pp10^{13}$ per degree Celsius.

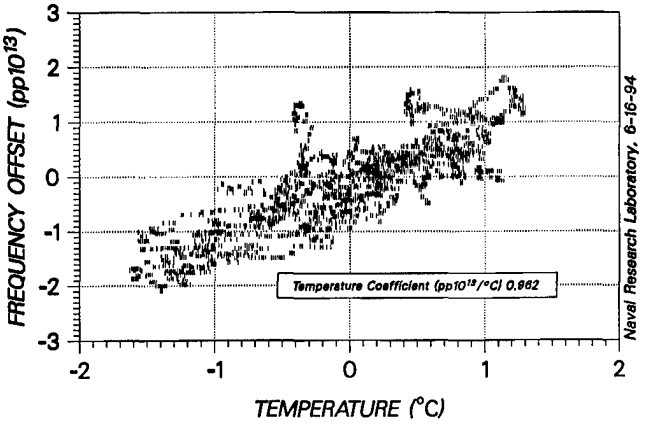


Figure 20

The degree to which the plotted points cluster about a straight line is an indication of how well the variations

in the frequency that are independent of the temperature have been identified and removed.

NAVSTAR 16

Figure 21 is a plot of the raw temperature measurements for the Block II Navstar 16 cesium clock. Notable in the behavior of the temperature is an initial decrease from the highest value achieved after which the temperature oscillates in a band between 34 and 39 degrees Celsius.

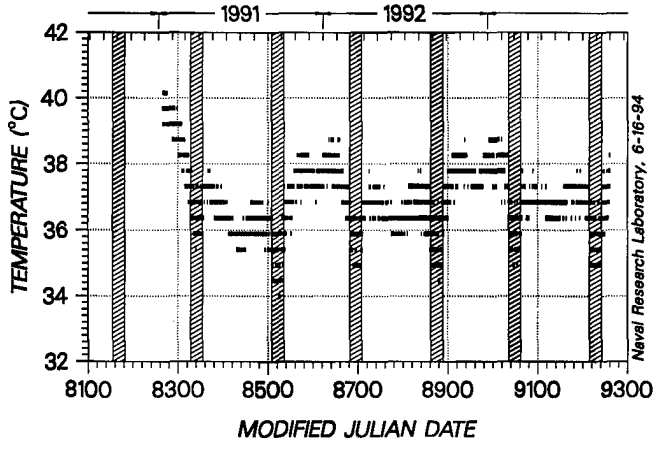


Figure 21

Figure 22 presents the raw frequency offset measurements for the Navstar 16 cesium clock. The frequency has the same initial decrease after which it experiences a long-term negative linear drift. The frequency can be seen to dip during each of the eclipse seasons.

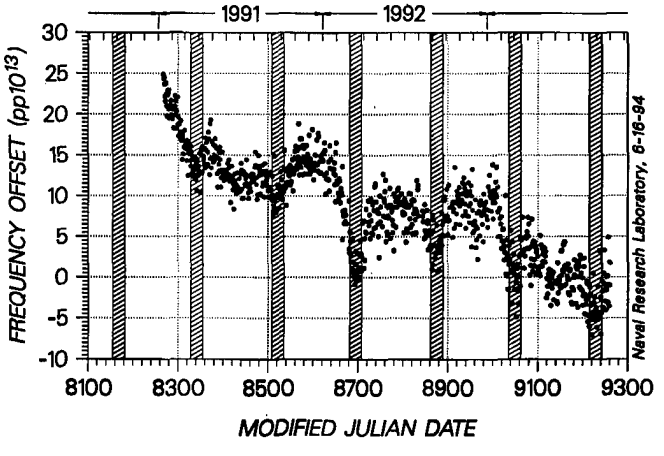


Figure 22

Figure 23 is a comparison of the temperature and frequency of the Block II cesium clock on Navstar 14. The upper trace is the 11-point moving average of the temperature history. The lower trace is the 11-point moving average of the frequency history with the linear drift of 1.5 $pp10^{15}$ per day removed. The purpose of applying the moving average filter to the frequency offset is to suppress the noise that is not expected to correlate with the seasonal temperature variations.

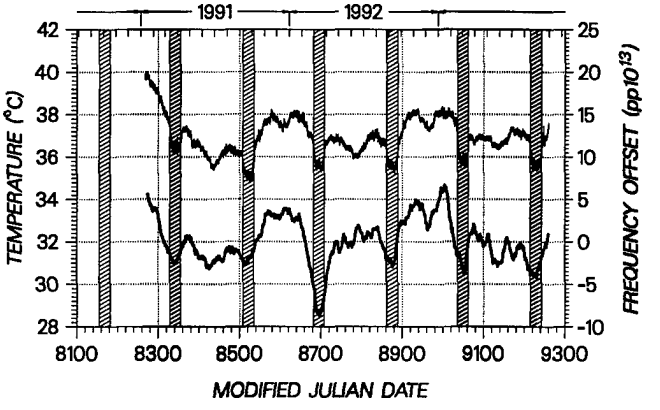


Figure 23

Figure 24 is a scatter plot showing the results of plotting the frequency of the Block II Navstar 16 cesium clock as a function of the temperature. The temperature coefficient is estimated to be 2.36 $pp10^{13}$ per degree Celsius. This value is not quite twice as large as the Rockwell provided prelaunch measured value of 1.5 $pp10^{13}$ per degree Celsius.

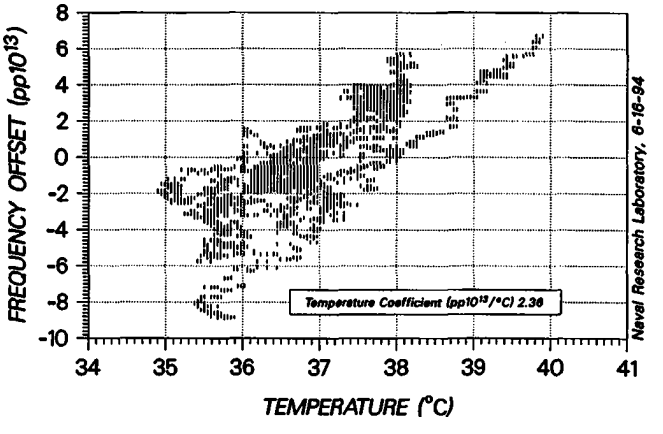


Figure 24

CONCLUSIONS

The sparse temperature measurements, i.e., two to four measurements per orbit, are sufficient to determine the orbital temperature profile. Because of the similarity of successive orbits the measurements may be superimposed on a single orbit to fully populate the profile. The

granularity of the raw measurements due to the quantization level can be compensated for by smoothing with a moving average filter. A filter having a window equal to a multiple of the orbital period and including only those successive orbits having similar temperature profiles yields the daily average temperature which has been found to correlate with the daily average frequency.

The Navstar 10 Block I rubidium clock, in spite of additional temperature control, shows eclipse season, or temperature, dependence. The Navstar 14 and 16 Block II cesium clocks are the first Navstar cesium clocks to show eclipse season, or temperature, dependence. The temperature coefficients for these two on-orbit clocks, estimated to be 1.0 $pp10^{13}$ and 2.4 $pp10^{13}$ per degree Celsius, are larger than the prelaunch measured values of 0.47 $pp10^{13}$ and 1.5 $pp10^{13}$ per degree Celsius.

REFERENCES

[1] T.B. McCaskill and J.A. Buisson, A., "NTS-1 (TIMATION III) Quartz and Rubidium Oscillator Frequency Stability Results", *Proceedings of the 29th Annual Symposium on Frequency Control*, Atlantic City, New Jersey, pp. 425–435, 28–30 May 1975

[2] T.B. McCaskill, J.A. Buisson, and T.J. Hogan, "On-Orbit Frequency Stability Analysis of GPS Navstar Cesium and Rubidium Clocks", *Proceedings of the Twenty-first Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, Redondo Beach, California, pp. 239–258, 28–30 November 1989

[3] Interface Control Document (ICD), *ICD-GPS-200*, Navstar GPS Space Segment, ARINC Research Corporation, 11770 Warner Ave., Suite 210, Fountain Valley, California 92708, 3 July 1991, p. 78

[4] Reid, Wilson G., "Spectral Analysis of the Navstar 3 Rubidium Clock", presented to the Control Segment Data Analysis Working Group (CS-DAWG), Headquarters Space Division, P.O. Box 92960, Los Angeles, California 90009, 4–5 February 1985

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

WAVELET ANALYSIS FOR SYNCHRONIZATION AND TIMEKEEPING

D. A. Howe
Time and Frequency Division
National Institute of Standards and Technology
325 Broadway
Boulder, CO 80303

D. B. Percival
Applied Physics Laboratory
University of Washington
HN-10
Seattle, WA 98195

Abstract

We discuss the concept of the wavelet variance as a generalized formalism for representing variations in a time series on a scale by scale basis. In particular, we note that the wavelet variance corresponding to some of the recently discovered wavelets can provide a more accurate conversion between the time and frequency domains than can be accomplished using the Allan variance. This increase in accuracy is due to the fact that these wavelet variances give better protection against leakage than does the Allan variance.

I. Introduction and Summary

The analysis of a time-ordered set of phase measurements $\{x_t\}$ often falls into one of three categories. The first approach treats the time series as a function to be expressed in terms of a set of basis functions defined globally over a finite interval (one example of such basis functions are orthogonal polynomials). The purpose here is to summarize the phase measurements with a few coefficients in order to quantify underlying physical effects such as drift.

The second approach takes first or second-order differences of the time series in order to transform nonstationary variations (due to low frequency components) into stationary variations. The mean square of the second difference of phase measurements at various sampling intervals τ quantifies the incremental variability of the phase and for many clocks is stationary. This is the Allan (or pair) variance as a function of τ.

The third approach is the windowed discrete Fourier transform (DFT) which is used to determine the spectral features underlying the time series. A common quantity derived from this transform is the power spectral density.

The problems with the first approach (polynomial fits) are that the coefficients are sensitive to time shifts over the finite interval; many coefficients may be needed to adequately represent the phase measurements; and, because the coefficients are calculated with respect to global basis functions, local features in the phase measurements can be misrepresented. Potential problems with the second approach include sensitivity to deterministic drifts and leakage due to the fact that the transfer function for the finite impulse response (FIR) filter associated with the Allan variance has substantial sidelobes (this leakage is quite similar to that occurring in the unwindowed DFT). The problems with the third approach (the windowed DFT) are that the results can depend on the choice of the particular window and that, because the windowed DFT is inherently narrowband, it is necessarily also highly variable and hence — without further processing — does not summarize the salient features of broadband processes.

Wavelet analysis attempts to address the potential problems with polynomial fits, the Allan variance and spectral analysis in one unified approach. First, wavelet analysis is based upon the discrete wavelet transform, which provides a "time and scale" representation of time-ordered observations. The time series is still treated as a function on a finite interval, but the wavelet basis functions are hierarchical rather than global so that, in contrast to polynomial fits, localized features (such as a step or discontinuity in phase) can be easily represented. Second, the wavelet transform can be implemented using a variety of basis functions and is narrowband at low fre-

0-7803-1945-1/94 $3.00 © 1994 IEEE

quencies and broadband at high frequencies ("multiresolutional," as it is referred to in the wavelet literature). The analyzing function used in the Allan variance when it is computed using fractional frequency deviates is identical to the Haar wavelet, a common starting point in discussions on wavelets. The variances corresponding to wavelets beyond the Haar wavelet are natural extensions to the Allan variance. These wavelet variances have potential advantages over the Allan variance in terms of leakage and insensitivity to deterministic drifts: because the transfer functions for the FIR filters associated with wavelets beyond the Haar wavelet have substantially reduced sidelobes, these wavelet variances have substantially less leakage; and because the FIR filters for higher order wavelets are based in part on differencing operations, a wavelet variance of order n will be invariant to a polynomial drift of order n. Finally, plots of the square root of the wavelet variance versus averaging time (or scale) yield curves that are analogous to the usual "σ/τ" curve for the Allan variance. Users familiar with the Allan variance can thus readily interpret the wavelet variance. In particular, as is true for the Allan variance, the wavelet variance can be regarded as an octave-band estimate of the spectrum and hence does not suffer from the high variability of the windowed DFT. Because higher-order wavelets provide a better approximation to octave-band filters than does the Haar wavelet, it is easier to translate higher-order wavelet variances into reasonable spectral estimates.

Even though wavelets are a relatively new topic, there is already an enormous literature about them — see [4] and references therein. In what follows, we merely attempt to motivate the use of wavelet analysis for synchronization and timekeeping, with particular emphasis on the problem of leakage. Space precludes a full discussion of many important aspects of wavelet analysis (such as the existence of fast computational algorithms [9], procedures for determining confidence intervals for the wavelet variance [8], and the use of the scalogram as a diagnostic tool for monitoring oscillator stability in real time [7]).

II. Power-Law Noise Processes

It has often been claimed that, of all the physical measures, we can realize frequency or periodicity with the greatest accuracy. What we mean is that some basic periodic (repeating) event is very consistent in its recurrence so that, for example, it is independent of environmental influence. This event can thus be used to define consistent and repeatable intervals of time such as the second. Departures from this consistent recurrence are classed as noise. Perfectly recurrent noiseless events mean that events happen now exactly as happened before, that is, an observation now (the present moment) can be perfectly predicted based upon what has happened before (a past moment). Since reproduction is never quite exact, the degree with which present recurrence duplicates past recurrence indicates how well the events remember or duplicate themselves.

When the phase or time difference between two oscillators or clocks is measured as a function of the nominal time of the clocks, what we are measuring is the relative time deviation of the two clocks. Time deviation is generally modeled by two parts:

[1] a deterministic part quantified by a time offset, frequency offset, and frequency drift, and
[2] a random part quantified by various classes of noise processes.

Historically, power-law (or long memory) noise processes have played a vital role in characterizing the performance of clocks. The statement that the relative phase $\{x_t\}$ between a test clock and reference obeys a power-law noise process means that the power spectral density function $S_x(\cdot)$ for the process modelling $\{x_t\}$ is proportional to f^α for positive Fourier frequencies f. Correctly classifying a clock's power-law behavior, which is equivalent to determining α, is a primary objective of analysis techniques such as spectral analysis. Once the exponent α has been determined, we derive estimates of how a clock's timekeeping ability might evolve [11].

The spectrum of the residual time difference between two clocks or oscillators sometimes contains periodicities (such as from 50 or 60 Hz AC power) and always contains nonperiodic (stochastic) characteristics quantified as power-law processes. In timekeeping metrology, there are five commonly used models of power-law noise processes [11]: white PM ($S_x(f) \propto f^0$, a constant), flicker PM ($S_x(f) \propto f^{-1}$), white FM ($S_x(f) \propto f^{-2}$), flicker FM ($S_x(f) \propto f^{-3}$), and random walk FM ($S_x(f) \propto f^{-4}$). Examples of time series drawn from these five processes are shown in Fig. 1.

III. Narrowband vs. Broadband Processing

A common approach to estimating the power spectrum $S_x(\cdot)$ of phase residuals $\{x_t\}$ uses a digital processor to compute a windowed discrete Fourier transform (DFT) of $\{x_t\}$. Here "windowed DFT" refers to multiplying the phase residuals by a data window $\{h_t\}$ (sometimes called a data taper) to produce a windowed series $\{h_t x_t\}$, to which we then apply the DFT. The purpose of windowing is to reduce a potential bias known as leakage, in which power "leaks" from high power into low power portions of the spectrum, thus causing a significant positive bias in unwindowed spectral estimates. There is a

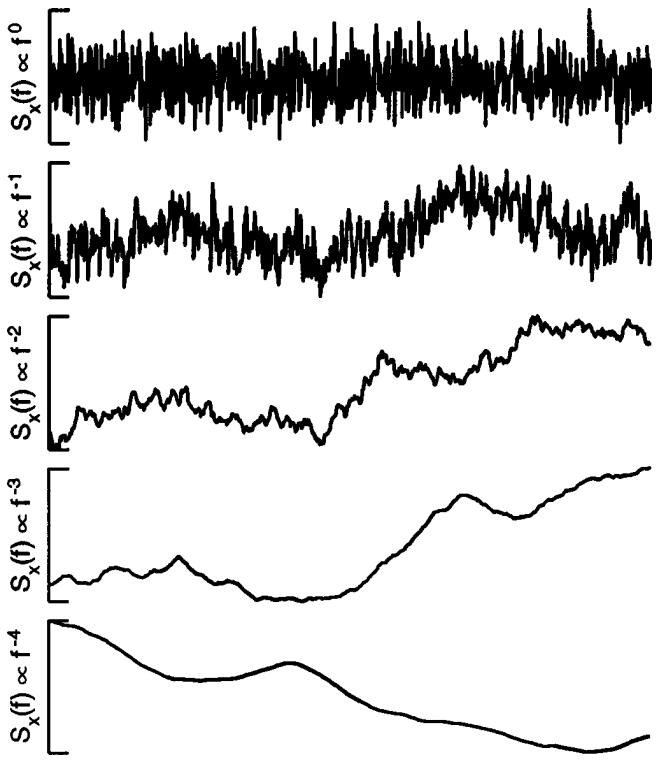


Fig. 1. Examples of time series that are portions of realizations from power-law processes $S_x(f) \propto f^\alpha$, with $\alpha = 0$, -1, -2, -3 and -4 (from top to bottom).

vast literature on the windowed DFT (see [10] for a review). Commercial digital spectrum analyzers typically compute a power spectral density estimate from a sampled varying voltage by converting the voltage to a time series $\{x_t\}$, windowing the series using a user-selected window $\{h_t\}$, and then taking the squared modulus of a properly normalized version of the DFT of $\{h_t x_t\}$. The windowed DFT is inherently narrowband and hence highly variable across frequencies, which makes straightforward interpretation of DFT-based spectral estimates somewhat problematic for the novice (these estimates can also be highly dependent on the choice of a particular window – see [12] for a discussion of this dependence in the case of power-law processes).

Because narrowband processing is not required for broadband processes such as power-law processes, time and frequency standards laboratories have alternatively handled power-law noise processes using the Allan variance [1] or in some cases a modified version of it [2]. These variances can be interpreted as the variance of a process after it has been subjected to an approximate bandpass filter of constant Q (the ratio of the center frequency of the analyzing filter to the width of the filter's pass band is constant [3]). The Allan variance can be used to construct a broadband spectral estimate using well-known conversion schemes [11], [6]. However, while broadband processing produces spectral estimates with inherently less variability than those of narrowband processing, both types of processing are subject to a bias known as leakage. Leakage has long been recognized as an important concern for narrowband spectral estimates (and in fact is the rationale for using windows), but its importance in broadband processing has not received much attention. As we argue below, one rationale for considering the wavelet variance is that higher-order wavelets effectively address the leakage problem.

IV. Wavelets and the "Scale Domain"

Suppose that $x_0, x_1, \ldots, x_{N-1}$ form a sequence $\{x_t\}$ of N time-ordered phase measurements. Let us define

$$\sum_{t=0}^{N-1} |x_t|^2 \equiv \mathcal{E}_x$$

to be the "energy" in our finite set of measurements. We can then trivially regard $|x_t|^2$ as the contribution to the energy $\mathcal{E}_x$ due to the component of $\{x_t\}$ with time index t. We can also regard $\{x_t\}$ as the "time domain" representation of our phase measurements.

Next, consider the discrete Fourier transform (DFT) of $\{x_t\}$, namely,

$$X_k \equiv \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x_t e^{-i2\pi f_k t}, \qquad k = 0, 1, \ldots, N-1,$$

where X_k is the kth DFT coefficient and is associated with the kth Fourier frequency $f_k \equiv k/N$. Parseval's theorem tells us that

$$\sum_{k=0}^{N-1} |X_k|^2 = \mathcal{E}_x.$$

Hence we can regard $|X_k|^2$ as the contribution to the energy $\mathcal{E}_x$ due to the component of $\{X_k\}$ with frequency index k, and we can regard $\{X_k\}$ as the "frequency domain" representation of our phase measurements. The time and frequency domain representations are equivalent in the sense that we can recover one given the other because of the inverse DFT, namely,

$$x_t = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{i2\pi f_k t}, \qquad t = 0, 1, \ldots, N-1.$$

As is true for the DFT, the discrete wavelet transform (DWT) of $\{x_t\}$ preserves the energy $\mathcal{E}_x$ in a set of

coefficients; however, unlike the DFT, these coefficients are not indexed by frequency, but rather doubly indexed by time shift j and "scale" τ. The DWT is defined in terms of a "mother wavelet" $\psi(\cdot)$ and an associated "scaling function" $\phi(\cdot)$, where $\psi(\cdot)$ can be any member of a large class of functions satisfying certain stringent conditions [4]. Assuming for convenience that $N = 2^p$ for some positive integer p, we define $\psi_{j,\tau}(\cdot)$ as a shifted and scaled version of $\psi(\cdot)$:

$$\psi_{j,\tau}(t) = \frac{1}{\sqrt{2\tau}} \psi\left(\frac{t}{2\tau} - j\tau\right),$$

where $\tau = 1, 2, 4, \ldots, N/2$ indexes a "power of 2" scale, while $j = 0, 2\tau, 4\tau, \ldots, N - 2\tau$ indexes shifts in time commensurate with scale τ. The DWT coefficients are the doubly indexed series $\{d_{j,\tau}\}$ defined by

$$d_{j,\tau} \equiv \sum_t x_t \psi_{j,\tau}(t)$$

along with $c \equiv \sum x_t \phi(t/N)/\sqrt{N}$ (depending upon the precise implementation of the DWT, c is typically proportional to either the average $\sum x_t/N$ of the sequence $\{x_t\}$ or a quantity that converges to the average as N gets large). Parseval's theorem tells us that

$$\sum_\tau \sum_j |d_{j,\tau}|^2 + |c|^2 = \mathcal{E}_x.$$

Hence we can regard $|d_{j,\tau}|^2$ as the contribution to the energy $\mathcal{E}_x$ due to the component of $\{d_{j,\tau}\}$ with time shift index j and scale index τ, and we can regard $\{d_{j,\tau}\}$ as the "scale domain" (or "time/scale" domain) representation of our phase measurements. This scale domain representation is fully equivalent to the time and frequency domain representations because of the inverse DWT, namely,

$$x_t = c + \sum_\tau \sum_j d_{j,\tau} \psi_{j,\tau}(t).$$

As an example of a scale domain representation, let us set our mother wavelet $\psi(\cdot)$ equal to the Haar wavelet $\psi^{(\text{Haar})}(\cdot)$, which we define here as

$$\psi^{(\text{Haar})}(t) = \begin{cases} -1, & 0 \le t < 1/2; \\ 1, & 1/2 \le t < 1; \\ 0, & \text{otherwise.} \end{cases}$$

The corresponding scaling function $\phi(\cdot)$ is given by

$$\phi^{(\text{Haar})}(t) \equiv |\psi^{(\text{Haar})}(t)|$$

(this simple relationship between the mother wavelet and the scaling function is unique to the Haar wavelet). For the Haar wavelet, we find that

$$d_{j,\tau} = \frac{1}{\sqrt{2\tau}} \left(\sum_{l=0}^{\tau-1} x_{(2j+2)\tau-1-l} - \sum_{l=0}^{\tau-1} x_{(2j+1)\tau-1-l} \right)$$
$$= \frac{\sqrt{\tau}}{\sqrt{2}} \left[\overline{x}_{(2j+2)\tau-1}(\tau) - \overline{x}_{(2j+1)\tau-1}(\tau) \right],$$

where $\overline{x}_t(\tau) \equiv \sum_{j=0}^{\tau-1} x_{t-j}/\tau$. Thus, at scale $\tau = 1$, we have $d_{j,1} = (x_{2j+1} - x_{2j})/\sqrt{2}$ for $0 \le j \le N/2 - 1$, while, at the largest scale $\tau = N/2$, we have the single coefficient

$$d_{0,N/2} = (x_{N-1} + \cdots + x_{N/2} - x_{N/2-1} - \cdots - x_0)/(\sqrt{2})^p.$$

Let us now define the wavelet variance for scale τ as

$$\sigma_x^2(\tau) \equiv \text{var}\,\{d_{j,\tau}\}/\tau.$$

Under the assumption that $E\{d_{j,\tau}\} = 0$ so that the variance of $d_{j,\tau}$ is equal to $E\{d_{j,\tau}^2\}$, an obvious estimator of this wavelet variance is

$$\tilde{\sigma}_x^2(\tau) = \frac{1}{\tau} \times \frac{1}{N/2\tau} \sum_{j=0}^{N/2\tau-1} d_{j,\tau}^2 = \frac{2}{N} \sum_{j=0}^{N/2\tau-1} d_{j,\tau}^2.$$

Specializing now to the case of the Haar wavelet, we find that

$$\tilde{\sigma}_x^2(\tau) = \frac{\tau}{N} \sum_{j=0}^{N/2\tau-1} \left[\overline{x}_{(2j+2)\tau-1}(\tau) - \overline{x}_{(2j+1)\tau-1}(\tau) \right]^2.$$

If the x_t's represented average fractional frequency deviations rather than phase measurements, then the above would be the well-known "nonoverlapped" estimator of the Allan variance. The Allan variance therefore corresponds to a wavelet variance when the Haar wavelet is used with average fractional frequency deviations. When viewed from the perspective of wavelets, the Allan variance is thus not a "time domain" quantity, but rather is a "scale domain" or "time/scale domain" quantity.

V. Determination of Power-Law Noise Types

As a function of time, two-oscillator phase deviations might look like one of the plots of a realization of a pure power-law process shown in Fig. 1. More realistically, these deviations resemble a linear combination of such processes, whose spectrum can be described mathematically as

$$S_x(f) = \sum_\alpha h_\alpha |f|^\alpha,$$

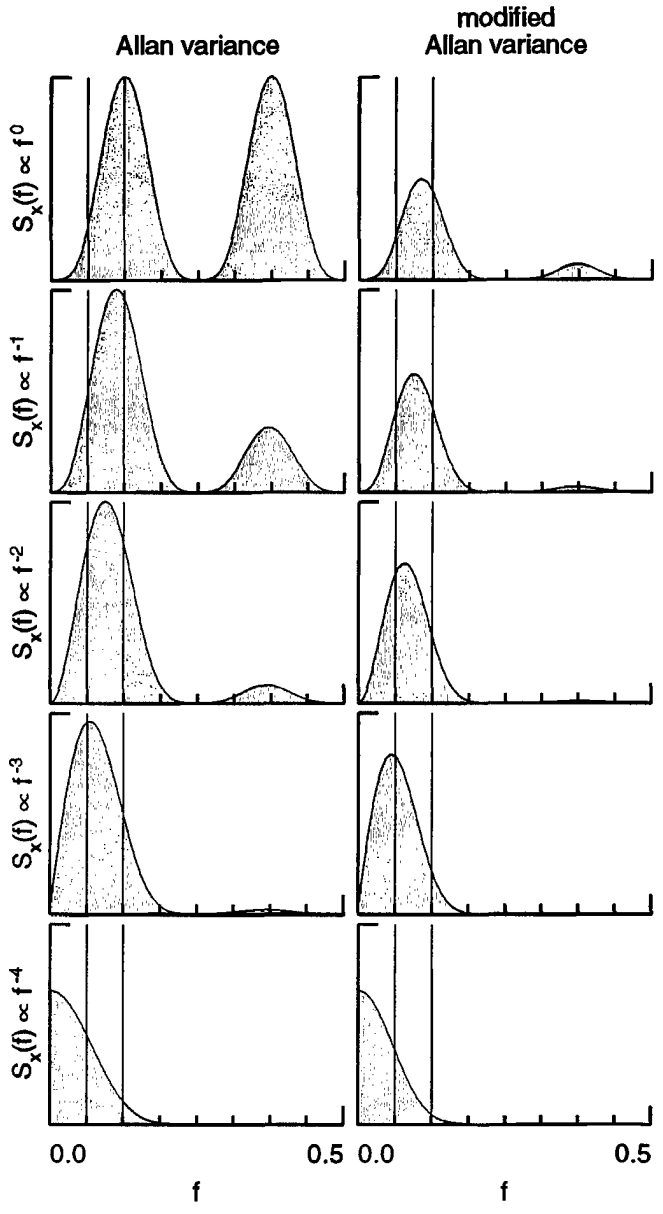


Fig. 2. Modulus squared of the transfer function for the Allan variance (left-hand column) and the modified Allan variance (right-hand column) times power-law spectra $S_x(\cdot)$ proportional to f^0 (top row), f^{-1} (second row), f^{-2} (third row), f^{-3} (fourth row) and f^{-4} (bottom row) for scale $\tau = 4$. The integrals of the shaded areas yield the Allan variance or modified Allan variance for scale $\tau = 4$.

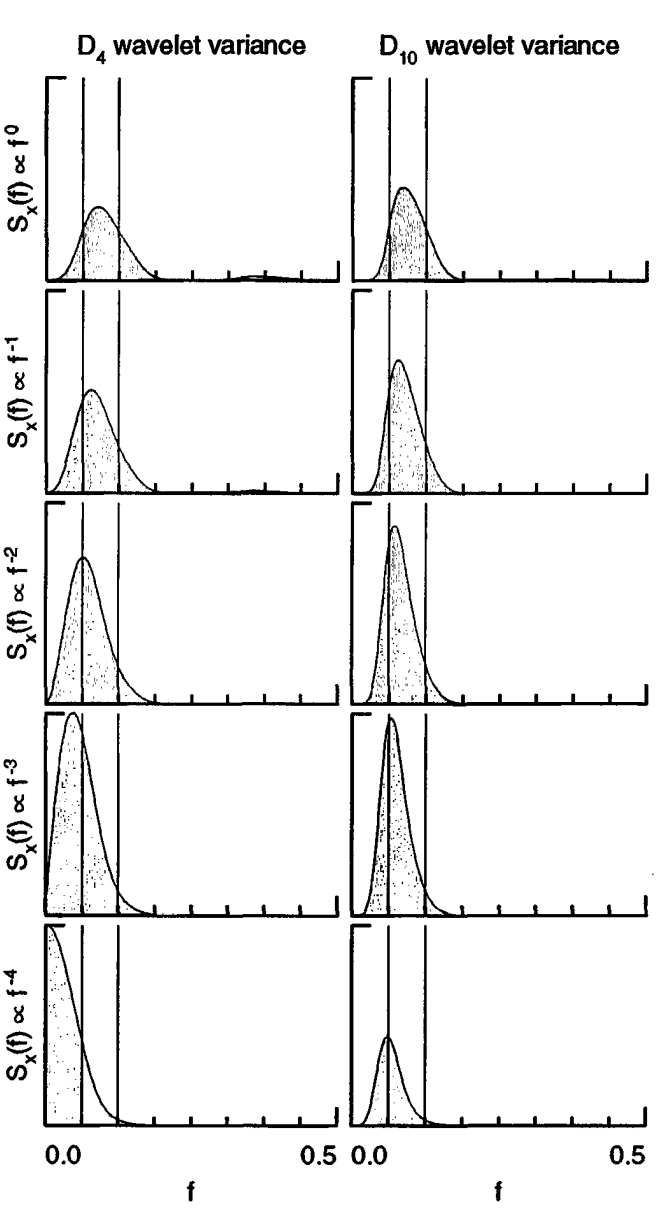


Fig. 3. Modulus squared of the transfer function for the D_4 wavelet variance (left-hand column) and the D_{10} wavelet variance (right-hand column) times the same set of power-law spectra shown in Fig. 2. The integrals of the shaded areas yield the D_4 or D_{10} wavelet variance for scale $\tau = 4$.

where the summation is over a finite number of different α's (usually a subset of $\alpha = 0, -1, -2, -3$ and -4), with h_α determining the relative contribution of the power-law process with exponent α. We refer to a process with the above spectrum as a composite power-law process. For pure power-law processes, there are well-known formulae for converting from the Allan variance to the frequency domain [11]. Here we argue that, for composite power-law processes, this conversion can become problematic for both the usual Allan variance and

the modified Allan variance. Let $\sigma_x^2(\tau)$ represent either of these variances (this notation should not be confused with similar notation for "TVAR" in [3]). We can then write

$$\sigma_x^2(\tau) = \int_{-1/2}^{1/2} \mathcal{F}_\tau(f) S_x(f)\, df,$$

where $\mathcal{F}_\tau(\cdot)$ is the modulus squared of the appropriate transfer function for the filters associated with these variances at scale τ [5]. The shaded areas in the plots of Fig. 2 show the product $\mathcal{F}_\tau(f)S_x(f)$ versus f for the Allan variance (left-hand column) and the modified Allan variance (right-hand column) for five pure power-law spectra and scale $\tau = 4$ (the power-law spectrum is constant in the top row of plots, so this row really shows just $\mathcal{F}_\tau(f)$ versus f). The integral of each shaded area gives $\sigma_x^2(4)$ for the appropriate pure power-law process. In the octave-band interpretation of these variances, either variance at scale τ should roughly reflect the power in the spectrum in the frequency interval $[1/4\tau, 1/2\tau]$. For $\tau = 4$, this interval is $[1/16, 1/8]$ and is delineated on each plot by a pair of thin vertical lines. If the filters associated with these variances were perfect octave-band filters, the shaded area in each plot would be entirely contained between the vertical lines. The amount of the shaded area that lies outside of the vertical lines represents the contribution to the Allan or modified Allan variance attributable to leakage. These plots indicate that there is substantial leakage for the Allan variance, but less so for the modified Allan variance. Leakage is most pronounced in the Allan variance for white PM ($S_x(f) \propto f^0$ in the top left-hand plot), a deficiency that in fact accounts for the development of the modified Allan variance [5]. If we now consider a composite power-law process dominated between the vertical lines by a power-law with a different exponent than the one displayed in the plots of Fig. 2, we can see the potential problem with leakage, namely, that the integral of $\mathcal{F}_\tau(f)S_x(f)$ (the Allan or modified Allan variance) can be influenced mainly by values of f outside of the vertical lines and hence cannot accurately reflect the values of $S_x(f)$ between the vertical lines.

Fig. 3 shows corresponding plots for the wavelet variance using the D_4 (left-hand column) and D_{10} (right-hand column) "extremal phase" wavelets [4]. The D_4 wavelet was chosen because it is "one order up" from the Haar wavelet (and in fact rather closely mimics the behavior of the modified Allan variance), while the D_{10} wavelet is an example of a higher-order wavelet. The main point to notice here is that the D_{10} wavelet variance for scale $\tau = 4$ reflects the spectrum in the passband $[1/16, 1/8]$ to a much better degree than the other variances because the shaded areas are concentrated between the vertical lines to a higher degree for the D_{10} wavelet variance than for the Allan or modified Allan variances.

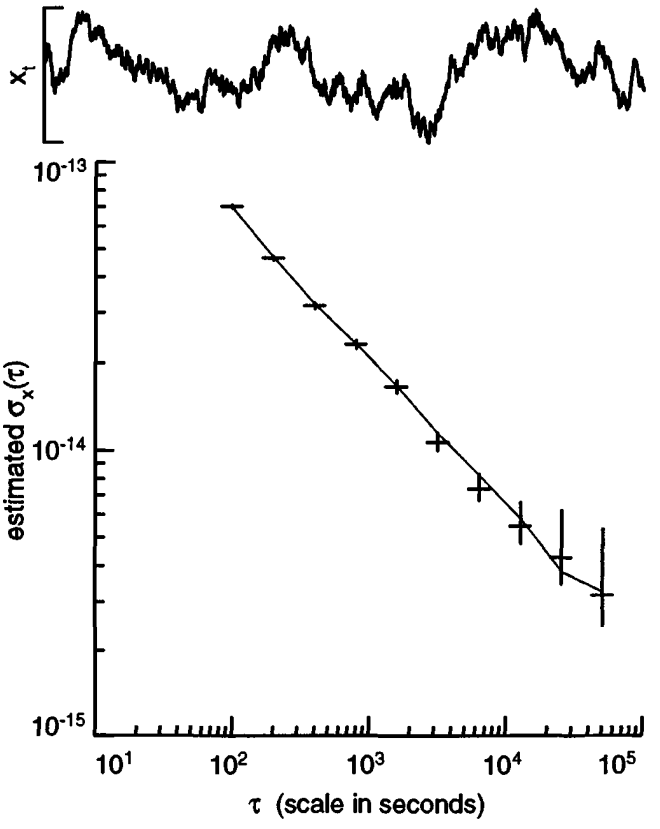


Fig. 4. NIST–7 vs. hydrogen maser phase measurements (top plot) and estimated $\sigma_x(\tau)$ versus τ (bottom plot) for the Allan variance (connected curve) and the D_4 wavelet variance (crosses).

VI. Examples

We present two examples of the limited tests we have made to date using the wavelet variance with phase measurements (see [9] for an example involving geophysical data, for which the D_8 wavelet variance performed considerably better than the Allan variance). The top plot of Fig. 4 shows phase measurements recorded every 100 seconds over a 3.7 day interval comparing NIST–7 to a hydrogen maser. The bottom plot shows the estimated Allan standard deviation (which is just the square root of the Allan variance) versus scale τ (the connected curve) and also the estimated D_4 wavelet standard deviation versus τ (the crosses). The center of each cross indicates the appropriate D_4 estimate, whereas the vertical portion of the cross delineates a "one sigma" (68.3%) confidence interval for the true D_4 wavelet standard deviation [8]. The Allan and D_4 wavelet standard deviations agree fairly well here, although there are two scales

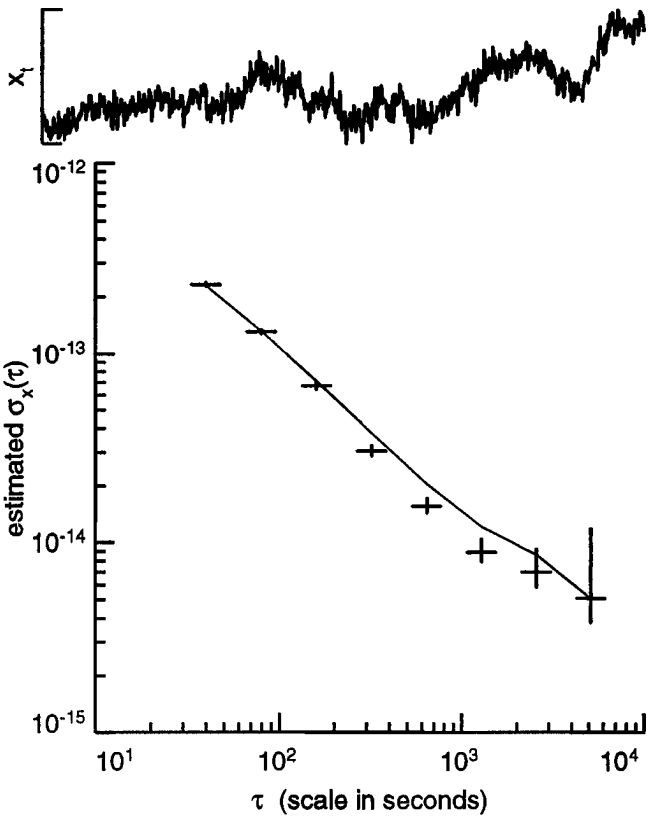


Fig. 5. Time-synchronization phase measurements using the NIST satellite two-way transfer modem configured in an in-cabinet loop test (top plot) and estimated $\sigma_x(\tau)$ versus τ (bottom plot) for the Allan variance (connected curve) and the D_4 wavelet variance (crosses).

(τ = 3200 and 6400 seconds) for which the Allan standard deviation is just inside the "one sigma" confidence limits for the D_4 wavelet standard deviation. Use of the D_4 wavelet here tells us that leakage is not a major problem with the Allan variance for this set of phase measurements.

The top plot of Fig. 5 shows phase measurements recorded every 40 seconds over a half day interval reflecting time-synchronization using the NIST satellite two-way transfer modem configured in an in-cabinet loop test. The bottom plot here shows the same quantities as in the bottom plot of Fig. 4. While the Allan and D_4 standard deviations agree quite well in the smallest three and largest scales shown, there is significant difference in the middle three scales; moreover, the difference is consistent with an interpretation of leakage in the Allan variance (because the Allan variance is higher than the D_4 variance). Use of the higher-order D_6 wavelet yields good agreement with the D_4 wavelet.

References

[1] D.W. Allan, "Statistics of Atomic Frequency Standards," *Proc. IEEE*, vol. 54, pp. 221–230, 1966.

[2] D.W. Allan and J.A. Barnes, "A Modified 'Allan Variance' with Increased Oscillator Characterizaton Ability," *Proc. 35th Ann. Frequency Control Symp.*, pp. 470–475, 1981 (reprinted in [11]).

[3] D.W. Allan, M.A. Weiss, and J.L. Jespersen, "A Frequency-Domain View of Time-Domain Characterizaton of Clocks and Time and Frequency Distribution System," *Proc. 45th Ann. Frequency Control Symp.*, pp. 667–678, 1991.

[4] I. Daubechies, *Ten Lectures on Wavelets.* Philadelphia: SIAM, 1992.

[5] C.A.Greenhall,"A Shortcut for Computing the Modified Allan Variance," *Proc. 46th Ann. Frequency Control Symp.*, pp. 262–264, 1992.

[6] D.B. Percival, "Characterization of Frequency Stability: Frequency-Domain Estimation of Stability Measures," *Proc. IEEE*, vol. 79, pp. 961–972, 1991.

[7] D.B. Percival, "An Introduction to Spectral Analysis and Wavelets," in *Advanced Mathematical Tools in Metrology*, (Edits P. Ciarlini, M. Cox, R. Monaco and F. Pavese). Singapore: World Scientific, pp. 175—186, 1994.

[8] D.B. Percival, "On Estimation of the Wavelet Variance," submitted to *Biometrika*, 1994.

[9] D.B. Percival and P. Guttorp, "Long Memory Processes, the Allan Variance and Wavelets," in *Wavelets in Geophysics*, (Edits E. Foufoula-Georgiou and P. Kumar). Boston: Academic Press, pp. 325–343, 1994.

[10] D.B. Percival and A.T. Walden, *Spectral Analysis for Physical Applications.* Cambridge, U.K.: Cambridge Univ. Press, 1993.

[11] D.B. Sullivan, D.W. Allan, D.A. Howe, and F.L. Walls (Edits), *Characterization of Clocks and Oscillators*, NIST Technical Note 1337, 1990 (available from NIST, 325 Broadway, Boulder, CO 80303–3328).

[12] F.L. Walls, D.B. Percival, and W.R. Ireland, "Biases and Variances of Several FFT Spectral Estimators as a Function of Noise Type and Number of Samples," *Proc. 43rd Ann. Frequency Control Symp.*, pp. 336–341, 1989 (reprinted in [11]).

1994 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ACCUMULATION OF RANDOM NOISE IN A CHAIN OF SLAVE CLOCKS

G. M. Garner

AT&T Bell Laboratories
101 Crawfords Corner Rd., Room 2G-533A
Holmdel, NJ 07733

Abstract

This paper considers random noise accumulation in a chain of clocks using a time-domain, state-space approach. In this configuration, the output phase signal of one clock is the input phase signal to the next clock. The problem is relevant to synchronization distribution is SONET and SDH networks.

Both linear and nonlinear phase-locked loop models with additive noise sources are considered. For the latter, the particular nonlinearity considered is phase detector quantization. Simulation results for Time Variance (TVAR) are obtained in all cases. Explicit analytic results for the asymptotic behavior of TVAR are obtained for the linear case.

For short time intervals, TVAR is proportional to time interval and is independent of the number of clocks in the chain and the quantization size. In addition, for short time intervals TVAR increases with the square of bandwidth for phase detector white phase modulation and is independent of bandwidth for VCO white phase modulation and white frequency modulation. For long time intervals, TVAR is inversely proportional to time interval for phase detector white phase modulation and inversely proportional to the square of time interval for VCO white phase modulation and white frequency modulation. In addition, for long time intervals TVAR increases with the number of clocks in the chain; the increase is linear for phase detector white phase modulation and VCO white frequency modulation. Finally, for long time intervals TVAR increases when phase detector quantization is introduced and as the quantization size is increased, for the uniform phase detector quantization considered here.

Introduction

This paper considers random noise accumulation in a chain of clocks. In this configuration, the output phase signal of one clock is the input phase signal to the next clock; timing is passed from the initial clock in the chain through the intermediate clocks to the final clock. Each clock contains various sources of random noise. Due to these sources, noise is added to the timing signal at each clock in the chain. The noise is then filtered by succeeding clocks in the chain. The overall objective is to maintain synchronization of all the clocks in the chain, i.e., to minimize the phase difference between any two clocks in the chain. The phase differences depend on the type and strength of each noise source and on the various clock parameters.

This problem is relevant to synchronization of digital telecommunications networks, where a variety of clocks in different geographic locations must be synchronized. In the evolving synchronous digital networks based on the Synchronous Optical Network (SONET) [1], [2], [3] or Synchronous Digital Hierarchy (SDH) [4], [5], [6] architectures, digital signals at all levels must be synchronized.[1] Typically, timing is distributed from a high quality (i.e., very accurate) clock, referred to as a Primary Reference Clock (PRC), through chains of lower quality (i.e., less accurate) clocks. One method for doing this in SONET or SDH networks is to transport the timing over the digital facilities via the respective SONET or SDH signals. The philosophy and architecture for network synchronization are different for SONET and SDH. A brief description for each case follows (more detail may be found in the references).

In a SONET network, synchronization distribution is hierarchical [7]. Timing for each office is provided by a single clock, referred to as the Building Integrated Timing Supply (BITS) clock. If the office contains a PRC, the BITS clock receives timing from the PRC. If the office does not contain a PRC, the BITS clock receives timing from another office whose BITS clock has the same or higher quality. This interoffice timing

1. This is in contrast to existing asynchronous digital networks, where only the lowest level digital signals (1.544 Mbps or 2.048 Mbps) must be synchronized.

0-7803-1945-1/94 $3.00 © 1994 IEEE

distribution is done via SONET facilities. There is no clock in the office whose quality is higher than that of the BITS clock, except for the PRC if the office contains one. The BITS clock supplies timing to all the network elements (e.g., digital multiplexers, digital cross-connect systems, etc.) in the office, which in turn time the respective SONET facilities. The Network Element (NE) clocks typically are of lower quality and have wider bandwidth than the BITS clock. Note that in a timing chain, there is at most one wider bandwidth NE clock between two narrower bandwidth BITS clocks. Typical bandwidths are 0.1 Hz for the SONET NE clock [8], [2], [3] and 0.01 Hz for the BITS clock [9], [10]. In addition, timing chains tend to be rather short. In one Interexchange Carrier (IEC) Network the number of BITS clocks in a chain is limited to two [11], i.e., every office either contains a PRC or receives timing from an office that contains a PRC. In general, it is expected for arbitrary networks that the number of BITS clocks in a timing chain should rarely exceed five or six.

In an SDH network, synchronization distribution is not hierarchical and timing chains can be much longer (by a factor of 10 or more) than in SONET networks [12]. Timing originates at a PRC and is supplied directly to a slave clock, or Synchronization Supply Unit (SSU). The SSU is analogous to a BITS clock, although it generally has different requirements. The SSU supplies timing to another SSU (in another office) through a chain of one or more SDH Equipment Clocks (SECs). Each SEC times an SDH NE and respective facilities. The SEC is analogous to the SONET NE clock, although it has different requirements. In addition, the SECs between two successive SSUs need not be in the same office or in the offices of either of the SSUs. It is not required that every office have an SSU (in contrast to SONET networks, where every office has a BITS clock). The SECs typically are of lower quality and have wider bandwidth than the SSUs. Typical bandwidths are 1-10 Hz for the SEC [13], [14] and 0.01 Hz for the SSU. Timing chains are limited to 10 SSUs, 60 SECs, and 20 SECs between any two SSUs [12].

Note that SONET and SDH networks may have multiple PRCs. In this case, the phase difference between the outputs of two clocks whose timing is traceable to different PRCs depends on both noise accumulation in the two chains and phase differences between the two PRCs. A schematic of a SONET or SDH timing chain is shown in Figure 1.

In this work, analytic and simulation results for random noise accumulation in a chain of clocks are obtained. Both linear and nonlinear phase-locked loop (PLL) models with additive noise sources are considered, although for the latter only simulation results are obtained. The linear model is a simplified version of the model in [15].[2] A time domain, state space model is derived for the chain of PLLs, where the state variables are the phase output and derivative of phase output (i.e., frequency output) for each clock (in addition, state variables are added to model some of the noise sources). For the linear case, the steady-state autocorrelation function for the state vector is obtained; from this the time variance (TVAR) and time deviation (TDEV) stability parameters (the latter is the square root of the former) may be obtained at any point in the chain. TVAR and TDEV are used to characterize the stability of a timing signal in a telecommunications network [17]. Explicit analytic results are given for the asymptotic behavior of TVAR and TDEV (for $\tau \to \infty$ and $\tau \to 0$). Simulation results for TDEV are also given.

The nonlinear PLL model considered here represents the specific nonlinearity of quantization in the phase detector (PD). Only the very simple case of fixed, uniform quantization is considered. The effect of this quantization is similar, but quantitatively different, to that of additive noise. Simulation results are given for TDEV.

The analytic and simulation models are used to study the dependence of TVAR and TDEV on clock bandwidth and number of clocks in the chain, for specific noise sources with and without PD quantization. Results are given for each noise source individually; the behavior of TVAR and TDEV is not necessarily the same for all the noise sources.

Analytic models for the linear case were developed in [18] and [19] using a frequency domain approach (these are based on previous models for chains of clocks or regenerators given in [20], [21], and [22]). In [18], total mean-square phase output (at the end of the chain) is given, without regard to time interval. In [19], root-mean-square Time Interval Error (TIErms) is given. Both references concentrate on the SDH reference chain (i.e., narrow-band clocks separated by some number of wide-band clocks). In addition, the eventual calculations

2. In this paper, the term "linear phase-locked loop" is used for a phase locked loop that operates in a region where it can be represented by a linear model with a continuous state space and additive noise. This is different from the usage in [16], where "linear phase-locked loop" refers to an analog PLL with a product phase detector.

in both references are numerical; no explicit analytic results are given. In [23], simulation and measurement results are given for both the linear case and cases involving PD quantization. Details of the quantization are not given, except for the statement that the quantization is nonlinear. Since the results there show a decrease in TDEV for cases with quantization compared to otherwise identical cases without quantization, it is presumed that the quantization is not the simple, fixed and uniform type considered here.[3] In addition, [23] considers only random-walk frequency modulation VCO noise and no PD noise, in contrast to the noise sources considered here (see the next section). The results in [23] were all obtained via simulation or measurement; no analytic results are given.

The work done in the present paper builds upon the above previous work. Analytic results for asymptotic behavior of TVAR and TDEV are given; these are believed to be new. Results are given in terms of TDEV rather than TIErms as in [19]. A more extensive set of simulation results for TDEV are given (for somewhat different cases) compared to [23].

The paper is organized as follows. In the next section, the state-space model is developed for the linear and nonlinear cases. In the section after that, TVAR is related analytically to the state-space model for the linear case and asymptotic results are given. Simulation results for the linear and nonlinear cases are given in the following section. The work is summarized and conclusions are given in the final section.

State-Space Model

Consider a chain of N clocks, each modeled as a PLL (In Figure 1, $N = P(R + 1)$). Two successive clocks, labeled j and $j+1$, are shown in Figure 2. The clock model is the same as the model in [19], which is based on the model in [15]. The output phase for clock j, ϕ_O^j, is equal to the input phase for clock $j+1$, ϕ_I^{j+1} (note that $\phi_I^1 = 0$, i.e., the first clock in the chain has no input reference). The effect of the transmission link between two successive clocks is neglected.[4] In this work, we are interested in the effect of the clocks rather than the transmission links. The loop filter, $F^j(s)$ is assumed to be an active (i.e., high-gain) filter[5] with transfer function

$$F^j(s) = 1 + \frac{a^j}{s}, \tag{1}$$

where a^j is a filter parameter (all variables and parameters are indexed with the superscript j to indicate the j^{th} clock in the chain). For the linear case, the transfer characteristic for the j^{th} phase detector is assumed to be

$$e^j(t) = K_d^j[\phi_I^j(t) - \phi_O^j(t)], \tag{2}$$

where e^j is the phase detector output. For phase detectors with quantization, the transfer function is assumed to be

$$e^j(t) = K_d^j \left\lfloor \frac{\phi_I^j(t) - \phi_O^j(t)}{q^j} \right\rfloor q^j, \tag{3}$$

where q^j is the size of the quantization for the j^{th} phase detector ($q^j \neq 0$).

There are two additive noise sources, n_D^j and n_{VCO}^j, which represent noise present in the phase detector and loop filter and in the voltage controlled oscillator (VCO), respectively. As described in [19], the VCO noise consists of white phase modulation (WPM), flicker phase modulation (FPM), white frequency modulation (WFM), and flicker frequency modulation (FFM) sources. The phase detector/loop filter noise consists of WPM and FPM sources. The VCO and PD power spectral densities are

$$S_{VCO}^j(f) = \frac{b_{-1}^j f_0^4}{f^3} + \frac{b_0^j f_0^4}{f^2} + \frac{b_1^j}{f} + b_2^j \tag{4}$$

$$S_D^j(f) = \frac{c_1^j}{f} + c_2^j, \tag{5}$$

where f_0 is the oscillator center frequency and the constants b_n^j and c_n^j, which represent noise strengths, are given in [15] and [19]. It is pointed out in [19] that the transfer function between the VCO noise and PLL output is of a high-pass nature (i.e., the transfer function approaches one for large frequencies and zero for small frequencies) and, therefore, the VCO WPM noise must be band limited. It should be noted that this is also true for the VCO FPM component. This band limiting is accomplished here by passing the VCO WPM and FPM components through first-order, low-pass filters with time constants τ_{wj} and τ_{fj}, respectively.

3. In the present work, an increase in TDEV was observed in all cases when the simple, fixed and uniform quantization was considered.
4. The transfer function of the link is assumed to be unity and noise added by the link is assumed to be zero.
5. A general lead-lag filter could have been assumed; this would have made the state equations slightly more complicated. It is felt that the case of an active filter is the most important case.

Linear Analytic Model

Derivation of State Equations The input-output transfer characteristic for a single clock can be obtained in a straightforward manner; the result is

$$\Phi_O^j(s) = H^j(s)\Phi_I^j(s) + \frac{H^j(s)N_{DF}^j(s)}{K_d^j} +$$

$$H_{VCO}^j(s)\,[N_{VCO,FFM}^j(s) + N_{VCO,WFM}^j(s)] +$$

$$\frac{H_{VCO}^j(s)\,N_{VCO,WPM}^j(s)}{\tau_{wj}s + 1} +$$

$$\frac{H_{VCO}^j(s)\,N_{VCO,FPM}^j(s)}{\tau_{fj}s + 1}, \tag{6}$$

where

$$H^j(s) = \frac{2\zeta_j\omega_{nj}s + \omega_{nj}^2}{s^2 + 2\zeta_j\omega_{jn}s + \omega_{nj}^2}, \tag{7}$$

$$H_{VCO}^j(s) = \frac{s^2}{s^2 + 2\zeta_j\omega_{jn}s + \omega_{nj}^2}, \tag{8}$$

$\omega_{nj}^2 = K_o^j K_d^j a^j$, $\zeta_j = K_o^j K_d^j/2\omega_{nj}$, and the $\Phi(s)$ and $N(s)$ terms are Laplace transforms of the phases and noise sources, respectively. This transfer characteristic is equivalent to the following differential equation for $\phi_O^j(t)$:

$$\ddot{\phi}_O^j + 2\zeta_j\omega_{nj}\dot{\phi}_O^j + \omega_{nj}^2\phi_O^j = 2\zeta_j\omega_{nj}\dot{\phi}_I^j + \omega_{nj}^2\phi_I^j +$$

$$f_0^4[b_{-1}^j\dot{f}_a^j + b_0^j\dot{w}_a^j] + b_1^j\ddot{z}_f^j + b_2^j\ddot{z}_w^j +$$

$$\frac{2\zeta_j\omega_{nj}(c_1^j\dot{f}_c^j + c_2^j\dot{w}_c^j)}{K_d} + \frac{\omega_{nj}^2(c_1^j f_c + c_2^j w_c)}{K_d}, \tag{9}$$

with

$$\tau_{wj}\dot{z}_w^j + z_w^j = w_b^j \tag{10}$$

$$\tau_{fj}\dot{z}_f^j + z_f^j = f_b^j. \tag{11}$$

In Eqs. (9)-(11), the quantities w_a, w_b, and w_c are independent white noise sources and represent the VCO WFM, VCO WPM, and PD WPM sources (note that WFM is the integral of WPM). Likewise, the quantities f_a, f_b, and f_c are independent flicker (i.e., $1/f$) noise sources and represent the VCO FFM, VCO FPM, and PD FPM sources (note that FFM is the integral of FPM).

Using standard techniques (see, for example, Section 2.5 of [24]), a set of state equations, equivalent to Eqs. (9)-(11), may be derived. Note that the single output variable is the output phase, ϕ_O^j. The input variables are the input phase, ϕ_I^j, and the various noise sources. Let the output phase, its derivative (i.e., the instantaneous output frequency), and the outputs of the VCO WPM and VCO WFM band-limiting filters (i.e., z_w^j and z_f^j, respectively) be chosen as states, i.e., the state vector, $\mathbf{x}^j$, is

$$\mathbf{x}^j = [\phi_O^j \;\; \dot{\phi}_O^j \;\; z_w^j \;\; z_f^j]^T. \tag{12}$$

Then the state and output equations are

$$\dot{\mathbf{x}}^j = \mathbf{A}^j\mathbf{x}^j + \mathbf{B}^j\mathbf{u}^j + \mathbf{B}_w^j\mathbf{w}^j + \mathbf{B}_f^j\mathbf{f}^j \tag{13}$$

$$\mathbf{y}^j = \mathbf{C}\mathbf{x}^j, \tag{14}$$

where

$$\mathbf{A}^j = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\omega_{nj}^2 & -2\zeta_j\omega_{nj} & b_2^j/\tau_{wj}^2 & b_1^j/\tau_{fj}^2 \\ 0 & 0 & -1/\tau_{wj} & 0 \\ 0 & 0 & 0 & -1/\tau_{fj} \end{bmatrix} \tag{15}$$

$$\mathbf{B}^j = [2\zeta_j\omega_{nj} \;\; \omega_{nj}^2(1 - 4\zeta_j^2) \;\; 0 \;\; 0]^T \tag{16}$$

$$\mathbf{u}^j = [\phi_I^j] \tag{17}$$

$$\mathbf{B}_w^j = [\mathbf{B}_{w1}^j \;\; \mathbf{B}_{w2}^j \;\; \mathbf{B}_{w3}^j] \tag{18}$$

$$\mathbf{B}_{w1}^j = (f_0^4 b_0^j)\,[1 \;\; -2\zeta_j\omega_{nj} \;\; 0 \;\; 0]^T \tag{19}$$

$$\mathbf{B}_{w2}^j = \frac{[b_2^j \;\; -b_2^j(2\zeta_j\omega_{nj} + 1/\tau_{wj}) \;\; 1 \;\; 0]^T}{\tau_{wj}} \tag{20}$$

$$\mathbf{B}_{w3}^j = \frac{c_2^j\,[2\zeta_j\omega_{nj} \;\; \omega_{nj}^2(1 - 4\zeta_j^2) \;\; 0 \;\; 0]^T}{K_d^j} \tag{21}$$

$$\mathbf{w}^j = [w_a^j \;\; w_b^j \;\; w_c^j]^T \tag{22}$$

$$\mathbf{B}_f^j = [\mathbf{B}_{f1}^j \;\; \mathbf{B}_{f2}^j \;\; \mathbf{B}_{f3}^j] \tag{23}$$

$$\mathbf{B}_{f1}^j = (f_0^4 b_{-1}^j)\,[1 \;\; -2\zeta_j\omega_{nj} \;\; 0 \;\; 0]^T \tag{24}$$

$$\mathbf{B}_{f2}^j = \frac{[b_1^j \;\; -b_1^j(2\zeta_j\omega_{nj} + 1/\tau_{fj}) \;\; 1 \;\; 0]^T}{\tau_{fj}} \tag{25}$$

$$\mathbf{B}_{f3}^j = \frac{c_1^j\,[2\zeta_j\omega_{nj} \;\; \omega_{nj}^2(1 - 4\zeta_j^2) \;\; 0 \;\; 0]^T}{K_d^j} \tag{26}$$

$$\mathbf{f}^j = [f_a^j \;\; f_b^j \;\; f_c^j]^T \tag{27}$$

$$\mathbf{y}^j = [\phi_O^j] \tag{28}$$

$$\mathbf{C} = [1 \;\; 0 \;\; 0 \;\; 0]. \tag{29}$$

It may be verified that the state and output equations

(Eqs. (13)-(14)) are equivalent to the differential equations (Eqs. (9)-(11)) by evaluating the transfer matrices $\mathbf{C}(s\mathbf{I} - \mathbf{A}^j)^{-1}\mathbf{B}^j$, $\mathbf{C}(s\mathbf{I} - \mathbf{A}^j)^{-1}\mathbf{B}_w^j$, and $\mathbf{C}(s\mathbf{I} - \mathbf{A}^j)^{-1}\mathbf{B}_f^j$, and the individual transfer functions between the inputs and outputs of Eqs. (9)-(11). In fact, the elements of $\mathbf{B}^j$, $\mathbf{B}_w^j$, and $\mathbf{B}_f^j$ were determined by requiring equality of the former and latter transfer functions; see [24] for details.

For a chain of N clocks, Eqs. (13)-(14) are repeated N times (once for each clock). The equations may be combined into a single system of $4N$ state equations by using the condition that the input phase of clock j is the output phase of clock $j-1$, i.e., $\phi_I^j = \phi_O^{j-1}$ (and noting that clock 1 has no input, i.e., $\phi_I^1 = 0$). Define the $4N\times1$ state, white noise, and flicker noise vectors

$$\mathbf{x} = [(\mathbf{x}^1)^T \ (\mathbf{x}^2)^T \ \cdots \ (\mathbf{x}^N)^T]^T. \tag{30}$$

$$\mathbf{w} = [(\mathbf{w}^1)^T \ (\mathbf{w}^2)^T \ \cdots \ (\mathbf{w}^N)^T]^T. \tag{31}$$

$$\mathbf{f} = [(\mathbf{f}^1)^T \ (\mathbf{f}^2)^T \ \cdots \ (\mathbf{f}^N)^T]^T. \tag{32}$$

Then the state equations for the chain of N clocks are

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}_w\mathbf{w} + \mathbf{B}_f\mathbf{f}. \tag{33}$$

The matrix $\mathbf{A}$, whose dimension is $4N\times4N$, is conveniently specified in terms of N^2 4×4 submatrices:

$$\mathbf{A} = [\mathbf{A}_{ij}] \tag{34}$$

$$\mathbf{A}_{ii} = \mathbf{A}^i \tag{35}$$

$$\mathbf{A}_{i+1,i} = \mathbf{B}^i\mathbf{C} \tag{36}$$

$$\mathbf{A}_{ij} = \mathbf{0} \quad (i \neq j, j+1), \tag{37}$$

with $\mathbf{A}^i$, $\mathbf{B}^i$, and $\mathbf{C}$ given by Eqs. (15), (16), and (29), respectively. The matrix $\mathbf{B}_w$, whose dimension is $4N\times3N$, is conveniently specified in terms of N^2 4×3 submatrices:

$$\mathbf{B}_w = [(\mathbf{B}_w)_{ij}] \tag{38}$$

$$(\mathbf{B}_w)_{ii} = \mathbf{B}_w^i \tag{39}$$

$$(\mathbf{B}_w)_{ij} = \mathbf{0} \quad (i \neq j), \tag{40}$$

with $\mathbf{B}_w^i$ given by Eqs. (18)-(21). The matrix $\mathbf{B}_f$ is given by equations analogous to Eqs. (38)-(40) with the subscript w replaced by f.

Eigenvalues and Stability The eigenvalues of the system are obtained by straightforward evaluation of the characteristic polynomial of $\mathbf{A}$. The result is

$$\det(s\mathbf{I} - \mathbf{A}) = \prod_{j=1}^{N}(s^2 + 2\zeta_j\omega_{nj}s + \omega_{nj}^2)$$

$$(s + 1/\tau_{wj})(s + 1/\tau_{fj}). \tag{41}$$

The characteristic polynomial for the chain of clocks factors into the product of the characteristic polynomials for the individual clocks; this is due to the banded structure of $\mathbf{A}$. Then the eigenvalues for the system are the same as the eigenvalues for the individual clocks and band-limiting VCO WPM and FPM filters. For the case where the clocks are overdamped (the usual case), these are given by (for $j = 1, 2, \ldots, N$)

$$\lambda_{1j} = [-\zeta_j + (\zeta_j^2 - 1)^{1/2}]\omega_{nj} \tag{42}$$

$$\lambda_{2j} = [-\zeta_j - (\zeta_j^2 - 1)^{1/2}]\omega_{nj} \tag{43}$$

$$\lambda_{3j} = -1/\tau_{wj} \tag{44}$$

$$\lambda_{4j} = -1/\tau_{fj}. \tag{45}$$

Note that all the eigenvalues are real and negative; i.e., the system is stable.

It can be shown that the system is diagonalizable, i.e., has N distinct eigenvectors, if and only if the clocks are distinct. If a clock, and therefore its eigenvalues, is repeated M times in the chain, the two eigenvalues given by Eqs. (43) and (44) each correspond to a single eigenvector (the other two eigenvalues each correspond to M linearly independent eigenvectors. In this case, the Jordan form of $\mathbf{A}$ contains two Jordan blocks of size M corresponding to the two eigenvalues. Note that there still exists a similarity transformation that transforms $\mathbf{A}$ to its Jordan form.

Linear and Nonlinear Simulation Models

Simulation models were developed for the chain of clocks given by Figure 2, with the phase detector specified by Eqs. (2) or (3) and the loop filter by Eq. (1). An implicit (i.e., backward difference) integration scheme was used for the loop filter, while an explicit integration scheme (i.e., forward difference) was used for the VCO. Although the stability of implicit schemes is more robust as the integration time step is increased, the advantage of of using an implicit scheme for the VCO is offset by the additional computation required in determining the state in the nonlinear case.

Gaussian white noise (i.e., WPM) was simulated by generating a stream of independent, Gaussian random numbers (techniques for doing this are described in [25]). WFM was simulated by integrating (explicitly)

WPM. The relationship between the discrete time white noise variance and equivalent continuous time white noise variance parameter and power spectral density is [26]

$$\sigma^2_{\text{discrete}} = \frac{\sigma^2_{\text{continuous}}}{\Delta t} = \frac{b_2}{2\Delta t}, \qquad (46)$$

where Δt is the integration time step and b_2 is the magnitude of the power spectral density (see Eq. (4)). In Eq. (46), the units of b_2, $\sigma^2_{\text{continuous}}$ and $\sigma^2_{\text{discrete}}$, are compatible; e.g., if the former has units rad^2/Hz then the latter two have units rad^2/Hz and rad^2, respectively. Flicker noise (both FPM and FFM) was not included in the simulation model as it was also omitted from the analytical determination of TDEV (see the next section). The VCO WPM band-limiting filter was simulated using an implicit integration scheme.

TVAR and TDEV

TVAR and TDEV are obtained by first calculating the covariance and autocorrelation matrices for the state $\mathbf{x}$; TVAR is easily related to the latter. However, if the flicker noise terms are present in the state equations (Eq. (33)), the process $\mathbf{x}(t)$ can, in general, be nonstationary. In this case, the equation for the covariance matrix is a matrix differential rather than algebraic equation. Therefore, to simplify the analysis and allow the steady-state, algebraic equation for the covariance matrix to be used, the flicker noise terms are neglected. The inclusion of flicker noise in this type of analysis is an area for future work.

If the flicker noise terms are not present, the steady-state covariance matrix $\mathbf{X}$ for the state is given by the solution to the steady-state Lyapunov Equation [26]

$$\mathbf{AX} + \mathbf{XA}^T + \mathbf{B}_w\mathbf{Q}_w\mathbf{B}_w^T = \mathbf{0}, \qquad (47)$$

where $\mathbf{X} \triangleq \lim_{t\to\infty} E[\mathbf{x}(t)\mathbf{x}^T(t)] = \mathbf{X}^T$ is the steady state-covariance matrix and $\mathbf{Q}_w$ is a $3N\times3N$ diagonal matrix whose diagonal entries are the strengths of the $3N$ white noise sources (these are related to the power spectral density by Eq. (46)). In addition, the autocorrelation matrix $\mathbf{R}$ is given by [26]

$$\mathbf{R}(t+\tau, t) = e^{\mathbf{A}\tau}\mathbf{X}(t), \quad \tau \geq 0 \qquad (48)$$

$$\mathbf{R}(t, t+\tau) = \mathbf{X}(t)e^{\mathbf{A}^T\tau}, \quad \tau \geq 0 \qquad (49)$$

where $\mathbf{R}(t+\tau, t) \triangleq \mathrm{E}[\mathbf{x}(t+\tau)\mathbf{x}^T(t)]$. Eqs. (48) and (49) hold for arbitrary, nonstationary processes; for stationary processes the explicit dependence on t goes away (i.e., $\mathbf{R}$ depends only on τ and $\mathbf{X}$ is constant).

Define the $4N\times4N$ matrix $\mathbf{T}(t, \tau)$, whose components are given by

$$T_{rs}(t, \tau) \triangleq$$

$$\frac{1}{6}\mathrm{E}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}[x_r(t + (i+2n)\tau_0) + x_r(t + i\tau_0) -$$

$$2x_r(t + (i+n)\tau_0)]\right\}$$

$$\left\{\frac{1}{n}\sum_{j=1}^{n}[x_s(t + (j+2n)\tau_0) + x_s(t + j\tau_0) -$$

$$2x_s(t + (j+n)\tau_0)]\right\}\Bigg], \qquad (50)$$

where $\tau = n\tau_0$ and τ_0 is the sampling time. Then $T_{4r-1,\,4r-1}$, $r = 1, 2, \ldots, N$ is TVAR for clock r. The other components of $\mathbf{T}$ are not of interest at present. Expanding the double summation in Eq. (50), interchanging the order of expectation and summation, using the definition of autocorrelation matrix, and assuming $\mathbf{x}$ is stationary produces (note that the t dependence is dropped from $\mathbf{T}$)

$$\mathbf{T}(\tau) = \frac{1}{6n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}[6\mathbf{R}((i-j)\tau_0) -$$

$$4\mathbf{R}((i-j+n)\tau_0) - 4\mathbf{R}((i-j-n)\tau_0) +$$

$$\mathbf{R}((i-j+2n)\tau_0) + \mathbf{R}((i-j-2n)\tau_0)] \qquad (51)$$

Substituting for $\mathbf{R}$ from Eqs. (48) and (49) and evaluating the summations produces

$$\mathbf{T}(\tau) = \mathbf{U}(\tau) + \mathbf{U}^T(\tau), \qquad (52)$$

where

$$\mathbf{U}(\tau) = \frac{1}{n}[(\mathbf{I} - e^{\mathbf{A}\tau_0})^{-1}\mathbf{X} + \frac{1}{6n^2}(\mathbf{I} - e^{\mathbf{A}\tau_0})^{-2}$$

$$e^{\mathbf{A}\tau_0}[-10\mathbf{I} + 15e^{n\mathbf{A}\tau_0} - 6e^{2n\mathbf{A}\tau_0} + e^{3n\mathbf{A}\tau_0}]\mathbf{X}. \qquad (53)$$

For the limit of infinite sampling rate ($\tau_0\to0$, $n\to\infty$, $n\tau_0 = \tau$), Eq. (53) simplifies to

$$\mathbf{U}(\tau) = -\frac{\mathbf{A}^{-1}\mathbf{X}}{\tau} + \frac{\mathbf{A}^{-2}}{6\tau^2}$$

$$[-10\mathbf{I} + 15e^{\mathbf{A}\tau} - 6e^{2\mathbf{A}\tau} + e^{3\mathbf{A}\tau}]\mathbf{X}. \quad (54)$$

The general evaluation of $\mathbf{T}(\tau)$ via Eqs. (52)-(54) requires the evaluation of $\mathbf{X}$ and the matrix exponential $e^{\mathbf{A}\tau}$. However, explicit analytic expressions may be obtained for the asymptotic cases $\tau\to\infty$ and $\tau\to 0$. The former is obtained by putting $e^{\mathbf{A}\tau}\to 0$ (using the fact that the eigenvalues of $\mathbf{A}$ are all real and negative; the latter is obtained by expanding $e^{\mathbf{A}\tau}$ to order τ^2 for small τ. The results are

$$\mathbf{T}(\tau) = \frac{\mathbf{A}^{-1}\mathbf{B}_w\mathbf{Q}_w\mathbf{B}_w^T\mathbf{A}^{-T}}{\tau} - \frac{5[\mathbf{A}^{-2}\mathbf{X} + \mathbf{X}\mathbf{A}^{-2T}]}{3\tau^2}, \quad \tau\to\infty \quad (55)$$

$$\mathbf{T}(\tau) = \frac{\tau}{6}\mathbf{B}_w\mathbf{Q}_w\mathbf{B}_w^T, \qquad \tau\to 0. \quad (56)$$

The first term of Eq. (55) is derived using the result $\mathbf{A}^{-1}\mathbf{X} + \mathbf{X}\mathbf{A}^{-T} + \mathbf{A}^{-1}\mathbf{B}_w\mathbf{Q}_w\mathbf{B}_w^T\mathbf{A}^{-T} = 0$, which is easily obtained from Eq. (47). Note that the second term of Eq. (55) still depends explicitly on $\mathbf{X}$.

Finally, Eqs. (55) and (56) may be evaluated explicitly for the cases of PD WPM, VCO WPM, and VCO WFM. For an arbitrary chain of clocks, with possibly different bandwidths and noise strengths, TVAR for the j^{th} clock in the chain, for the case $\tau\to 0$, is given by

$$\mathrm{TVAR}_j(\tau) = \frac{\tau}{6}[2\zeta_j^2\omega_{nj}^2 c_2^j + 2\pi^2 b_0^j + \frac{b_2^j}{2\tau_{wj}^2}]. \quad (57)$$

For PD WPM, TVAR for the j^{th} clock in the chain, for the case $\tau\to\infty$, is given by (to order $1/\tau$)

$$\mathrm{TVAR}_j(\tau) = \frac{\sum_{i=1}^{j} c_2^i}{2\tau}. \quad (58)$$

For the asymptotic behavior of VCO WPM and WFM as $\tau\to\infty$, the terms of order $1/\tau$ vanish; TVAR for these behaves asymptotically like $1/\tau^2$ and is given by the second term of Eq. (55). Unfortunately, it was not possible to find an explicit, analytical expression for $\mathbf{X}$. Finally, the second term of Eq. (57) contains a factor of $4\pi^2$ that relates the WFM power spectral density coefficient with the equivalent white noise variance parameter in the state space description and each term of Eqs. (57)-(58) contains a factor of ½ that relates the white noise PSD and variance parameter (see Eq. (46)).

Eq. (57) shows that, over short time intervals, TVAR for a clock in the chain is independent of any preceding clocks. This is because each clock has a low-pass input-output characteristic (see Eqs. (6) and (7)); over short time intervals the input to a clock (from preceding clocks) is attenuated. In addition, TDEV for PD WPM is proportional to clock bandwidth (TVAR is proportional to ω_{nj}^2); TDEV is lower for smaller bandwidths due to the low-pass transfer characteristic between the PD noise and the clock output (see Eqs. (6) and (7)).

Finally, VCO WPM and WFM are unattenuated over short times (TVAR for these is independent of bandwidth) due to the high-pass transfer characteristic between the VCO noise and the clock output (see Eqs. (6) and (8)).

Eq. (58) shows that, over long time intervals, TVAR for PD WPM at the output of a particular clock in the chain accumulates as the sum of the PD WPM noise strengths for all the clocks in the chain up to and including that clock. If the clocks all have the same PD WPM noise strength, then TVAR at the end of the chain is proportional to the number of clocks in the chain. As above, this is due to the low-pass transfer characteristic. Note also that for PD WPM, TVAR over long times is independent of bandwidth.

Simulation Results

A number of clock chain cases were simulated to study the dependence of TDEV on noise type, clock bandwidth, number of clocks, and quantization. Noise strengths were chosen to be $c_2 = 1.0\times10^{-17}$ rad^2/Hz, $b_2 = 3.2\times10^{-16}$ rad^2/Hz, and $b_0 = 4.0\times10^{-34}$ rad^2/Hz^3. The first two are the same as values used in [19]; the third is somewhat larger. A bandwidth of 100 KHz was chosen for the PD WPM band-limiting filter; this is consistent with [19]. Note that for this study, the main interest was in the qualitative behavior of TDEV rather than the exact magnitudes. Each simulation case considered only a single noise type. A 1 ms time step was used in each simulation; this was also used for the sampling time when TDEV was calculated from the simulation output data.

Figure 3 shows results for PD WPM noise accumulation for chains consisting of 1, 10, and 100 linear PLLs. Each case was run for bandwidths of 0.1 Hz, 1 Hz, and 10 Hz. The analytical expressions for the asymptotic results Eqs. (57)-(58) are also shown (as solid lines) for comparison. As expected, TDEV increases like $\tau^{1/2}$ for

small τ and decreases like $\tau^{-1/2}$ for large τ. In addition, TDEV is proportional to bandwidth for small τ and to the square root of the number of clocks for large τ. Note that for each simulation, TDEV is larger than the corresponding analytic asymptotic result for short times approaching 1 ms; this is because the analytic results assume infinite sampling rate while a 1 KHz sampling rate (1 ms time step) was used in the simulations. Aside from this, the agreement between the simulation and analytic asymptotic results is excellent. Finally, as clock bandwidth decreases the time of the peak TDEV increases; the time of the peak varies approximately inversely with bandwidth. This is consistent with the notion that, when τ is considered to be large or small, it is always in comparison with one over the bandwidth (i.e., with a characteristic time constant of the system).

Figure 4 shows results for VCO WPM noise accumulation for the same cases as in Figure 3. For large τ, TDEV is proportional to τ^{-1} as expected. For small τ, TDEV appears to be proportional to $\tau^{-1/2}$ rather than $\tau^{1/2}$. To observe the $\tau^{1/2}$ behavior, a time step and sampling time small compared to the time constant of the 100 kHz band-limiting filter would have to be used. However, for short times, TDEV is still independent of number of clocks and clock bandwidth (consistent with Eq. (57)). For large τ, TDEV appears to be inversely proportional to clock bandwidth. This is due to the high-pass characteristic between the VCO and clock output. In addition, TDEV increases more slowly than the square root of the number of clocks. This is in contrast to the PD WPM case; apparently, VCO WPM is filtered somewhat by subsequent clocks over longer time intervals.

Figure 5 shows results for VCO WFM noise accumulation for the same cases as in Figure 3. For large τ, TDEV is proportional to τ^{-1} as expected. For small τ, TDEV is proportional to $\tau^{-1/2}$ as expected. For short times, TDEV is independent of number of clocks and clock bandwidth (consistent with Eq. (57)). For large τ, TDEV appears to increase proportionally with the square root of the number of clocks. Apparently, VCO WFM is not filtered significantly by subsequent clocks over long time intervals. Finally, TDEV decreases with clock bandwidth for large τ; the decrease appears to be faster than one over the bandwidth.

Figure 6 shows results for PD WPM noise accumulation for chains consisting of 1, 10, and 100 PLLs with phase detector quantization. In all cases the clock bandwidth was 1 Hz. For each clock chain, cases were run with quantization size (q^j in Eq. (2)) equal to one and two times the PD WPM standard deviation ($\sigma_{discrete}$ in Eq. (46)).

For comparison, corresponding cases without PD quantization are shown (these are reproduced from Figure 3). These values were chosen for PD quantization so that neither the quantization nor PD WPM noise would dominate. It was found that the effect of quantization is qualitatively like the effect of noise; if q^j is large, quantization dominates and the noise is negligible; if q^j is small, quantization is negligible.

For small τ, the behavior in Figure 6 is similar to that in Figure 3; TDEV is proportional to $\tau^{1/2}$ and is independent of the number of clocks. Note that for small τ TDEV is independent of q^j (and of whether quantization is even done). For large τ, quantization causes TDEV to increase. As in the case of no quantization, TDEV increases with the number of clocks; however, no conclusion can be drawn as to the exact dependence on the number of clocks. There appears to be greater peaking in TDEV compared to the case of no quantization; the peaking increases with both the number of clocks and q^j.

Figure 7 shows results for VCO WPM noise accumulation for chains consisting of 1, 10, and 100 PLLs with phase detector quantization. In all cases the clock bandwidth was 1 Hz. In correspondence with the PD WPM cases, q^j was chosen to be equal to one and two times the VCO WPM standard deviation. For comparison, corresponding cases without PD quantization are shown (these are reproduced from Figure 4). For small τ, the behavior in Figure 7 is similar to that in Figure 4; TDEV is proportional to $\tau^{-1/4}$ and is independent of the number of clocks. As in the cases with no quantization, the $\tau^{1/2}$ behavior would be observed only if a time step small relative to the 100 kHz VCO WPM filter bandwidth was used. For large τ, quantization causes TDEV to increase. As in the case of no quantization, TDEV increases with the number of clocks; however, no conclusion can be drawn as to the exact dependence on the number of clocks.

Figure 8 shows results for VCO WFM noise accumulation for chains consisting of 1, 10, and 100 PLLs with phase detector quantization. In all cases the clock bandwidth was 1 Hz. q^j was chosen to be equal to one and two times the VCO WFM standard deviation at 5.6 ms (i.e., 5.6 time steps); note that the VCO WFM is a random walk process and has a standard deviation that grows with time. For comparison, corresponding cases without PD quantization are shown (these are reproduced from Figure 5). For small τ, the behavior in Figure 8 is similar to that in Figure 5; TDEV is

proportional to $\tau^{1/2}$ and is independent of the number of clocks. For large τ, quantization causes TDEV to increase. As in the case of no quantization, TDEV increases with the number of clocks; however, no conclusion can be drawn as to the exact dependence on the number of clocks.

In Figures 6-8, TDEV is larger with PD quantization compared to cases without quantization and increases with q^j. This was observed for all three noise types. This is contrary to the results in [23], which show a decrease in TDEV when PD quantization is present compared to cases without quantization. The type of quantization assumed in [23] was not given, except for the statement that the quantization was nonlinear. Presumably, the nature of the nonlinearity is such that TDEV decreases rather than increases.

Finally, note that for the noise types considered here, TVAR decreases as τ^{-1} or τ^{-2} as $\tau \to \infty$. In fact, the only noise type mentioned here for which TVAR does not decrease as $\tau \to \infty$ is PD FPM; for this noise type TVAR remains constant as $\tau \to \infty$. For TVAR to increase as $\tau \to \infty$, noise such as PD WFM or PD FFM would have to be present. Of course, there is an increase in TVAR as $\tau \to \infty$ that reflects the long-term behavior of the PRC (the initial clock in the chain); this was neglected in this paper.

Summary and Conclusions

In this paper, noise accumulation in an arbitrary chain of clocks was considered. Both linear phase-locked loop models and and nonlinear PLL models with uniform phase detector quantization were developed. In both cases, the models were formulated as time-domain, state equations. For the linear models, both analytic results for the asymptotic behavior of TDEV and simulation results were obtained. For the nonlinear models, only simulation results were obtained. The noise types considered included PD WPM, VCO WPM, and VCO WFM.

Analytic and simulation results were consistent in all cases and were in excellent agreement when comparisons could be made. From the results, the following can be concluded:

1. For times short compared to the smallest time constant of the system (the reciprocal bandwidth of the clock or VCO WPM filter), TVAR is proportional to τ and is independent of the number of clocks in the chain and the size of the quantization. For PD WPM, TVAR is proportional to the square of the bandwidth; for VCO WPM and WFM, TVAR is independent of bandwidth.
2. For times long compared to the largest time constant of the system, TVAR is proportional to τ^{-1} for PD WPM and τ^{-2} for VCO WPM and WFM. For PD WPM and VCO WFM, there is little or no filtering by subsequent clocks over long times; if the clocks all have the same noise strength, TVAR is proportional to the number of clocks in the chain. For VCO WPM, TVAR increases more slowly than the number of clocks in the chain. In all cases over long times, TVAR increases when PD quantization is introduced and when the quantization size is increased.
3. For $\tau \to \infty$, TVAR does not increase asymptotically for any of the noise types considered here. The only noise type mentioned here where TVAR does not decrease (it remains constant) is PD FPM. For TVAR to decrease as $\tau \to \infty$, noise such as PD WFM or PD FFM would have to be present. Of course, there is an increase in TVAR as $\tau \to \infty$ that reflects the long-term behavior of the PRC (the initial clock in the chain); this was neglected in this paper.

References

[1] *Digital Hierarchy, Optical Interface Rates and Formats Specifications (SONET),* American National Standards Institute, ANSI Standard T1.105-199_ (Revision of T1.105-1988).

[2] *Synchronous Optical Network (SONET) Transport Systems: Common Generic Criteria (Incremental Proposed Requirements to TR-NWT-000253, Issue 2),* Bellcore, Technical Advisory TA-NWT-000253, Issue 8, October, 1993.

[3] *Synchronous Optical Network (SONET) Transport Systems: Common Generic Criteria (A Module of TSGR, FR-NWT-000440)* Bellcore, Technical Reference TR-NWT-000253, Issue 2, December, 1991.

[4] CCITT, Revised Recommendation G.707, *Synchronous Digital Hierarchy Bit Rates,* Geneva, 1992.

[5] CCITT, Revised Recommendation G.708, *Network Node Interface for the Synchronous Digital Hierarchy,* Geneva, 1992.

[6] CCITT, Revised Recommendation G.709, *Synchronous Multiplexing Structure,* Geneva, 1992.

[7] *Digital Network Synchronization Plan,* Bellcore, Technical Advisory TA-NWT-000436, Issue 2, June, 1993.

[8] *Synchronous Optical Network (SONET): Jitter at Network Interfaces,* ANSI T1.105.03-1994 (previously T1 LB 364/T1X1 LB93-03 DEFAULT).

[9] *Synchronization Interface Standard, ANSI T1.101-1994 (Revision of ANSI T1.101-1987),* (previously T1 Letter Ballot 312).

[10] *Clocks for the Synchronized Network: Common Generic Criteria,* Bellcore, Technical Reference TR-NWT-001244, Issue 1, June, 1993.

[11] E. W. Butterline, *Reach Out and Time Someone,* GPS WORLD, January, 1993, pp. 32-40.

[12] *ITU Draft Recommendation G.803, Architectures of Transport Networks Based on the Synchronous Digital Hierarchy (SDH),* ITU, COM XVIII-R 106-E, Geneva, July, 1992.

[13] G. H. Manhoudt, *Comparison Between ETSI and ANSI Requirements Concerning SDH Clock Bandwidths,* Netherlands Contribution, Delayed Document D.360, ITU-T SG XIII/WP 4, Q21, Geneva, Switzerland, March, 1994.

[14] P. Uythoven (contact), *New Edition of Draft Recommendation G.81s,* Contribution to ITU-T Q21/13 Expert Meeting, Rome, December 1-3, 1993.

[15] V. F. Kroupa, *Noise Properties of PLL Systems,* IEEE Transactions on Communications, Vol. CM-30, No. 10, October, 1982, pp. 2244-2252.

[16] R. E. Best, *Phase-Locked Loops, Theory, Design, and Applications,* McGraw-Hill, New York, 1993, 2nd Edition.

[17] 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,* Proceedings of the 45th Annual Symposium on Frequency Control, IEEE, May 29-31, 1991, pp. 667-678.

[18] P. Uythoven, *Accumulation of Random Wander in a Non-Homogeneous Chain of Slave Clocks,* CCITT Study Group XVIII, Working Party 6, Question 7/XVIII, Working Document, Melbourne, Australia, December, 1991.

[19] M. Carbonelli, D. DeSeta, D. Perucchini, *Slave Clock Noise Modeling for SDH Synchronization Network Design: Some Results on Clock Chain Performance,* ETSI/STC TM3, Temporary Document 63, Bristol, October 12-16, 1992.

[20] J. T. Harvey and J. W. Rice, *Random Timing Noise Growth in a Cascaded Digital Regenerator Chain,* IEEE Transactions on Communications, August, 1973, pp. 969-971.

[21] D. L. Duttweiler, *The Jitter Performance of Phase-Locked Loops Extracting Timing From Baseband Data Waveforms,* Bell System Technical Journal, Vol. 55, No. 1, January, 1976, pp. 37-58. IEEE Transactions on Communications, August, 1973, pp. 969-971.

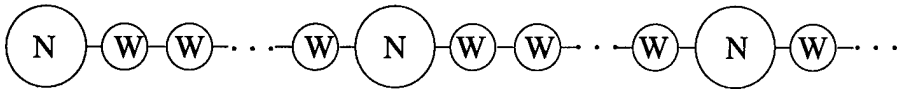
[22] T. Shimamura and I. Eguchi, *An Analysis of Jitter Accumulation in a Chain of PLL Timing Recovery Circuits,* IEEE Transactions on Communications, Vol. COM-25, No. 9, September, 1977, pp. 1027-1032.

[23] G. Zampetti and D. Chislow, *Cascaded Clocks Measurement and Simulation Findings,* Time and Frequency Workshop on Synchronization Standards for Telecommunications Systems, NIST, Time and Frequency Division, February 28 - March 2, 1994. D. W. Allan, M. A. Weiss, and J. L. Jespersen,

[24] D. G. Schultz and J. L. Melsa, *State Functions and Linear Control Systems,* McGraw-Hill, New York, 1967, 2nd Edition.

[25] Averill M. Law and W. David Kelton, *Simulation Modeling & Analysis,* McGraw-Hill, New York, 1991.

[26] A. E. Bryson, Jr. and Yu-Chi Ho, *Applied Optimal Control, Optimization, Estimation, and Control,* Hemisphere Publishing Corporation, New York, 1975, Revised Printing.



N = Narrow Band Clock (e.g., BITS, SSU)
W = Wide Band Clock (e.g., SONET NE Clock, SEC)
Number of Narrow Band Clocks = P
Number of Wide Band Clocks Between Two Narrow Band Clocks = R
Total Number of Clocks in Chain = $P(R + 1)$
SONET Networks: $R \leq 1$, $P \leq 5$ or 6
SDH Networks: $R \leq 20$, $P \leq 10$, $PR \leq 60$

Figure 1. Schematic of SONET or SDH Timing Chain.

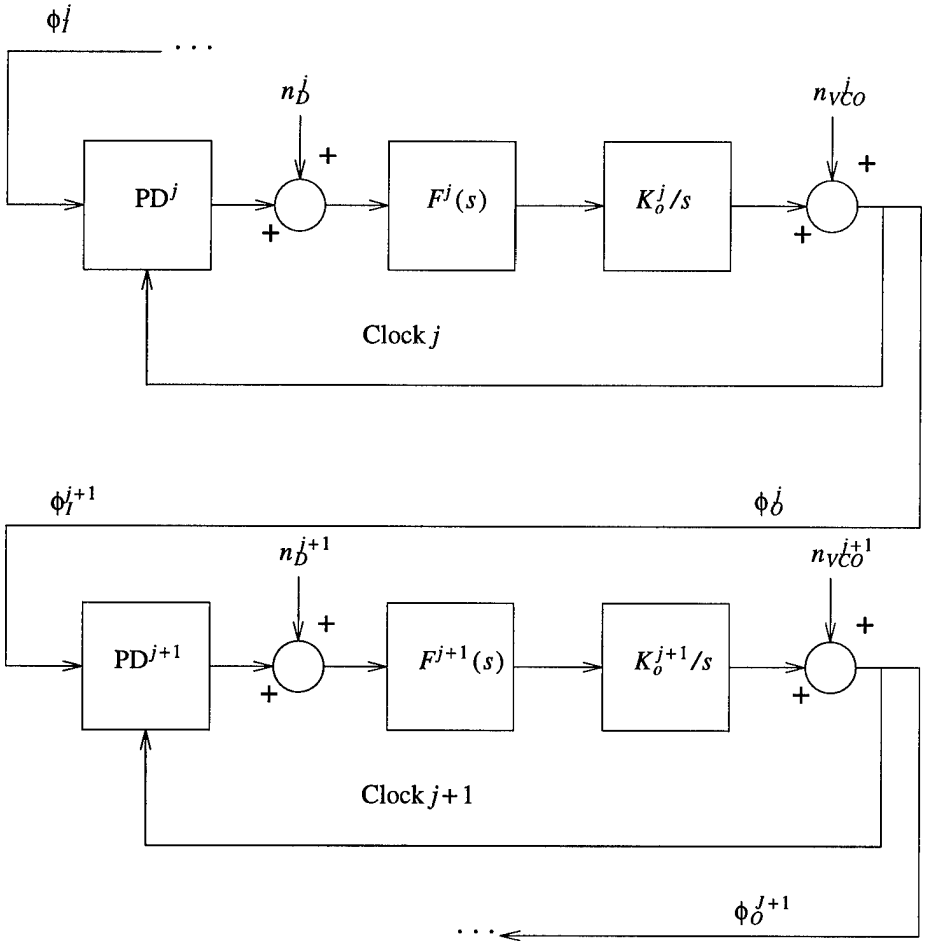


Figure 2. Two Successive Clocks in a Chain of N Clocks

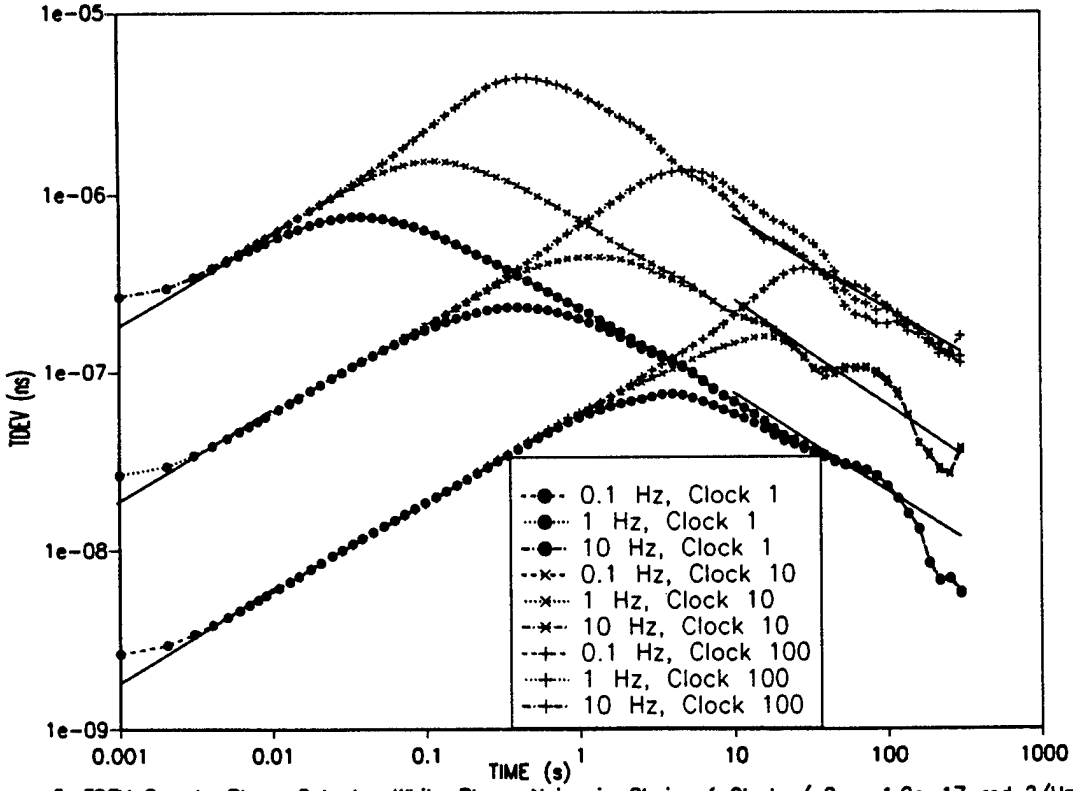


Figure 3. TDEV Due to Phase Detector White Phase Noise in Chain of Clocks (c2 = 1.0e-17 rad^2/Hz)

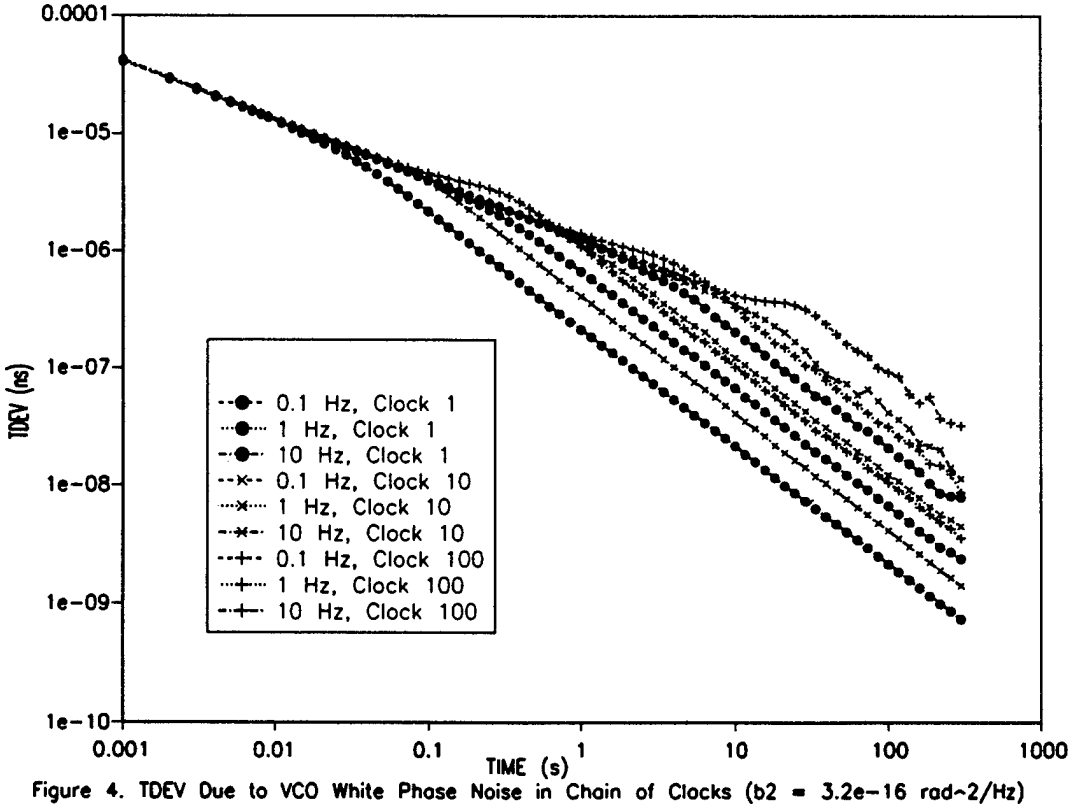


Figure 4. TDEV Due to VCO White Phase Noise in Chain of Clocks (b2 = 3.2e-16 rad^2/Hz)

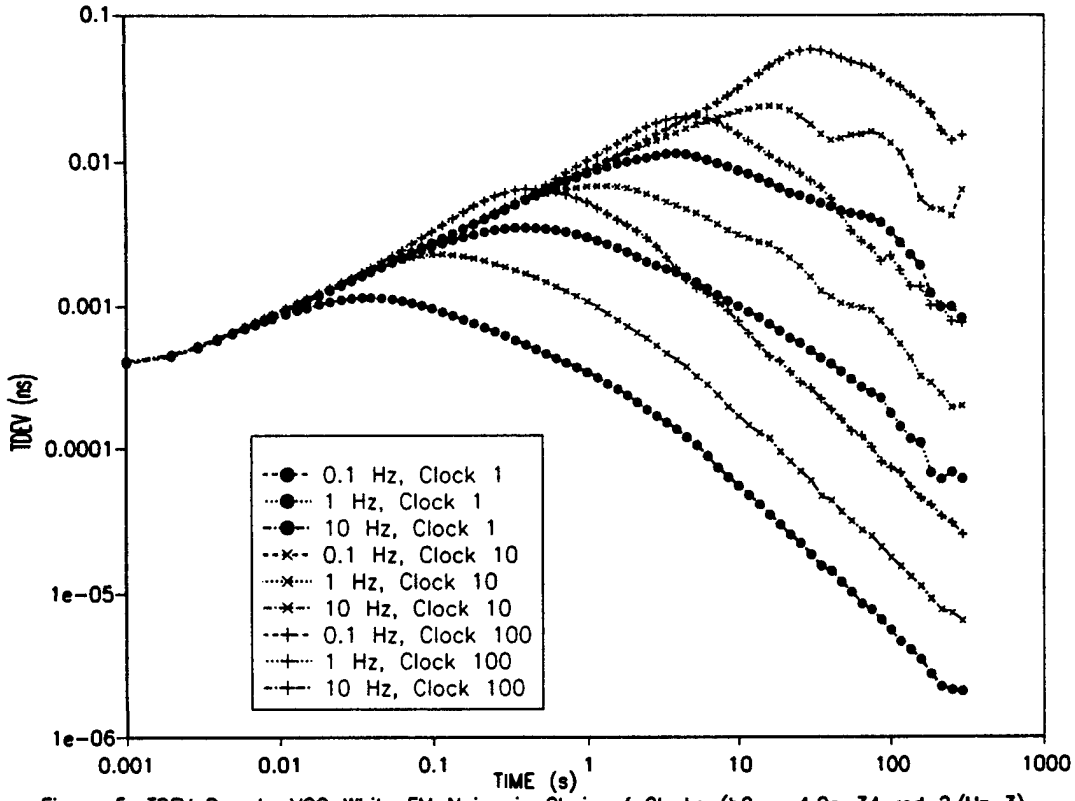


Figure 5. TDEV Due to VCO White FM Noise in Chain of Clocks (b0 = 4.0e-34 rad^2/Hz^3)

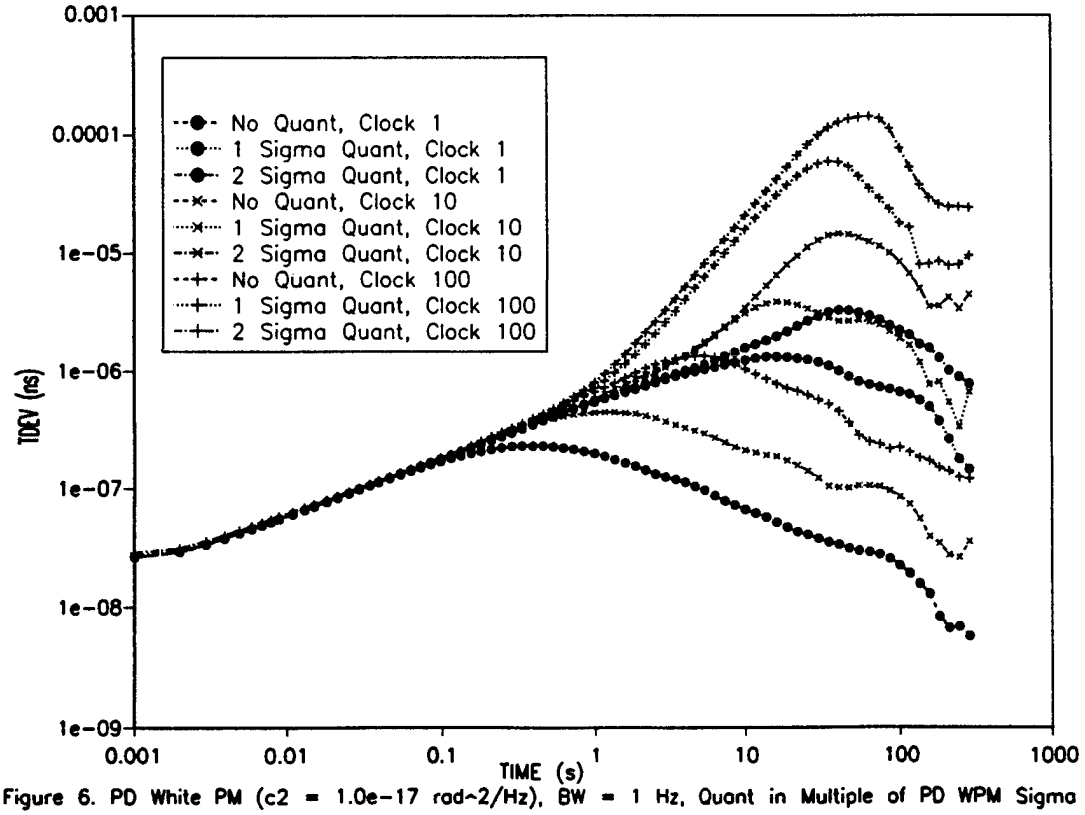


Figure 6. PD White PM (c2 = 1.0e-17 rad^2/Hz), BW = 1 Hz, Quant in Multiple of PD WPM Sigma

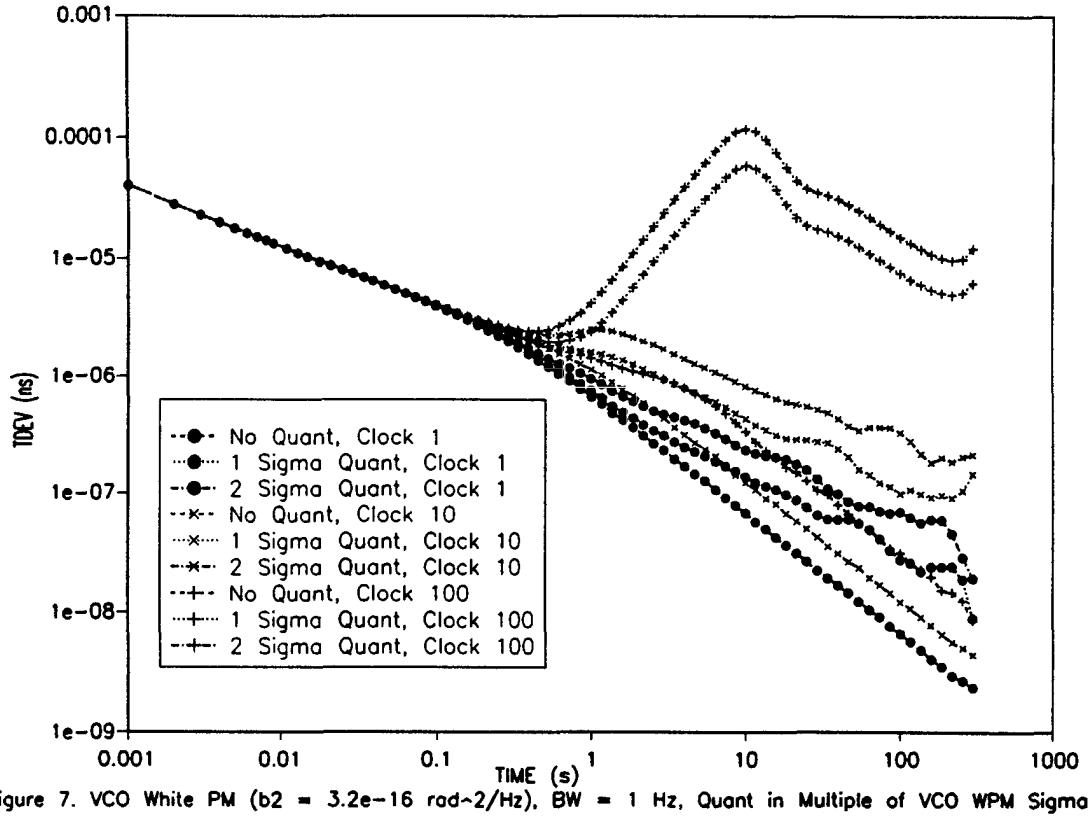


Figure 7. VCO White PM (b2 = 3.2e-16 rad^2/Hz), BW = 1 Hz, Quant in Multiple of VCO WPM Sigma

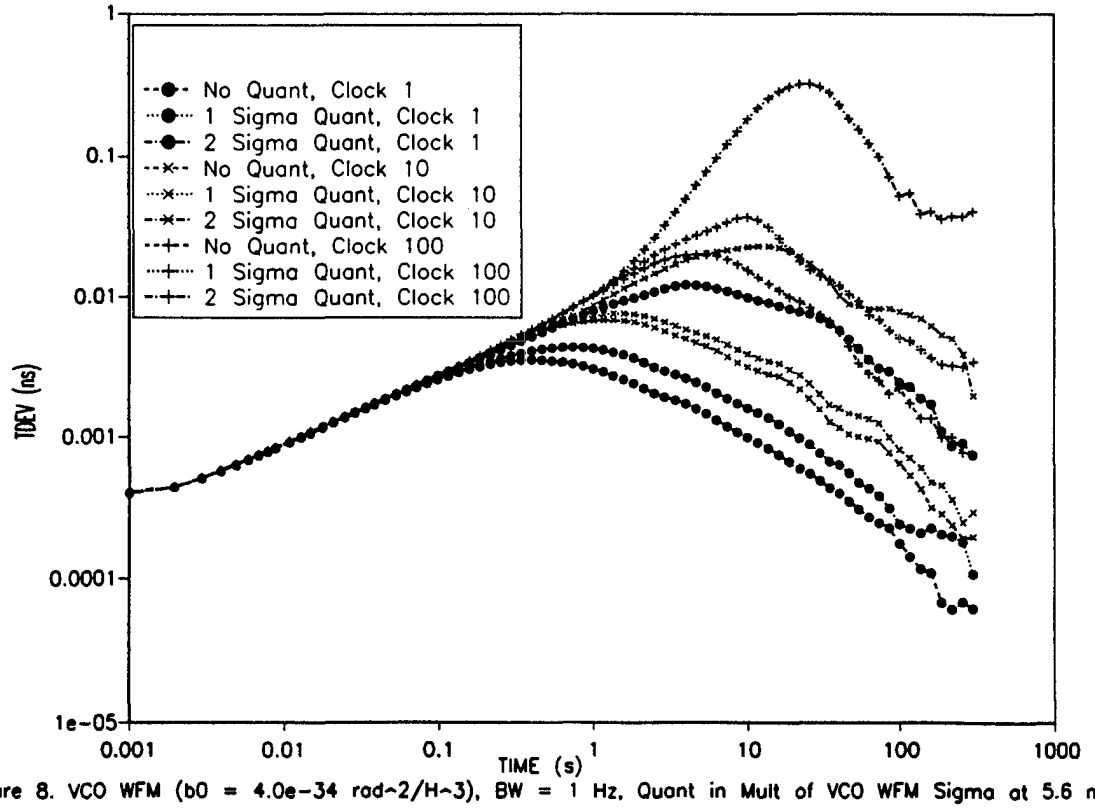


Figure 8. VCO WFM (b0 = 4.0e-34 rad^2/H^3), BW = 1 Hz, Quant in Mult of VCO WFM Sigma at 5.6 ms

IEEE International Frequency Control Symposium
PROCEEDINGS ORDERING INFORMATION

NO.	YEAR	DOCUMENT NO.	OBTAIN FROM*	PRICE MICROFICHE / HARDCOPY
10	1956	AD-298322	NTIS	$9.00 /$59.00
11	1957	AD-298323	NTIS	9.00 / 66.00
12	1958	AD-298324	NTIS	9.00 / 66.00
13	1959	AD-298325	NTIS	9.00 / 73.00
14	1960	AD-246500	NTIS	9.00 / 50.00
15	1961	AD-265455	NTIS	9.00 / 43.00
16	1962	AD-285086	NTIS	9.00 / 50.00
17	1963	AD-423381	NTIS	9.00 / 66.00
18	1964	AD-450341	NTIS	9.00 / 66.00
19	1965	AD-471229	NTIS	9.00 / 66.00
20	1966	AD-800523	NTIS	9.00 / 66.00
21	1967	AD-659792	NTIS	9.00 / 59.00
22	1968	AD-844911	NTIS	9.00 / 66.00
23	1969	AD-746209	NTIS	9.00 / 43.00
24	1970	AD-746210	NTIS	9.00 / 43.00
25	1971	AD-746211	NTIS	9.00 / 43.00
26	1972	AD-771043	NTIS	9.00 / 43.00
27	1973	AD-771042	NTIS	9.00 / 50.00
28	1974	AD-A011113	NTIS	9.00 / 50.00
29	1975	AD-A017466	NTIS	9.00 / 50.00
30	1976	AD-A046089	NTIS	9.00 / 59.00
31	1977	AD-A088221	NTIS	9.00 / 66.00
32	1978	AD-A955718	NTIS	9.00 / 59.00
33	1979	AD-A213544	NTIS	9.00 / 73.00
34	1980	AD-A213670	NTIS	9.00 / 59.00
35	1981	AD-A110870	NTIS	19.00 / 73.00
36	1982	AD-A130811	NTIS	9.00 / 59.00
37	1983	AD-A136673	NTIS	9.00 / 59.00
38	1984	AD-A217381	NTIS	12.50 / 43.00
39	1985	AD-A217404	NTIS	19.00 / 59.00
40	1986	86CH2330-9	IEEE	70.00 / 70.00
41	1987	AD-A216858	NTIS	17.00 / 59.00
42	1988	AD-A217275	NTIS	17.00 / 59.00
43	1989	AD-A235629	NTIS	19.00 / 66.00
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	TBD / TBD

*NTIS - National Technical Information Service
5285 Port Royal Road, Sills Building
Springfield, VA 22161, U.S.A.
Tel: 703-487-4650

*IEEE - Inst. of Electrical & Electronics Engineers
445 Hoes Lane
Piscataway, NJ 08854, U.S.A.
Tel: 800-678-4333 or 908-981-0060

An **index** to all the papers in the Proceedings since 1956, and other Symposium information are available as ASCII files on a 5 1/4" (13 cm) MS-DOS format floppy disk, for $5-, from Synergistic Management, Inc., 3100 Route 138, Wall Township, NJ 07719.

SPECIFICATIONS AND STANDARDS RELATING TO FREQUENCY CONTROL

INSTITUTE OF ELECTRICAL AND ELECTRONIC ENGINEERS (IEEE)

Order from: IEEE Service Center
445 Hoes Lane
Piscataway, NJ 08854
(201) 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: Military Specifications and Standards
700 Robbins Ave., Bldg. 4D
Philadelphia, PA 19111-5094
(215) 697-2667/2179 - Customer Service
(215) 697-1187 to -1195 Telephone Order Entry System

MIL-C-3098 Crystal Unit, Quartz, Gen Spec for

MIL-C-49468 Crystal Units, Quartz, Precision, 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 Pas, Band Pass Suppression and Dual Processing, Gen Spec for

MIL-F-28861 Filters and Capacitors, Radio Frequency/Electromagnetic 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-683 Crystal 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
(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 ELECTROTECHNICAL COMMISSION (IEC)

Order from: American Nat'l. Standard Inst.
(ANSI), 1430 Broadway
New York, NY 10018
(212) 354-3300

IEC PUBLICATIONS

122: Quartz crystal units for frequency control and selection.

122-1 (1976) Part 1: Standard values and test conditions. Amendment No. 1 (1983).

122-2 (1983) Part 2: Guide to the use of quartz crystal
units for frequency control and selection.

122-3 (1977) Part 3: Standard outlines and pin connections. Amendment No. 1 (1984)

122-3A (1979) First supplement.

122-3B (1980) Second supplement.

122-3C (1981) Third supplement.

283 (1986) Methods for the measurement of frequency & equivalent resistance of unwanted resonances of filter crystal units.

302 (1969) Standard definitions & methods of measurment 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 (1982) Part 1: General information, standard values and test conditions.

368A (1973) First supplement. Amendment No.1 (1977). Amendment No.2 (1982).

368B (1975) Second supplement.

368-2 Part 2: Guide to the use of piezoelectric filters.

368-2-1 (1988) Section One-Quartz crystal filters

368-3 (1979) Part 3: Standard outlines.

368-3A (1981) First supplement.

444: Measurement of quartz crystal unit parameters by zero phase technique in a π-network.

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 Co.

444-4 (1988) Part 4: Method for the measurement of the load resonance frequency fL, load

resonance resistance RL 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.

679: Quartz crystal controlled oscillators.

679-1 (1980) Part 1: General information, test conditions & methods.

679-2 (1981) Part 2: Guide to the use of quartz crystal controlled oscillators. Amendment No. 1 (1985).

689 (1980) Measurements and test methods for 32 kHz quartz crystal units for wrist watches and standard values.

758 (1983) Synthetic quartz crystal. Chapter I: Specification for synthetic quartz crystal. Chapter II: Guide to the use of synthetic quartz crystal. Amendment No. 1 (1984).

862: Surface acoustic wave (SAW) filters.

862-1-1 (1985) Part 1: General information, test conditions and methods.

862-3 (1986) Part 3: Standard outlines.

CONSULTATIVE COMMITTEE ON INTERNATIONAL RADIO (CCIR)

Order from: International Telecommunications Union
General Secretariat - Sales Section
Place des Nations
CH-1211 Geneva
SWITZERLAND

Ask for CCIR 17th Plenary Assembly, Volume VII, "Standard Frequencies and Time Signals (Study Group 7)", which contains all of the following documents:

RECOMMENDATION 457-1	Use of the Modified Julian Date by the Standard-Frequency and Time-Signal Services
RECOMMENDATION 458-1	International Comparisons of Atomic Time Scales
RECOMMENDATION 460-4	Standard Frequency and Time Signal Emissions [Note: defines the UTC system]
RECOMMENDATION 485-1	Use of Time Scales in the Field of Standard-Frequency and Time Services
RECOMMENDATION 486-1	Reference of Precisely Controlled Frequency Generators and Emissions to the International Atomic Time Scale
RECOMMENDATION 535-1	Use of the Term UTC
RECOMMENDATION 536	Time Scale Notations
RECOMMENDATION 538	Frequency and Phase Stability Measures
REPORT 580-2	Characterization of Frequency and Phase Noise
RECOMMENDATION 583	Time Codes
RECOMMENDATION 685	International Synchronization of UTC Time Scales
RECOMMENDATION 686	Glossary

AUTHOR INDEX

Abe, H. 289
Abramson, I.V. 578
Adachi, M. 296
Agatsuma, K. 99
Albert, W.C. 415
Allan, D.W. 572
Andersen, B.D. 308
Andres, D. 352,530
Andreyev, I.A. 43
Araki, A. 223
Arnold, D. 684
Avramov, I.D. 379

Bagaev, V.P. 578
Ballandras, S. 184
Ballato, A. 22,85,229
Bassevich, A.B. 776
Batra, R.C. 201
Baudrand, H. 555
Beard, R. 724
Beijerinck, H.C.W. 651
Belkerdid, M.A. 308
Benz, S.P. 666
Bergquist, J.C. 744
Bigler, E. 184
Blair, D.G. 441,459,500,731
Blin, T. 559
Bogdanov, P.P. 776
Bohuszewicz, T.V. 122
Booi, P.A.A. 666
Boulanger, J.-S. 695
Boyd, D.A. 709
Bradaczek, H. 237
Brendel, R. 260,559
Brunet, M. 559
Buisson, J.A. 782
Busca, G. 716
Butler, C.T. 600

Caldwell, S. 641
Capelle, B. 48,58,315
Cavin, M. 308
Change, C-R. 548
Choi, W-S. 139
Cole, A.L. 677
Coles, C. 731
Costa, M.E. 441
Coyle, L.M. 684,709
Crampton, S.B. 677
Cretin, B. 260
Croitoru, N. 343
Cutler, L.S. 572

Danicki, E. 337
Danzy, F. 600,724
DeMarchi, A. 567
Denis, J.P. 48,58
Detaint, J. 48,58,315
Dick, G.J. 421,482,493,739,755
Dobershtein, S. 330
Dothee, D. 91
Doughty, D.H. 122
Douglas, R.J. 695
Driessen, J.P.J. 651
Driscoll, M.M. 647
Drullinger, R. 658,662
Dubovik, M.F. 43
Dulmet, B. 172

Eer Nisse, E.P. 286
ElHabti, A. 260
Endres, T.J. 624
Ermakov, Y.M. 405

Fan, S-J. 78
Ferris, J.E. 115
Fisk, P.T.H. 731
Foerster, J. 401
Frank, A. 724
Frolov, V.S. 405
Frye, G.C. 122

Garner, G.M. 798
Gavignet, E. 184
Gerber, J. 548
Gevorkyan, A.G. 776
Giffard, R. 572
Gil, L.M. 482
Gniewinska, B. 608
Golding, W.M. 600,655,724
Gouzhva, Y.G. 776
Grigorjev, L.V. 405
Gualtieri, J.G. 22

Habic, D. 587
Hall, R.B. 624
Handel, P.H. 8,539
Hanji, M. 223
Hardy, W.N. 670
Haruvi-Busnach, E. 343
Hayden, M.E. 670
Hazzard, A.C. 647
Heighway, J. 374
Hendrickson, M. 245
Hikita, K. 72
Hildebrandt, G. 237
Ho, K. 572
Hoffman T.E. 709
Hongwei, S. 544
Hoogerland, M.D. 651
Howe, D.A. 791
Howe, R.T. 2,127
Hürlimann, M.D. 670
Hurrell, J.P. 769
Huynh, D. 401

Ibanez, A. 48
Ishigami, S. 99
Itano, W.M. 744
Ivanov, E.N. 433,500

Jaduszliwer, B. 769
Janushevsky, G.D. 524
Jendrock, R.F. 709

Kalachev, A. 687
Kalashnikova, I.I. 40
Kalinichev, V.I. 500
Kaneko, T. 142
Karaulnik, A.E. 281
Karaulnik, M.E. 281
Kastberg, A. 655
Kawabata, A. 296
Kerman, A.J. 677
Kolosovsky, E.A. 345
Komatsu, R. 72
Kondratyev, S.N. 374
Kosinski, J.A. 85,229
Kosykh, A.V. 578
Kozlov, A.S. 345,395
Kroupa, V.F. 613
Krupczak, J.J. 677
Kuleshov, V.N. 524
Kusters, J.A. 572

Laffey, S. 107,245
Lareau, R. 107
Lawn, M.A. 731
Lee, P.C.Y. 192
Lee, W.D. 658,662
Liang, J-F. 478
Lindberg, L. 273
Lindenmuth, R. 401
Lopez, A.M. 624
Lowe, J.P. 662
Lu, Y. 85,401
Luiten, A.N. 441
Lukaszek, T. 401
Lwo, B.J. 165,172

Maleki, L. 739,747,755,761
Malocha, D.C. 308

Malyukhov, V. 330
Mann, A.G. 441,731
Mansfeld, G.D. 35
Marianneau, G. 559
Marihin, V. 687
Martin, J.J. 115
Martin, S.J. 122
Masiukiewicz, A. 608
Mateescu, I. 229
Mattison, E.M. 684,709
McAllaster, D.R. 677
McCaskill, T.B. 782
McGowan, R. 401
Megens, H.J.L. 651
Meguro, T. 251
Mercer, C.D. 286
Merchant, P. 135
Mileti, G. 716
Mizan, M. 466
Momosaki, E. 142
Montress, G.K. 359,365,379
Morys, B. 237
Mostiaev, V.A. 405
Murray, R.A. 107
Muzhdoba, D. 687
Myasnikov, A. 687
Myasnikova, L. 687

Nakase, T. 323
Nakazawa, M. 223
Naumov, V.S. 40
Nelson, C.W. 567
Nefedov, V.A. 301
Neubig, B. 268
Nguyen, C.T-C. 127
Nystrom, G.U. 709

Oaks, O.J. 782
Odagawa, H. 251
Ohmura, M. 289
Ohtsuka, K. 296
Oomura, Y. 255
Opdycke, D.G. 647
Opie, D. 600
Orlov, A. 411
Osegueda, R.A. 482,493
Oura, N. 323

Parker, T. 352,359,365,379,530
Pashkov, S.S. 40
Peng, H. 459
Percival, D.B. 791
Philippot, E. 48,58
Phillips, W.D. 655
Pierce, D. 107
Pierluissi, J.H. 482,493
Pilard, M. 315
Plessky, V.P. 374
Poitzsch, M.E. 744
Portnoff, G. 273
Postnikov, I.I. 209
Powers, E. 724
Prestage, J.D. 739,747,755

Reid, W.G. 782
Repinsky, S.M. 395
Revilla, A. 482,493
Rochat, P. 716
Rohde, U.L. 548
Rolston, S.L. 655
Ruby, R. 135

Saitou, H. 289
Santiago, D.G. 421,482
Sato, M. 99
Schuwer, M.P. 651
Schwartzel, J. 48,58,315
Schweda, I. 716
Searls, J.H. 433
Seidel, D.J. 761
Seidman, A. 343
Sekimoto, H. 255
Sellars, M.J. 731
Sharp, M. 600
Shestopalov, K.V. 301
Shimoda, T. 142
Shiosaki, T. 296
Shirley, J.H. 658
Shmaly, Yu. S. 43
Siccardi, M. 567
Skripov, V. 411
Slawsby, N. 633
Smagin, A. 541
Smith, R.C. 684
Smits, J.G. 139
Solal, M. 315
Spreeuw, R. 655
Stemniski, J.R. 241
Stewart, J.T. 149
Sugihara, T. 72
Sun, Q-F. 478
Sun, R-Y. 78
Svetlov, V. 687
Szulc, W. 608

Takahashi, E. 594
Takakuwa, S. 223
Tanski, W.J. 472
Taziev, R.M. 345
Teshigawara, O. 594
Tiersten, H.F. 165,172
Tjoelker, R.L. 739,755
Tobar, M.E. 433
Tsarapkin, D.P. 447,451
Tsukamoto, K. 99
Tyulyakov, A.E. 776

Uchiyama, F. 99
Uda, S. 72

van Leeuwen, K.A.H. 651
Vasiljevic, D.M. 587
Vessot, R.F.C. 684,709
Viard, B. 91
Vig, J.R. 107,245,506
Villalva, G.J. 482,493
Vlassov, Yu.N. 395
Volkov, S.V. 405
Vredenbregt, E.J.D. 651

Wagatsuma, Y. 251
Walls, F.L. 379,506,539,567,662
Wang, C. 478
Wang, J. 78
Wang, M. 604
Wang, R.T. 421,482
Warren, H.E. 782
Watanabe, Y. 255
Weinert, R.W. 478
Weisenbach, L. 122
White, J. 724
Wineland, D.J. 744
Woode, R.A. 433
Wu, X.Y. 486

Yagoub, M.C.E. 555
Yakovkin, I.B. 395
Yamamoto, K. 251
Yamanouchi, K. 251
Yamashita, K. 594
Yamni, K. 91
Yang, J.S. 201,486
Yates, J. 600
Yu, J.D. 192

Zadneprovsky, B. 301
Zaki, K.A. 478
Zarka, A. 48,58,315
Zecchini, P. 91
Zhou, W. 604
Zhu, J. 359
Zilberstein, M. 343